Coherent

Coordinated Control and Spectrum Management
for 5G Heterogeneous Radio Access Networks

Grant Agreement No.: 671639
Call: H2020-ICT-2014-2

Deliverable D2.2
System Architecture and Abstractions for Mobile Networks

Version:	1.0
Due date:	30.06.2016
Delivered date:	20.07.2016
Dissemination level:	PU

The project is co-funded by

European Commission

Authors

Alexandros Kostopoulos, George Agapiou, (OTE)
Deng Junquan (AALTO)
Dorin Panaitopol (TCS)
Fang-Chun Kuo (EICT, Editor-in-Chief)
Kostas Katsalis, Navid Nikaein (EUR)
Mariana Goldhamer (4GC)
Tao Chen (VTT)
Rebecca Steinert (SICS)
Roberto Riggio (CNET)

Coordinator

Dr. Tao Chen

VTT Technical Research Centre of Finland Ltd

Tietotie 3

02150, Espoo

Finland

Email: tao.chen@vtt.fi

Disclaimer

Acknowledgement

This report is funded under the EC H2020 5G-PPP project COHERENT, Grant Agreement No. 671639.

Version history

Version	Date	Remarks
0.0	20.04.2016	ToC created.
0.1.0	23.05.2016	Integrated inputs from EICT, 4GC and SICS
0.1.2	01.06.2016	Integrated inputs from VTT, OTE, AALTO, EUR, CNET
0.2.0	13.06.2016	Integrated inputs from EUR, 4GC, SICS, TCS
0.2.1	14.06.2016	Editorial work with figures and tables
0.3.0	16.06.2016	Internal review
0.3.1	21.06.2016	Updated inputs for OTE, EUR, TCS, 4GC (6.2 is not included), INEA's review (except 2.3.1).
0.3.2	28.06.2016	Updated input from TCS, EUR, CNET (section 3), EICT
0.3.3	29.06.2016	Integrated internal reviews
0.3.4	29.06.2016	Integrated 4GC input on 29.06.2016
0.3.5	30.06.2016	Integrated inputs from SICS and TCS
0.4.0	03.07.2016	External review
0.4.1	04.07.2016	Modified Sec 2.3.1, 4.1, 5.3, 5.4, added Annex A
0.4.2	07.07.2016	Updated figure in sec 7 (4GC) and removed figure in section 5.3 (VTT)
0.4.3	13.07.2016	Integrated external reviews from Antonio and Adrian. considerable changes in Section 4.1, 5.1, 5.2, 5.3 and 7 (conducted by Tao)
1.0	17.07.2016	Final version

Executive summary

Mobile network operators face a transformation from the "always best connected, faster, smarter bit pipes" paradigm that dominated the 3G/4G evolution, toward 5G which is envisioned by NGMN as an inclusive "end-to-end ecosystem". The envisioned 5G mobile networks must be agile in the spatial, frequency and temporal dimensions in order to allow elastic adaptation to network infrastructure particulars, flexible arrangement of spectrum use and dynamic adjustment to service changes. This deliverable presents the design of system architecture and abstractions in the COHERENT project that will enable the realisation of agile 5G mobile networks by embracing RAN heterogeneity, introducing programmability in spectrum management and deploying RAN service orchestration.

The COHERENT system architecture is designed based on three key concepts: i) **control separation** (logically centralised control and real-time control), ii) **network abstractions** and iii) **network slicing**. More specifically, a programmable *control and coordination* is required for heterogeneous 5G radio access networks. *Abstraction* is the key to drive such programmability in control and coordination for 5G systems. Accordingly, any RAN operations desired by the *network slices* could be facilitated because the abstracted (physical and virtualised) infrastructure simplifies the implementation and deployment of advanced control and coordination functions by hiding the configuration details of RAN hardware.

A number of different use cases, including multi-tenancy, Device-to-Device (D2D) communications, broadcast operation, UE-relaying operation for public safety service, multi-connectivity and mobility management are presented for showing how the COHERENT architecture can support them. Furthermore, COHERENT intends to influence the philosophy behind the next 5G 3GPP system. Therefore, in this document, the contribution of COHERENT to the 3GPP framework is also introduced based on COHERENT principles. Finally, we introduce the network abstractions which are used in the COHERENT system for heterogeneous mobile networks programmability, including the concepts of domain specific languages (DSLs), the COHERENT semantic model and the northbound interface as well as the primitives which allow the flexibility on manipulating the semantic models in order to configure the state of the networks for the network programmer.

Table of contents

List of abbreviations

2G	2nd Generation Mobile Networks
3D	Three-Dimensional
3GPP	Third Generation Partnership Project
4G	4th Generation Mobile Networks
5G	5th Generation Mobile Networks
5GPoAs	5G Point of Access
5G PPP	5G Infrastructure Public Private Partnership
AI	Air Interface
ANDSF	Access Network Discovery and Selection Function
API	Application Programming Interface
ARPU	Average Revenue Per User
AS	Access Stratum
ASA	Authorised Spectrum Access
BBU	Baseband Unit
BM-SC	Broadcast Multicast Service Centre
BNetzA	Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway in Germany
BS	Base Station
BSS	Base Station System
C3	Central Controller and Coordinator
CA	Carrier Aggregation
CAL	CHARISMA Aggregation Level
CAPEX	Capital Expenditure
CBRS	Citizens Broadband Radio Service
CC	Component Carrier
CDN	Content Delivery Network
CEPT	European Conference of Postal and Telecommunications Administrations
CESC	Cloud-Enabled Small Cell
CESCM	Cloud Enabled Small Cell Manager
CMOS	Complementary Metal Oxide Semiconductor
CN	Core Network
CoMP	Coordinated Multipoint
CP	Control Plane
CPE	Customer Premises Equipment
C-RAN	Cloud Radio Access Network
CriC	Critical Communications

D2D	Device-to-Device
DC	Data Centre
DAS	Distributed Antenna System
DMO	Direct Mode Operation
DPI	Deep Packet Inspection
D-RAN	Distributed Radio Access Network
ECC	Edge Cloud Computing
eICIC	enhanced Inter-Cell Interference Coordination
eNB	Evolved Node B
eNodeB	Evolved Node B
eMBB	enhanced Mobile Broadband
eMBMS	enhanced Multimedia Broadcast Multicast Service
EMS	Element Management System
EPC	Evolved Packet Core
E-UTRAN	Universal Terrestrial Radio Access Network
eV2X	enhanced Vehicle-to-X Communications
EPC	Evolved Packet Core
ESO	Essential Service Operator
ETP	European Technology Platform
ETSI	European Telecommunications Standards Institute
EU	European Union
FDD	Frequency-Division Duplex
FICORA	Finnish Communications Regulatory Authority
ForCES	Forwarding and Control Element Separation
GPP	General Purpose Processors
GSM	Global System for Mobile Communications
GW	GateWay
GWCN	Gateway Core Network
H2H	Human-to-Human
HAN	Heterogeneous Access Network
HMN	Heterogeneous Mobile Network
HSS	Home Subscriber Server
ICIC	Inter-Cell Interference Coordination
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMT	International Mobile Telecommunications
IMU	Intelligent Management Unit

IoT	Internet of Things
IP	Internet Protocol
IRTF	Internet Research Task Force
ISI	Integral SatCom Initiative
ITRS	International Technology Roadmap for Semiconductors
ITU	International Telecommunication Union
ITU-R	ITU – Radio Communication Sector
IWF	Interworking Function
KPI	Key Performance Indicator
L1	Layer 1
L2	Layer 2
LSA	Licensed Shared Access
LSN	Last Sequence Number
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution Advanced
M2M	Machine-to-Machine
MAC	Media Access Control
MANO	MANagement and Orchestration
MBMS	Multimedia Broadcast Multicast Service
MCS	Modulation and Coding Schemes
MEC	Mobile Edge Computing
MIMO	Multiple-Input Multiple-Output
MIoT	Massive Internet of Things
MMC	Massive Machine Communication
MME	Mobility Management Entity
MN	Mobile Network
MOCN	Multi-Operator Core Network
MTC	Machine-Type Communication
MTD	Machine-Type Device
NEO	Network Operation
mmWave	Millimeter Wave
MNO	Mobile Network Operator
MOCN	Multi-Operator Core Network
MoI	Ministry of Interior
MVNO	Mobile Virtual Network Operator
Naas	Network as a service
NAS	Non-Access Stratum
NEO	Network Operation

NFV	Network Function Virtualisation
NFVO	Network Functions Virtualisation Orchestrator
NFVI	Network Function Virtualisation Infrastructure
NGMN	Next Generation Mobile Networks Alliance
NRA	National Regulation Agency
NVS	Network Virtualisation Substrate
Ofcom	Office of Communications
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
ONF	Open Networking Foundation
OPEX	Operational Expenditure
OSI	Open Systems Interconnection
OSS	Operations support systems
PCC	Primary Component Carrier
P-GW	PDN gateway
PHY	Physical Layer
PLMN	Public Land Mobile Network
PMR	Private (or Professional) Mobile Radio
PNF	Physical Network Function
PON	Passive Optical Network
PoP	Point of Presence
PPDR	Public Protection and Disaster Relief
PRACH	Physical Random Access Channel
ProSe	Proximity Services
PTT	Push-To-Talk
PTS	Swedish Post and Telecom Agency
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RANaaS	Radio Access Network as a Service
RAT	Radio Access Technology
RAU	Radio Access Unit
RB	Resource Block
RF	Radio Frequency
RN	Relay Node
RT	Radio Transceiver
R-TP	Radio Transmission Point
RRH	Remote Radio Head

RRM	Radio Resource Management
RRU	Radio Remote Unit
RTC	Real-Time Controller
SA	Service and System Aspects
SAS	Spectrum Access System
SC	Small Cell
SCC	Secondary Component Carrier
SCS	Service Capability Server
SDK	Software Development Kit
SDN	Software Defined Network
SD-RAN	Software-Defined Radio Access Network
S-GW	Serving gateway
SID	Service Integration Driver
SLA	Service Level Agreement
SLP	SUPL Location Platform
SME	Small and Medium Enterprises
SON	Self Organizing Network
SONAC	Service-Oriented Virtual Network Auto-Creation
SUPL	Secure User Plane Location
TCO	Total Cost of Ownership
TDD	Time-Division Duplex
TEDS	TETRA Enhanced Data Service
TETRA	Terrestrial Trunked Radio
TN	Transport Node
TR	Technical Report
TDM	Time Division Multiplexing
TR	Technical Report
UAV	Unmanned Aerial Vehicles
UE	User Equipment
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunications System
UP	User Plane
V2X	Vehicle-to-X Communications
VIM	Virtualised Infrastructure Manager
VM	Virtual Machine
VNF	Virtual Network Function
VNFD	VNF Descriptors
VNFM	VNF Manager

vBBU	virtual Baseband Unit
vBSC	virtual Base Station Controller
VIM	Virtualised Infrastructure Manager
VNF	Virtual Network Function
VNO	Virtual Network Operator
VPL	Vehicular Penetration Loss
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WP	Work Package, Working Party
WRC	World Radio communication Conference
XCI	Xhaul Control Infrastructure,
XFE	Xhaul Packet Forwarding Element
XPU	Xhaul Processing Unit

List of figures

List of tables

1. Introduction

The COHERENT project focuses on the control and coordination of radio access functions in the 5G Radio Access Network (RAN). 5G scenarios and requirements bring particular challenges to the RAN design. The ultra-high data rate for enhanced mobile broadband services requires a new radio interface, wide spectrum bands, flexible carrier aggregation, high spectrum efficiency, and advanced interference management. The ultra-high capacity per geographical region demands ultra-dense network deployment, flexible spectrum access, and novel mobility management. Ultra-reliable and low latency communications require the delay as low as 1 millisecond at the RAN side. To accommodate a wide spectrum of services from different vertical sectors, as well as new features like network slicing across the RAN, RAN functions need to be extremely reliable, flexible and programmable. 5G networks will allow the flexible split and combination of logical RAN functions to physical network entities, so as to bring new features of Network Function Virtualisation (NFV), Software Defined Networking (SDN), and Mobile Edge Computing (MEC) to the RAN. The control and coordination of multiple network entities will be a common feature in the 5G RAN. Clearly, the efficient coordination of these network functions would be critical for end to end service provisioning in 5G networks.

The design of system architecture at RAN needs the following considerations. Firstly, the separation of the control [1]and data plane[2] needs to be carefully addressed. The separated control plane will enable the logical centralised control over heterogeneous RAN. It will be a necessary approach to provide a unified control framework for 5G networks. On the other hand, in RAN some radio control functions, e.g., the scheduling, are latency-sensitive, and thus have to be resided close to the data plane. The trade-off has to be made for real-time control functions and centralised control functions.

Secondly, the flexible function split needs to be supported at the data plane. It will allow different implementations of the radio transmission architecture. For instance, if some network function at the data plane are centralised in the cloud, the multiplexing gain among multiple base stations (especially dense small cell deployment) could be utilised in an optimal way by jointly processing data plane functions. The flexible function split together with NFV will enable a cost-efficient and service-agile 5G RAN.

Thirdly, the 5G RAN needs to have the native support for flexible spectrum management. In additional to the licensed spectrum below and above 6GHz, the 5G RAN will take advantage of different spectrum sharing techniques, e.g., Licensed Shared Access (LSA), or flexible duplex, for improved network capacity and service provisioning. The control plane of the 5G RAN should be designed to natively support for flexible spectrum.

Finally, programmability will be a necessary feature in the control and data plane of the 5G RAN. Programmability implies that the control and data plane functions can be tailored according to the network and service needs. For network slicing in the RAN, programmability will allow different control implementations per slice, which is extremely important to service isolation.

The key concepts of the COHERENT Architecture are summarised here:

- **Control Separation:** By applying SDN in COHERENT architecture, a centralised solution is adopted, which could achieve the global optimisation. However, the biggest challenge in creating such a software defined radio access network is the inherent delay between any control entity and the individual radio elements. By separating the control functionalities of centralised control and real-time control, COHERENT architecture could adjust to rapidly varying wireless networks while getting the benefit of performance gain introduced by centralised control.

[1] We use terms *control* and *coordination* interchangeably in this deliverable.
[2] We use terms *data plane* and *user plane* interchangeably in this deliverable.

- **Network Abstractions:** A coherent representation of the network state and infrastructure resources is the key to drive programmability in RAN control and coordination for 5G systems. In COHERENT, abstractions encompass representations and models of time-frequency resources, spatial capabilities (i.e. number of transmit and receive antennas), as well as throughput per network slice or per allocated resources. In the COHERENT architecture with control separation, the controller instances perform real-time control and centralised control with the provision of local and logically centralised view of the infrastructure, respectively.

- **Network Slicing**: SDN and NFV promise to reduce the cost to deploy and operate large networks by migrating Network Functions (NFs) from dedicated hardware appliances to software instances running on general purpose virtualised network and compute infrastructures. This, in time, shall improve the flexibility and scalability of mobile networks in that the deployment of new applications and services will be quicker (software vs. hardware development lifecycles) and different NFs can share the same resources paving the way to further economies of scale. This progressive process of network softwarisation is set to play a pivotal role in 5G. In this context the Network-as-a-Service (NaaS) business model shall allow operators to tap into new revenue streams by further abstracting the physical network into service specific slices possibly operated by different mobile virtual network operators (MVNOs). The network slices envisioned in COHERENT, span the whole protocol stack from the underlying (virtualised) hardware resources up to network services and applications running on top of them. This approach is aligned with the industry and telecom perspective, towards 5G [1][2], in order to meet the demands of extremely diverse Use Cases.

1.1 Structure of the document

The remaining document is organised as follows:

- **Section 2:** COHERENT aims to design a programmable 5G system based on emerging technology enablers (e.g., SDN and NFV) and mobile architectures. In Section 2, we survey the main outcomes from the academic community and industry recently.

- **Section 3:** For designing Programmable 5G Architecture, a set of fundamental requirements for designing programmable 5G architecture is introduced in Section 3. Furthermore, architecture requirements for network abstraction are also presented.

- **Section 4:** In Section 4, the key concepts that lie at the foundation of the COHERENT Architecture are introduced, namely i) **control separation** (logically centralised control and real-time control), ii) **network abstractions** and iii) **network slicing**.

- **Section 5:** In Section 5, the COHERENT architecture is proposed by applied three key concepts mentioned in Section 4. Furthermore, the control and coordination plane and the user plane are described in details. Finally, the functional blocks responsible for processing network information are proposed.

- **Section 6:** In Section 6 we show how the COHERENT Architecture can support a number of different use cases. The use cases discussed here are general and include multiple specific use cases described in D2.1, where a detailed description of COHERENT's reference scenarios has been presented. The use cases presented in this section include multi-tenancy, Device-to-Device (D2D) communications, broadcast operation, UE-relaying operation for public safety service, multi-connectivity and mobility management. Note that the use cases of spectrum management using COHERENT architecture are introduced in D4.1 of COHERENT.

- **Section 7:** While the generic COHERENT architecture introduced in Section5 covers basic concepts suitable for different RATs, including, but not limiting to, LTE and WiFi systems, it needs the necessary adaptation when applying to particular radio access technologies. Section 7 describes *3GPP oriented Base Station System (BSS) architecture* for future 5G 3GPP systems, which is the contribution of COHERENT to the 3GPP framework. The mapping between generic COHERENT architecture and 3GPP oriented BSS architecture is examined in details in this section.

- **Section 8:** The goal of Section 8 is to introduce the abstractions and interfaces for heterogeneous mobile networks programmability. Firstly, the concepts behind domain specific languages (DSLs) and the role DSLs play in COHERENT abstractions and interfaces are introduced. Moreover, the COHERENT semantic model is presented. Finally, we describe the northbound interface and the primitives which allow the flexibility on manipulating the semantic models in order to configure the state of the networks for the network programmer.

2. State of the Art

COHERENT aims to design a programmable 5G system based on novel concepts. In order to cope with the diverse 5G requirements summarized in D2.1 of COHERENT, COHERENT 5G systems will heavily rely on advanced resource abstraction, virtualisation and programmability combined with dynamic network service orchestration borrowing elements from i) Cellular network architecture, ii) Network Function Virtualisation (NFV) and iii) the Software Defined Networking (SDN).

In this section, the most relevant and recent collaborative projects which have invested significant efforts to enable some of the key enablers mentioned previously are presented in Section 2.1 and Section 2.2, respectively. More specifically, the collaborative projects outside the 5G-PPP initiative are described in Section 2.1 while the collaborative projects within the 5G-PPP initiative are introduced in Section 2.2. Furthermore, the 5G enabling technologies (e.g., SDN and NFV) and mobile architectures provided by standards or similar organisations are reviewed in Section 2.3. Finally, the academic literature review regarding 5G architecture is provided in Section 2.4.

2.1 Collaborative Projects outside the 5G-PPP initiative

In this section we shall survey the most relevant and recent collaborative projects in the software defined networking / programmable networks domain as well as spectrum management domain with a particular focus on wireless and mobile networks projects. The following collaborative projects are outside the 5G-PPP initiative.

2.1.1 CROWD

The CROWD (Connectivity management for eneRgy Optimised Wireless Dense networks) project [3] aims at defining and implementing a flexible architecture capable of accommodating the requirements of the extremely dense scenario, which will be common in future access networks. Software defined networking has been used as technical enabler and the necessity to split the control plane across local and regional controllers has been recognised.

CROWD pursues four key goals: i) bringing density-proportional capacity where it is needed, ii) optimising MAC mechanisms operating in very dense deployments by explicitly accounting for density as a resource rather than as an impediment, iii) enabling traffic-proportional energy consumption, and iv) guaranteeing mobile user’s quality of experience by designing smarter connectivity management solutions.

Among the projects surveyed in this section, CROWD is the closest in principle to COHERENT. However, unlike CROWD, COHERENT focuses on defining the common control and coordination abstractions for wireless and mobile networks, find the invariants on mobile network control, while isolating policies from mechanism and at putting the latter in the hands of network programmers.

2.1.2 Fed4FIRE

The Fed4FIRE [4] project aims at federating the various facilities for Future Internet Research and Experimentation (FIRE) that have been built so far under a single umbrella. Through the federation of these infrastructures, innovative experiments become possible that break the boundaries of these different domains (wired, wireless, IT). Such a goal is achieved by empowering the experimenters with common tools for the federation of the various experimental facilities, allowing them to focus on their core testbed activities.

The roster of experimental facilities and testbed made available by Fed4FIRE is significant and encompasses different technological networking domains (optical, WiFi, mobile networks, and sensor networks). Moreover, specialised facilities on cloud computing and high performance computing can also be found.

The Fed4FIRE is by objective radically different from COHERENT, in fact, while the former (Fed4FIRE) focuses on federation of testbed as well as setup of new facilities, the latter (COHERENT) is focusing on fundamental Research and Innovation in the deep programmable networks domain. Such activities may eventually result in new experimental facilities making on the COHERENT research outcomes available to a larger audience.

2.1.3 FLEX

FLEX (FIRE LTE testbeds for Open Experimentation) [5] aims to develop an open and operational LTE experimental facility. The goal of FLEX is to provide external experimenters with a facility that can be configured in order to suit their needs and that can run full end-to-end services. Such goals are pursued by leveraging on a combination of truly configurable commercial equipment, truly configurable core network software, full open source components, and on top of those, sophisticated emulation and mobility functionalities. The Flex facility allows researchers from academia and industry to test services and applications over real LTE infrastructure, or experiment with alternative algorithms and architectures of the core and access network.

Regarding the relationship between Flex and COHERENT, the consideration made for Fed4FIRE can be applied here as well.

2.1.4 VITAL

The VITAL (VIrtualised hybrid satellite-TerrestriAl systems for resilient and fLexible future networks) [6] project aims at addressing the combination of Terrestrial and Satellite networks by bringing Network Functions Virtualisation (NFV) into the satellite domain and by enabling Software-Defined-Networking (SDN)-based, federated resources management in hybrid Satellite Communication (SatCom)-terrestrial networks. The rationale behind such an approach is to pave way for a unified control plane that can allow operators to efficiently manage and optimise the operation of hybrid SatCom-Terrestrial networks. Likewise, enabling NFV into SatCom domain can provide the operators with appropriate tools and interfaces in order to establish end-to-end fully operable virtualised satellite networks to be offered to third-party operators/service providers.

VITAL addresses the development of a hybrid architectural framework, the required mechanisms to enable virtualisation of SatCom network components, including performance optimisation and implementation of a number of virtualised functions, and the design of an SDN-enabled, federated resources management framework, embedding strategies and algorithmic solutions to provide end-to-end communication services. Proof of concept validation of VITAL solutions and enabling technologies through a combination of real prototypes and emulators are also planned.

The VITAL project is conceptually similar to COHERENT in the way that both projects address the domain of deep programmable and softwarised networks. However, while VITAL targets mostly satellite networks, COHERENT focuses on terrestrial wireless and mobile networks. Moreover, while VITAL embeds also an NFV perspective, in the case of COHERENT we can find a particular focus on control and coordination aspects of a heterogeneous radio access network, including, but not limited to, spectrum management, RAN sharing, and traffic engineering.

2.1.5 ADEL

The FP7-ICT-2013-11 ADEL (Advanced Dynamic spectrum 5G mobile networks Employing Licensed shared access) [7] aims at providing fundamental contribution on flexible spectrum usage, which is a promising enabler of spectral efficiency for next generation wireless broadband networks.

With the emergence of heterogeneous and small cell networks, the original “licensed vs. unlicensed” spectrum usage model has recently given way to the “Licensed Shared Access (LSA)” (a.k.a. “Authorised Shared Access (ASA)”) paradigm wherein incumbent operators may allow other ones to share their spectrum at specific times and places, according to an agreed set of rules. However, the

LSA approach also sets specific technology challenges. The ADEL project pursues the following goals: i) the dynamic and optimised allocation of the spectral and power resources at a short time scale (on the order of seconds to even milliseconds); ii) the guarantee of Quality of Service to the users of all participating spectrum-sharing networks; and iii) the minimisation of the overall energy expenditure of LSA networks. To reach those goals the ADEL project aims to advance the state-of-the-art by developing innovative collaborative sensing techniques, resource allocation and flexible spectrum access protocols in the context of LSA-oriented, self-organised, hierarchical wireless access networks, thus paving the way for ever increased spectral efficiency, improved energy efficiency, and reduced deployment / operation cost in 5G mobile broadband networks.

In COHERENT, spectrum management and in particular the LSA paradigm will be integrated in the larger COHERENT control framework and in the COHERENT architecture, which will provide new tools to solve spectrum management issues.

2.2 Collaborative Projects within the 5G-PPP Initiative

According to the view from the 5G PPP Architecture white paper [8], 5G networks are programmable with respect to service, application, time, location or/and context as well. Such programmability could be applied in different network segments, including radio access networks, transport networks, core networks, etc., as it is depicted in Figure 2-1.

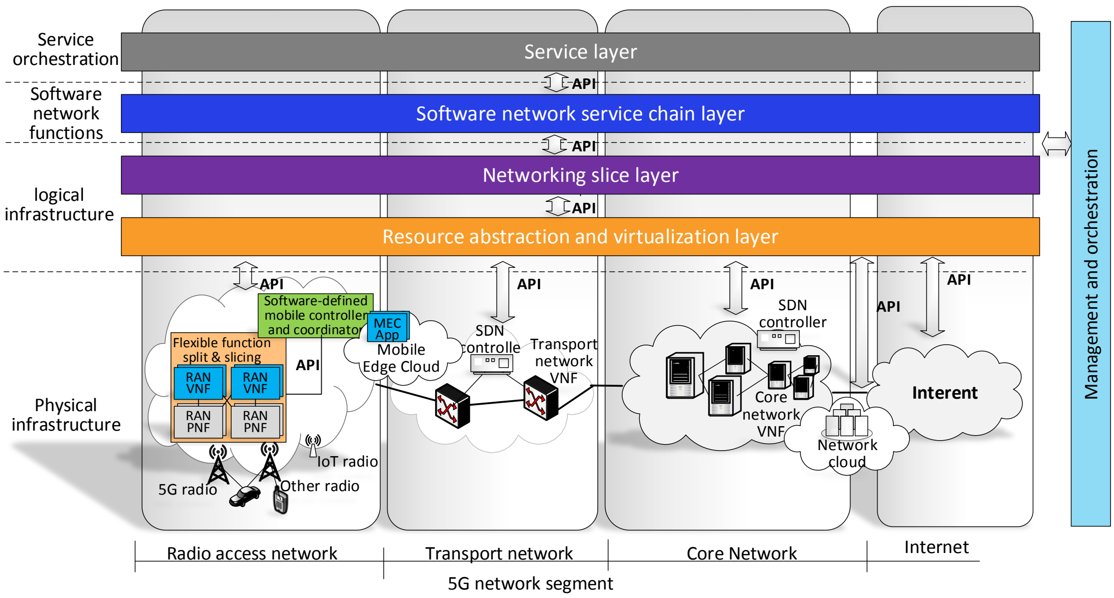

Figure 2-1 5G network segments and uniformed service platform [8]

Apart from the technological challenges, 5G architecture enables new business opportunities meeting the requirements of a wide range of use cases. Such use cases may include cost-efficient network slicing implementation, addressing both end user and operational services, softwarisation support, integration of heterogeneous access technologies, etc. A wide range of potential use cases were presented in D2.1.

Towards development of 5G research, there is already a lot of work delivered (or planned to be delivered in the following period) in the context of projects under the 5G-PPP initiative. COHERENT will leverage and exploit their knowledge generated regarding a number of issues, such as definition of use cases, user-driven requirements, network-driven requirements and technologies involved, as well as other layered architectures proposed towards 5G. We note that some COHERENT activities are orthogonal to the existing work delivered by other 5G-PPP projects and some COHERENT partners participate in some of 5G-PPP projects. In the following we provide a description of each project and possible interaction with COHERENT.

2.2.1 5G-Xhaul

5G-XHaul (Dynamically Reconfigurable Optical-Wireless Backhaul/Fronthaul with Cognitive Control Plane for Small Cells and Cloud-RANs) [9] aims at delivering a converged optical and wireless network solution able to efficiently address the demand for broadband connectivity in both dense RAN and Cloud-RAN architectures. 5G-XHaul solution intends to allow the dynamic allocation of network resources to predicted and actual hotspots and the flexible connection of small cells to the core network. Figure 2-2 depicts a conceptual view of 5G-XHaul network deployment. We observe that 5G-XHaul focuses on improving not only on current fronthaul, but also backhaul technologies.

As part of the 5G-Xhaul project, it is intended to be implemented dynamically programmable, high capacity, low latency, point-to-multipoint mm-Wave transceivers, cooperating with sub-6-GHz systems; a time shared optical network offering elastic and fine granular bandwidth allocation, cooperating with advanced Passive Optical Networks (PONs); as well as a software-defined cognitive control plane for forecasting traffic demand in time and space, and the ability to reconfigure network components.

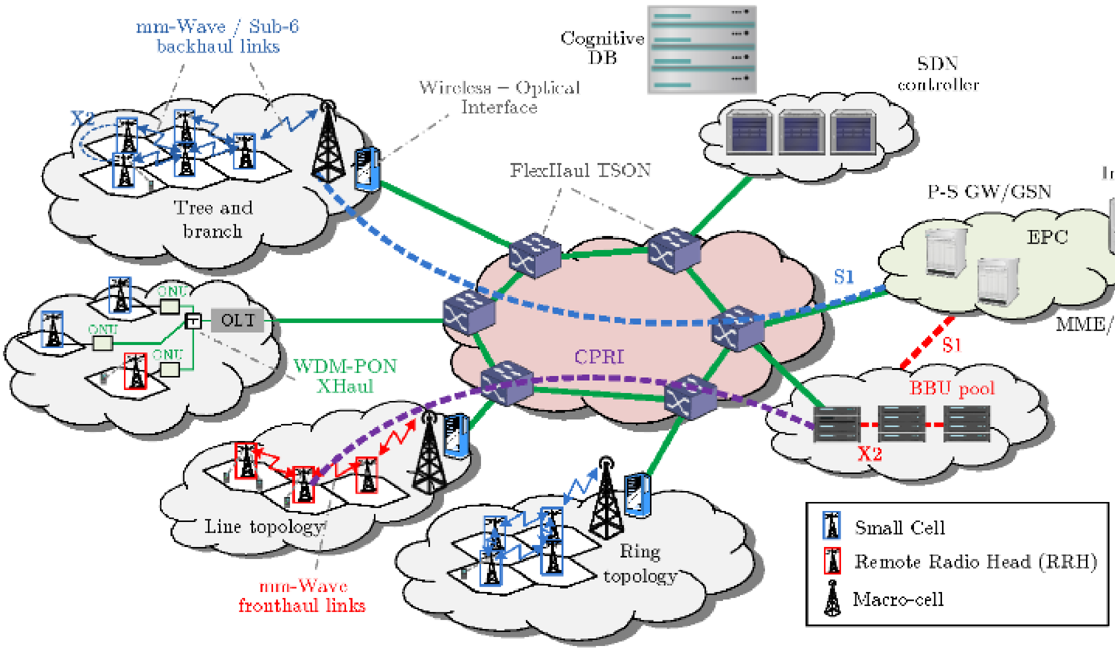

Figure 2-2 5G-XHaul network deployment [10]

2.2.2 5G-Crosshaul

In the same context as 5G-Xhaul, the 5G-Crosshaul project [11] aims at developing a 5G integrated backhaul and fronthaul transport network enabling a flexible and software-defined reconfiguration of all networking elements in a multi-tenant and service-oriented unified management environment.

The Xhaul transport network envisioned will consist of high-capacity switches and heterogeneous transmission links (e.g., fibre or wireless optics, high-capacity copper, mmWave) interconnecting Remote Radio Heads (RRHs), 5G Point of Access (e.g., macro and small cells), cloud-processing units (mini data centres), and points-of-presence of the core networks of one or multiple service providers. This transport network will flexibly interconnect distributed 5G radio access and core network functions, hosted on in-network cloud nodes, through the implementation of: i) a control infrastructure using a unified, abstract network model for control plane integration (Xhaul Control Infrastructure, XCI); ii) a unified data plane encompassing innovative high-capacity transmission technologies and novel deterministic-latency switch architectures (Xhaul Packet Forwarding Element, XFE).

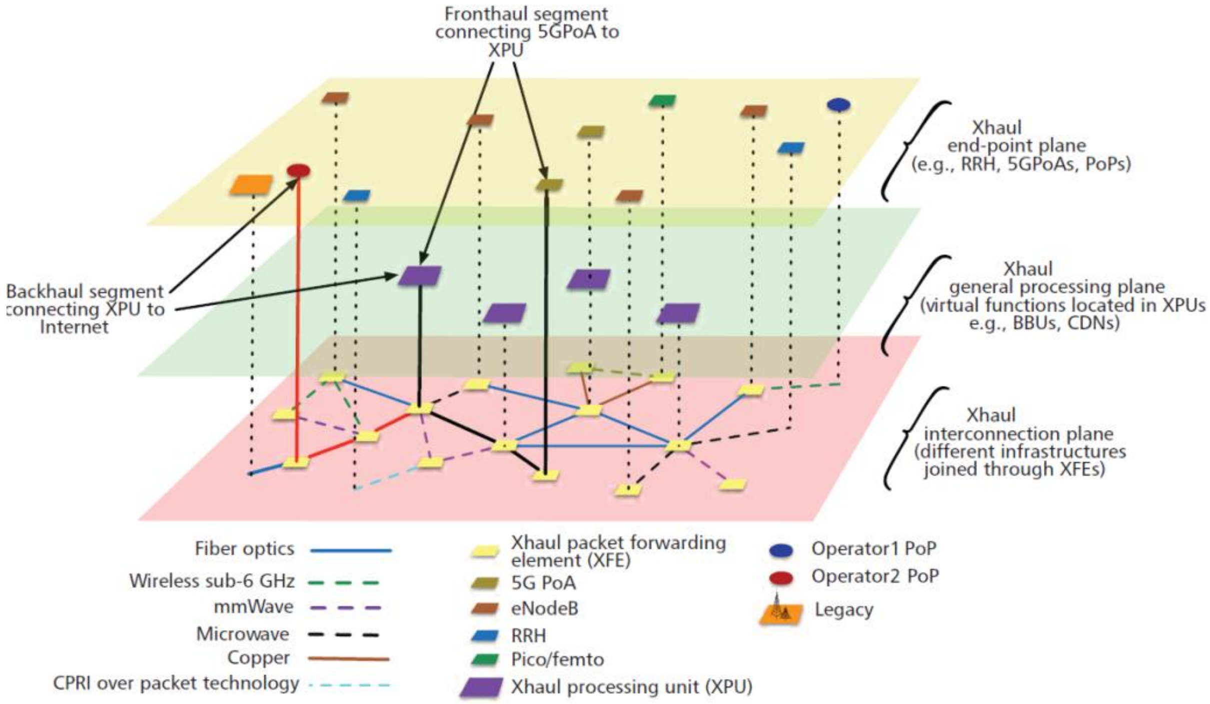

Figure 2-3 5G-Crosshaul network deployment [12]

Demonstration and validation of the Xhaul technology components developed will be integrated into a software-defined flexible and reconfigurable 5G Test-bed in Berlin. Mobility-related Xhaul experiments will be performed using Taiwan's high-speed trains. Xhaul KPI targets evaluated will include among others a 20% network capacity increase, latencies <1 ms and 30% Total Cost of Ownership (TCO) reduction.

2.2.3 CHARISMA

CHARISMA (Converged Heterogeneous Advanced 5G CloudRAN Architecture for Intelligent and Secure Media Access) [13] proposes an intelligent hierarchical routing and paravirtualised architecture that unites:

- Devolved offload with shortest path nearest to end-users and
- An end-to-end security service chain via virtualised open access physical layer security.

Within this project, a flexible and programmable control and management plane is enabled by the Open network-as-a-Service (OpenNaaS) platform, in order to offer cost savings enabled by Software-Defined Networking (SDN). Apart from focusing on the provision of a programmable control and management plane, security aspect is also of great importance.

The CHARISMA architecture intends to provide a cloud infrastructure platform with increased spectral/energy efficiency and enhanced performance as well as security as required for future converged wireless/wireline advanced 5G networking. The next figure depicts a high-level view of CHARISMA control, management, and orchestration plane. In this architectural figure, we see the following component groups; Virtualised Infrastructure (VI), Virtualised Network Functions (VNFs), Management and Orchestration (MANO), and Operations and Business Support Systems.

VI virtualises the hardware resources which are exposed for consumption by VNFs. MANO is responsible for implementing network services, VNF lifecycle management, resource management and control, slicing, etc. MANO communicates with the OSS/BSS of VNOs in order to report status or potentially receive requirements.

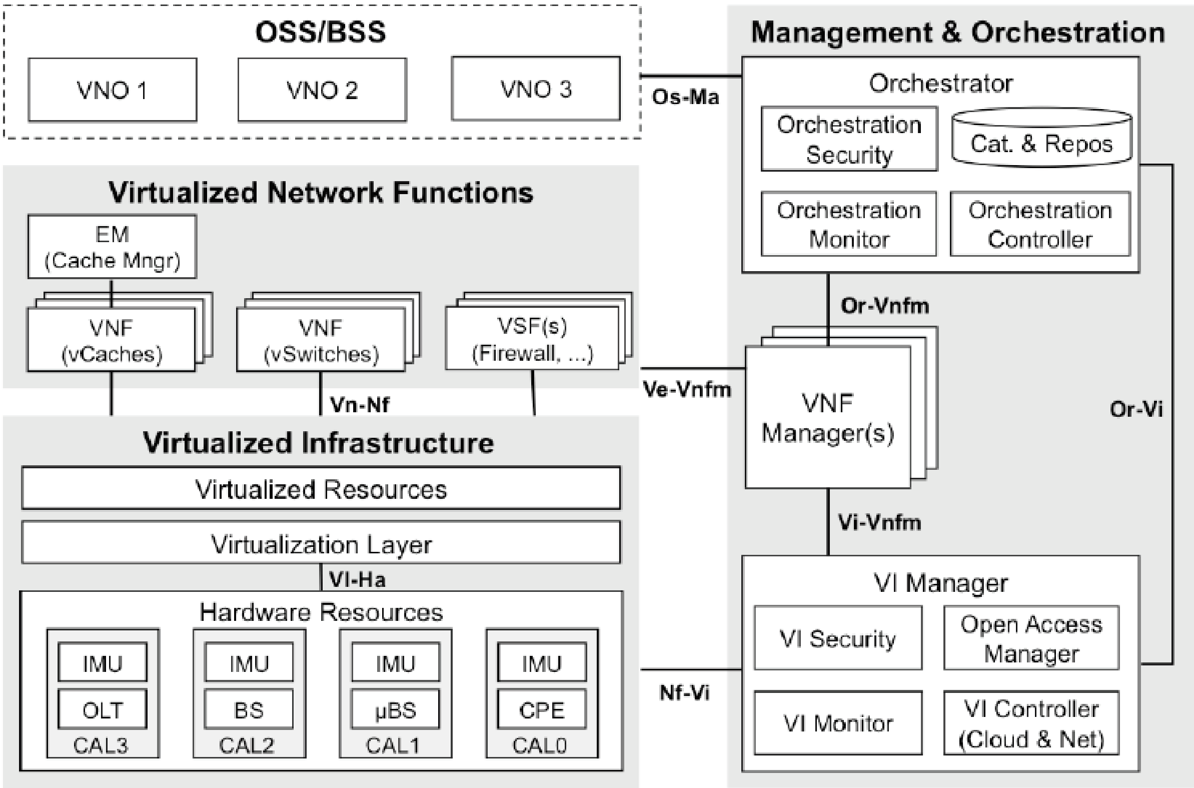

Figure 2-4 Conceptual CHARISMA control, management, and orchestration plane [14]

2.2.4 SELFNET

The SELFNET project [15] will design and implement an autonomic network management framework to achieve self-organizing capabilities in managing network infrastructures by automatically detecting and mitigating a range of common network problems that are currently still being manually addressed by network operators, thereby significantly reducing operational costs and improving user experience. SELFNET explores a smart integration of state-of-the-art technologies in Software-Defined Networking (SDN), Network Function Virtualisation (NFV), Self-Organizing Networks (SON), Cloud computing, Artificial intelligence, Quality of Experience (QoE) and Next-generation networking to provide a novel intelligent network management framework. Such a framework is capable of assisting network operators in key management tasks:

- Automated network monitoring by the automatic deployment of NFV applications to facilitate system-wide awareness of Health of Network metrics in order to have more direct and precise knowledge about the real status of the network.
- Autonomic network maintenance by defining high-level tactical measures and enabling autonomic corrective and preventive actions against existing or potential network problems.

SELFNET is driven by use cases designed to address major network management problems including self-protection capabilities against distributed cyber-attacks, self-healing capabilities against network failures, and self-optimisation to dynamically improve the performance of the network and the QoE of the users.

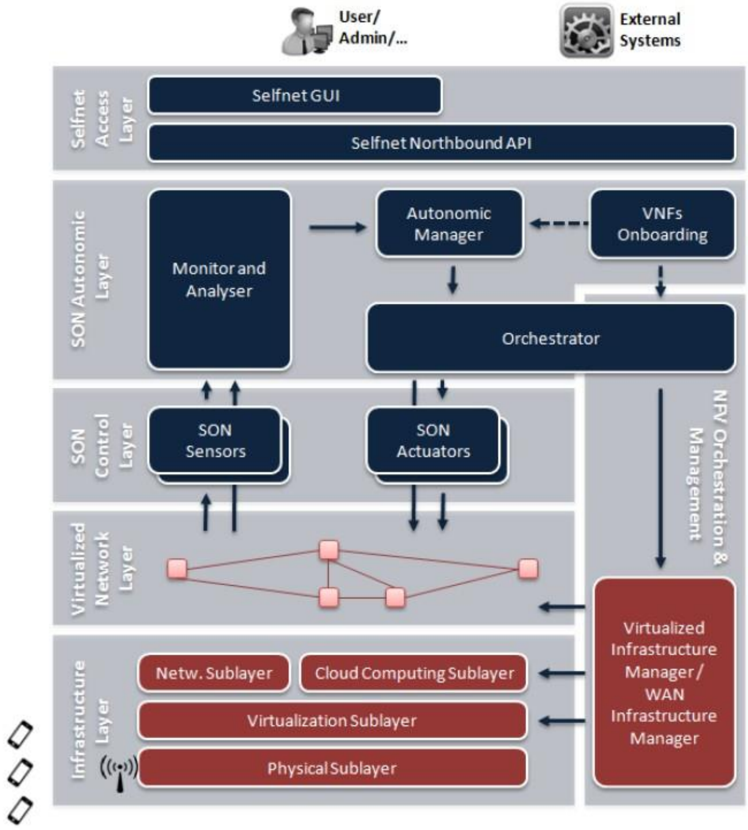

Figure 2-5 SELFNET architecture overview[16]

SELFNET is designed within this economic and business context to substantially reduce operational costs of network operators by automating a significant number of current labour-intensive network management tasks.

As COHERENT focus is not on the NFV orchestration layer as defined within the scope of the SELFNET, we are highly interested on exposing our *centralised control and coordination* mechanism, together with novel *network abstractions* for the RAN design, as a control realisation in the rest of the SELFNET solution. To this end liaison activities are expected with SELFNET project.

2.2.5 5G-NORMA

5G NORMA [17] aims to develop a novel mobile network architecture that provides the necessary adaptability in a resource efficient way, which is able to handle fluctuations in traffic demand resulting from heterogeneous and dynamically changing service portfolios and to changing local context. The developed "multi-service and context-aware adaptation of network functions" will allow for a resource-efficient support of these varying scenarios and help to increase energy-efficiency by always selecting the most energy efficient option.

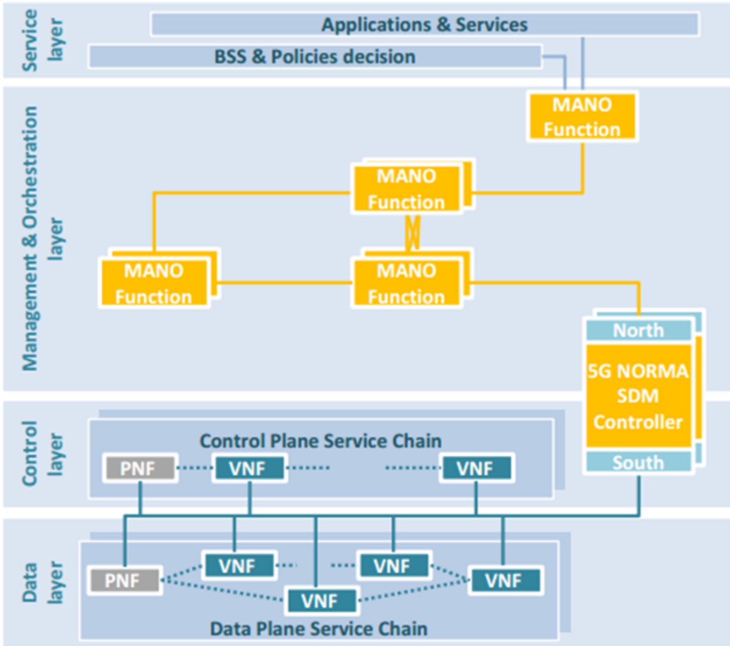

Figure 2-6 5G-NORMA Functional view [18]

The "mobile network multi-tenancy" approach to be developed by 5G NORMA will leverage the adaptability and efficiency of network functions and enable an inherent and dynamic sharing and distribution of network resources between operators. This will allow operators to increase their revenue through the new services, while leveraging the efficiency of the architecture to do so in a cost-effective way.

Similarly, to the case of SELFENT project we are highly interested on exposing our *centralised control and coordination* mechanism, together with novel *network abstractions* for the RAN design, as a control realisation in the rest of the NFV solution proposed by 5G-NORMA.

2.2.6 SESAME

SESAME (Small cEllS coordinAtion for Multi-tenancy and Edge services) [19] targets innovations within the context of 5G in the following issues (See also Figure 2-7):

- To place network intelligence and applications in the network edge through Network Functions Virtualisation (NFV) and Edge Cloud Computing (ECC).
- To evolve the small cell concept, by delivering its full potential in the challenging high dense 5G scenarios.
- To consolidate multi-tenancy in communications infrastructures, in order to allow network operators and service providers to collaborate in terms of sharing infrastructure, resources, as well as edge computing capabilities.

Furthermore, SESAME opens the door to 5G system design solutions by investigating, e.g., which core network functions are to be virtualised at the edge of the network.

The proposed architecture of SESAME project allows multiple network operators (tenants) to provide services to their users through a set of Cloud-Enabled Small Cell (CESC) deployed, owned and managed by a third party. This party integrates a virtualised execution platform for deploying Virtual Network Functions (VNFs), supporting management and executing novel services inside the access network infrastructure. The optimal management of a CESC deployment is a key challenge of SESAME, for which new orchestration, NFV management, virtualisation and radio access management techniques are intended to be developed.

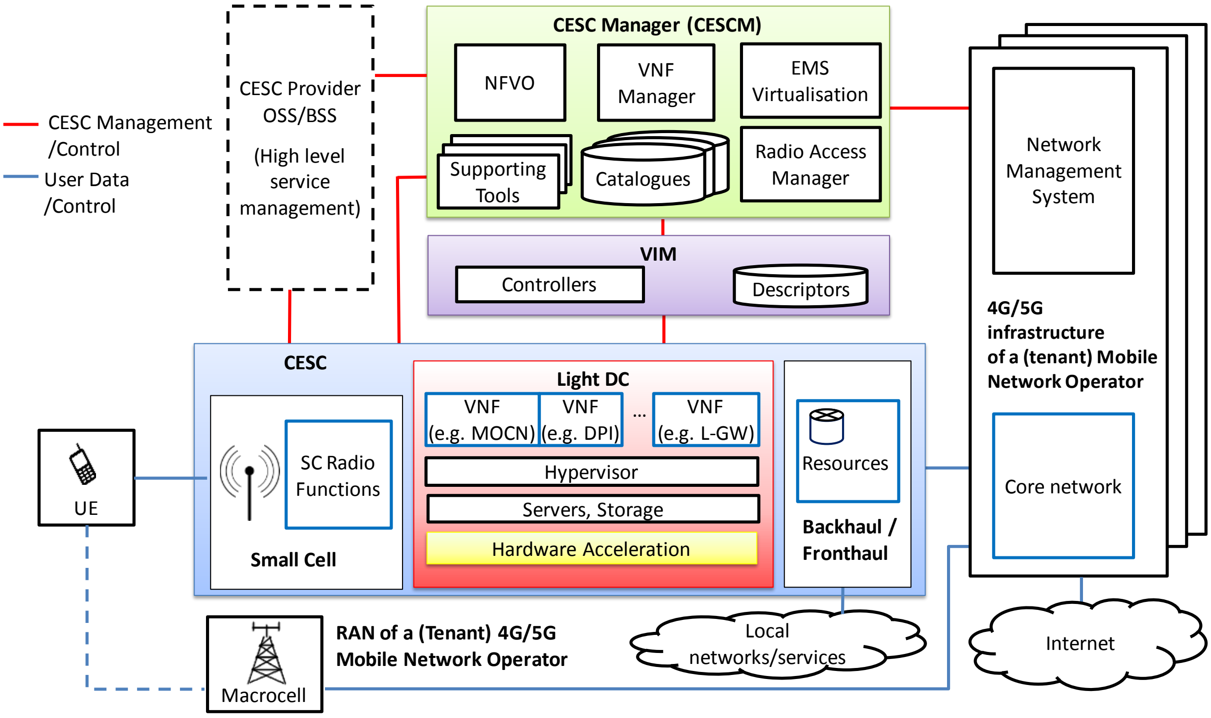

Figure 2-7 SESAME proposed system architecture [20]

2.2.7 5G-EXCHANGE (5GEX)

5GEX project [21] focuses on Multi-domain Orchestration for Software Defined Infrastructures. Particularly, the goal of the 5GEx project is to enable cross-domain orchestration of services over multiple administrations or over multi-domain single administrations. This will result in allowing end-to-end network and service elements to mix in multi-vendor, heterogeneous technology and resource environments.

One important aspect of 5GEX is that also focuses on economic and business challenges that arise with 5G technology. One such challenge is that current market fragmentation, as it is so far, makes difficult to create infrastructure services spanning multiple countries, such as virtual connectivity or compute resources, as no single operator has a footprint everywhere.

5GEx aims to enable collaboration between operators, regarding 5G infrastructure services in terms of introducing unification via NFV/SDN compatible multi-domain orchestration. As part of 5GEX's implementation part, an open platform enabling cross-domain orchestration of services over these multiple domains will be developed. Additionally, a Sandbox Network enabling experimentation and validation of the devised architecture, mechanisms, and business models will be implemented. The overall goal of the project is to actively promote the adoption of 5GEx's open solutions.

Such 5G infrastructure services will provide a crucial role in making 5G happen as they provide the foundation of all cloud and networking services. 5GEx aims to enable, through operator collaboration, a unified European infrastructure service market integrating multiple operators and technologies, where service provisioning is fast and automated and which results in stronger economy via economies of scale [22].

The Multi-domain logical interworking architecture of 5GEX is depicted in Figure 2-8.

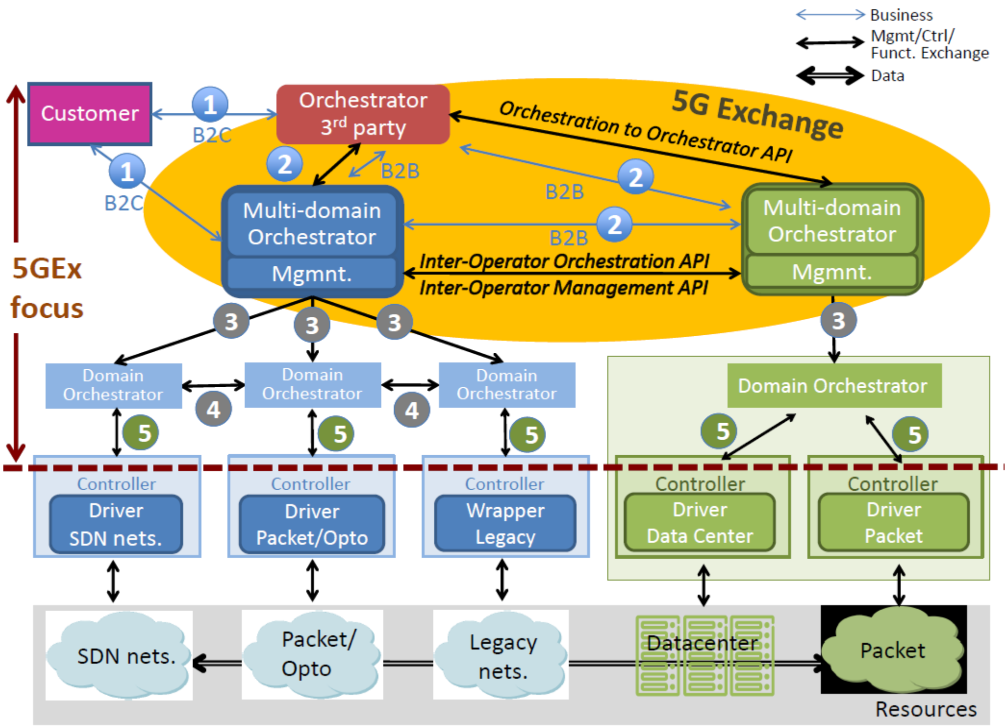

Figure 2-8 Multi-domain logical interworking architecture [23]

2.2.8 FANTASTIC-5G

FANTASTIC-5G (Flexible Air iNTerfAce for Scalable service delivery wiThin wIreless Communication networks of the 5th Generation) [24] focuses on developing a new multi-service Air Interface (AI) for below 6 GHz through a modular design.

FANTASTIC-5G intends to introduce new waveforms, new designs and control signalling in Physical Layer (PHY) and Medium Access Control (MAC), via a unified frame structure in up- and downlink handling a variety of traffic classes. FANTASTIC-5G also focuses on the provision of solutions with varying levels of macro-cell assistance for small cell deployment exploring network, as well as on supporting massive uncoordinated access of Massive Machine Communication (MMC) devices exploiting the sporadic traffic patterns.

Particularly, the research activities of the project focus on service-specific AI components for a set of core services. The concept of Service Integration Drivers (SIDs) is also adopted. The main role of a core service SIDs is collecting different components optimized for each service and integrating them into a service-specific air interface solution, including waveform, frame, control signalling, coding, modulation, procedures, retransmission schemes, Multiple-Input Multiple-Output (MIMO) etc. The term “Overall SID” is used in order to ensure that the co-existence of different services is achieved.

The next figure depicts the role of the SIDs in the air interface design.

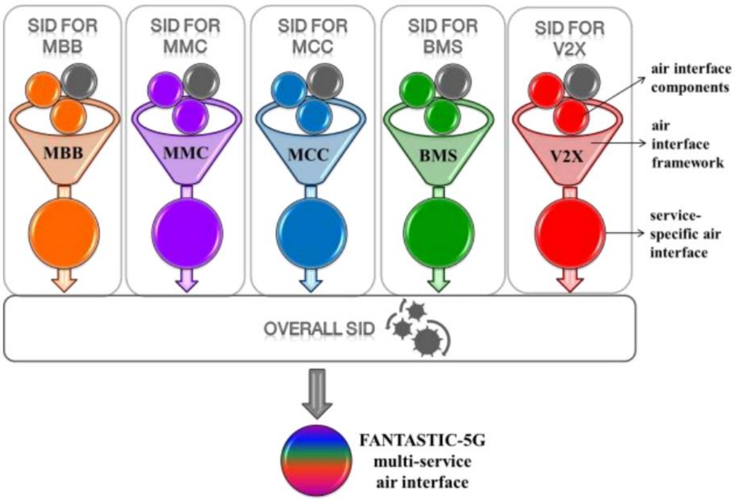

Figure 2-9 Role of Service Integration Drivers in the air interface design [25]

2.2.9 COHERENT Interaction with 5G-PPP

A summary for possible interactions with other 5G-PPP projects is provided in the following table:

Table 2-1 COHERENT interaction with other 5G-PPP projects

5G-PPP Project	Interaction with COHERENT
5G-Crosshaul: It is about a converging 5G fronthaul/backhaul infrastructure based on several data plane technologies and operated through SDN and NFV-based control and orchestration functions to provide multi-tenancy, dynamic virtual service function chains, energy efficient resource allocation and mobility management.	**Collaboration:** Technical areas of NFV orchestration for 5G oriented service chains. **Common Partner Involved:** CREATE-NET
5G-Xhaul: It is about dynamically programmable, point-to-multipoint mm-Wave transceivers, cooperating with sub-6-GHz systems; a TSON offering elastic and fine granular bandwidth allocation, cooperating with advanced passive optical networks; and a SDN cognitive control plane, able to reconfigure network components.	**Collaboration:** Technical areas of backhaul/fronthaul SDN developments and the opening of the RAN network. **Common Partner Involved:** OTE
SELFNET: It proposes an SDN- and NFV-based network management framework for advanced Self-Organizing Network (SON) capabilities in 5G infrastructures. SELFNET framework will be assessed through three use cases characterised by innovative proactive and reactive actions, like self-healing, self-protection and self-optimisation of SDN-based and virtualised 5G networks.	**Collaboration:** Possible cooperation may target automated orchestration functions for optimizing the VNF placement and automatically reacting in case of failures at the infrastructure level. We are highly interested in exposing our *centralised control and coordination* mechanism, together with our novel *network abstractions* for the RAN design, as a control realisation in the rest of the NFV solution proposed. **Common Partner Involved:** None

SESAME: It proposes the Cloud-Enabled Small Cell (CESC) concept, a new multi-operator enabled Small Cell that integrates a virtualised execution platform (i.e., the Light DC) for deploying Virtual Network Functions (VNFs)	**Collaboration:** Interest in the consolidation approach of SESAME and the way multi-tenancy effects are met in communications infrastructures. **Common Partner Involved:** OTE, CREATE-NET
5G-EXCHANGE (5GEX): 5GEx aims to enable collaboration between operators, regarding 5G infrastructure services, with the view to introducing unification via NFV/SDN compatible multi-domain orchestration by producing an open platform enabling cross-domain orchestration of services over these multiple domains, with a set of open source software tools and extensions that can be utilised outside the scope of 5GEx; a Sandbox Network enabling experimentation and validation of the devised architecture, mechanisms, and business models.	**Collaboration:** Since COHERENT investigates a convergent LTE/Wi-Fi solution, we are particularly interested on the way the SDK developments of WP2 can be potentially integrated as part of the 5GEX cross-domain orchestration of services over multiple administrations or over multi-domain single administrations. **Common Partner Involved:** None
5G-NORMA: The key objective of 5G NORMA is to develop a conceptually novel, adaptive and future-proof 5G mobile network architecture. The architecture is enabling unprecedented levels of network customisability, ensuring stringent performance, security, cost and energy requirements to be met; as well as providing an API-driven architectural openness, fuelling economic growth through over-the-top innovation	**Collaboration:** As COHERENT also proposes an architecture paradigm in order to facilitate the control and coordination with respect to 3GPPP activities, we will exploit how coordination is achieved in 5G-NORMA in terms of performance, security, cost and energy. We are also highly interested on exposing our *centralised control and coordination* mechanism, together with our novel *network abstractions* for the RAN design, as a control realisation in the rest of the NFV solution proposed **Common Partner Involved:** None
5G-CHARISMA: CHARISMA proposes an intelligent hierarchical routing and paravirtualised architecture that combines two important concepts: devolved offload with shortest path nearest to end-users and an end-to-end security service chain via virtualised open access physical layer security (PLS).	**Collaboration:** As CHARISMA lays the foundations of the open access network architecture in 5G. COHERENT can cooperate with CHARISMA on network slicing. **Common Partner Involved:** OTE/COSMOTE
FANTASTIC-5G: FANTASTIC-5G develops a new multi-service Air Interface (AI) for below 6 GHz through a modular design. In particular, it develops the technical AI components and integrates them into an overall AI framework where adaptation to the above described sources of heterogeneity will be accomplished.	**Collaboration:** As COHERENT focuses on flexible spectrum management, FANTASTIC-5G investigates research aspects related to the efficiency of resource utilisation and adaptation of anticipated heterogeneity due to the new Air Interface (AI) for below 6 GHz. **Common Partner Involved:**

	None

2.3 Standardisation Bodies

For the sake of providing an evolutionary path to adapt COHERENT project to industrial needs, it is important to review the 5G enabling technologies (e.g., SDN and NFV) and mobile architectures provided by standards or similar organisations, such as Third Generation Partnership Project (3GPP), the European Telecommunications Standards Institute (ETSI) and Open Network Foundation (ONF). Since 3GPP work on 5G system architecture has just started, the COHERENT project intends to synchronise with current 3GPP advancements for 5G and provide significant inputs to 3GPP, despite the fact that 3GPP focuses rather on near-term development. Furthermore, the European Telecommunications Standards Institute (ETSI) has defined a NFV architecture which is a key enabler for COHERENT architecture. Finally, ONF has launched a wireless and mobile working group (WMWG) to enable programmability in mobile and wireless networks by introducing SDN concept for the control and coordination of mobile network functions, which is precisely the goal of the COHERENT project.

2.3.1 3GPP

As far as wireless cellular systems are concerned, 3GPP is surely the main standardisation driver. It is well known that 3GPP is ambitioning (already starting from current Release 13 and previous ones for some case) to support services which do not fall into the tradition cellular ecosystem: broadcast (with eMBMS), mission critical communications (targeting PMR market), machine-type communications (targeting IoT market). 3GPP is hence going toward a higher differentiation of supported services. Starting from the previous consideration and also seeing current standardisation activities, 3GPP is most probably one of the main standardisation organisations in which 5G definition and design will happen, at least as far as wireless cellular systems and related supported services are concerned.

In this section, besides the architecture used for a traditional cellular communication in Section 2.3.1.1, we introduce also state of the art architectures for broadcast/multicast multimedia services in Section 2.3.1.3, direct communications between devices, and coverage extension enabled by user equipment in Section 2.3.1.2, since these are configurations of interest for COHERENT use cases described in COHERENT D2.1 and considered in the COHERENT technical work packages WP3, WP4 and WP5.

2.3.1.1 LTE Architecture

For synchronising the 5G architecture within the existing 3GPP context, we will first present below, based on [26], the existing LTE Architecture. The LTE architecture, including Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and Evolved Packet Core (EPC) is illustrated in Figure 2-10 below. The E-UTRAN consists of eNBs which are interconnected with each other through X2 interface. The eNB provides the necessary functionalities required in LTE systems to achieve the physical wireless connections between user devices (UEs) and the network. Furthermore, the eNB contains the radio interface and processes radio resource management, including radio bearer control, radio admission control, and scheduling of uplink and downlink radio resources for individual UEs. eNBs are connected to the EPC, including Mobility Management Entity (MME), PDN gateway (P-GW) and serving gateway (S-GW), through the S1 interface, which is split up into the user plane and the control plane. More details about the protocol stack for the user and control planes can be found in Annex A.1. 3GPP has defined the functional split between radio access and core networks, which is summarized in Annex A.2.

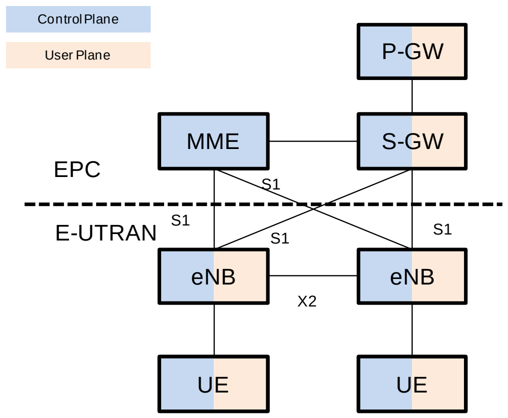

Figure 2-10 LTE network Architecture [26]

The S-GW serves as a local mobility anchor that enables seamless communication when the user moves from one eNB to another. The P-GW enforces quality-of-service policies and monitors traffic to perform billing. The P-GW also connects to the Internet and other cellular data networks, and acts as a firewall that blocks unwanted traffic. The policies at the P-GW can be very fine-grained, based on whether the user is roaming, properties of the user equipment, usage caps in the service contract, parental controls, and so on. As shown in Figure 2-10, the eNBs, S-GW, and P-GW participates not only the user-plane functionalities but also control-plane functionalities. From a core network perspective, the MME is the main entity for control of the LTE access network, handling all LTE-related control plane signalling, including mobility and security functions for devices and terminals attached over the LTE RAN.

2.3.1.2 3GPP Architecture for D2D and UE-to-Network Relay

In this section, the architecture design for D2D and UE-to-Network Relay in 3GPP is described. Due to the different architecture from traditional LTE, the protocol stack design is also tailored to the architecture for and UE-to-Network Relay. For more detailed information of the protocol stack design, please refer to Annex A.3. Please note that *Proximity Services* (ProSe) is the 3GPP term to indicate services built on top of D2D communication capability.

Architecture model using a ProSe UE-to-Network Relay has been taken into account by 3GPP work, for example in TS 23.303[27], Release-13, v13.4.0 Section 4.4.3, Figure 4.4.3-1) where it is showed a Remote UE connecting to a ProSe UE-to-Network Relay through PC5 interface. UE-to-Network Relay has a classical Uu interface with LTE network, and Remote UE can access the Public Safety AS (located outside the LTE network) through the UE-to-Network Relay. Further details about the user plane and control plane of PC5 interface are further provided in Annex A.3.

As described by Figure 2-11 below, and as a result of [28] (SA1 work in Release-12), six reference points have been introduced in 3GPP Release-13 and Release-12, as well as i) a ProSe Function responsible of Proximity Services and located in the Evolved Packet System or EPS (but outside the Evolved Packet Core or EPC), and ii) a ProSe APP Server (Proximity Services Application Server) that is outside of the EPS and connected through an SGi interface to EPC. The new component "ProSe Function" is created to support the EPS features of ProSe. This architecture has been further normalised in [27] (SA2) for both Release-12 and Release-13 for both roaming and non-roaming cases.

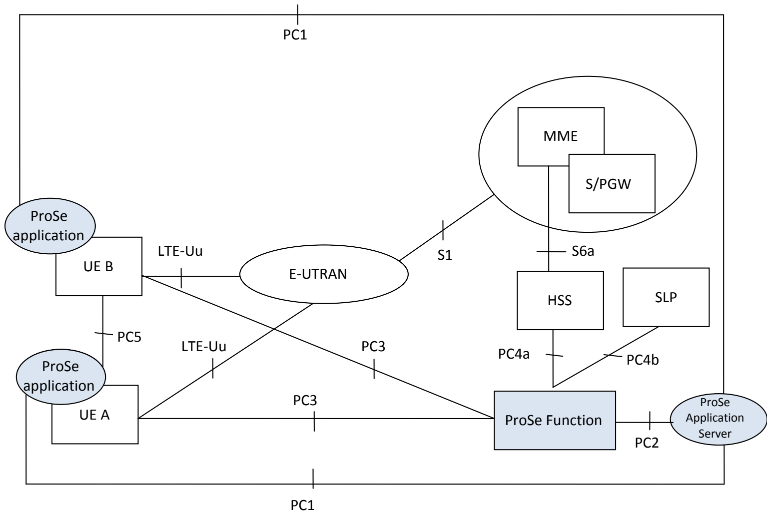

Figure 2-11 Non-Roaming Reference Architecture from [27]

The ProSe APP Server, which is located outside the 3GPP network, represents the application server of a ProSe provider. This architecture has to support a wide range of solutions and is thus very generic and flexible.

EPC-level Discovery ProSe Function has a reference point towards the Application Server (PC2), towards other ProSe Functions (PC6), towards the HSS (PC4a) and the UE (PC3). The functionality includes the following:

- Storage of ProSe-related subscriber data and/or retrieval of ProSe-related subscriber data from the HSS;
- Authorisation and configuration of the UE for EPC-level ProSe Discovery and EPC-assisted WLAN direct discovery and communication over PC3;
- Storage of a list of applications that are authorised to use EPC-level ProSe Discovery and EPC-assisted WLAN direct discovery and communication;
- Acting as location services client (SLP agent) to enable EPC-level ProSe Discovery;
- Providing the UE with information to assist WLAN direct discovery and communications;
- Handling of EPC ProSe User IDs and Application Layer User IDs;
- Exchange of signalling with 3rd party Application Servers over PC2 reference point for application registration and identifier mapping;
- Exchange of signalling with ProSe Functions in other PLMNs over PC6 reference points for sending proximity requests, proximity alerts and location reporting;
- Optional support for functionality for requesting UE location via the HSS.

The architecture allows a wide variety of implementations. The ProSe Function and the ProSe APP Server split can be very different. Some solutions can be based on the use of IMS in the ProSe APP Server, other solutions allow to use ProSe control plane messages over the PC3 interface between the UE and the ProSe Function by using IP user plane messages, while other solutions could use Non-Access Stratum (NAS) messages.

The architecture also supports all the features of ProSe, such as the discovery and the communication between UEs. This later one includes the communication required allowing a UE out-of-network coverage (often called as a Remote UE) to access to the network infrastructure through a UE-Relay (often called ProSe UE-to-Network relay), as we can see in Figure 2-12.

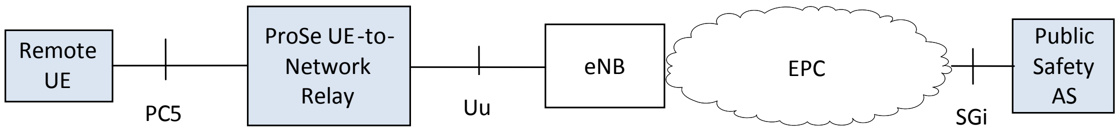

Figure 2-12 Architecture model using a ProSe UE-to-Network Relay [27]

Figure 2-12 shows a Remote UE connecting to a ProSe UE-to-Network Relay through PC5 interface. UE-to-Network Relay has a classical Uu interface with LTE network, and Remote UE can access the Public Safety AS (located outside the LTE network) through the UE-to-Network Relay.

With respect to previous architecture for Remote UE and ProSe UE-to-Network relay we can easily see that the same interface PC5 defined for ProSe applications is used also in the case of relayed UEs, but which presents a very poor level of control. Actually, the Remote UEs are no longer under the eNB control. Further details about the user plane and control plane of PC5 interface are further provided in Annex A.3.

2.3.1.3 Overview of eMBMS operation in 3GPP

The architecture for enhanced Multimedia Broadcast Multicast Service (eMBMS, and more generally referred to as MBMS) Operation described in this section is based on [26]. As represented in Figure 2-13, the main components for Multimedia Broadcast Multicast Service (MBMS) operation are BM-SC (Broadcast Multicast Service Center) and MBMS GW (MBMS Gateway). On the UE equipment side there is a MBMS user service and MBMS bearer service which are controlled by BM-SC. The mobile network (MN) is composed from an E-UTRAN part responsible of radio access and EPC part (evolved packet core, or core network). BM-SC can be considered as outside the mobile NW but is under the definition of 3GPP. The MBMS GW is the equivalent of the S/P-GW (Serving and PDN gateways) from classical 3GPP network and it acts as a gateway for MBMS service.

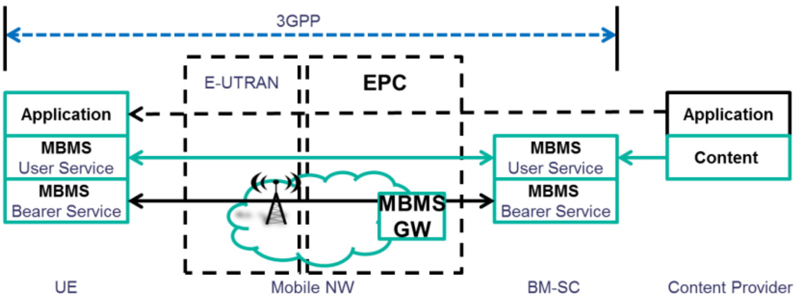

Figure 2-13 Simplified Architecture for MBMS Operation

A few functionalities are resumed below as follows:

- BM-SC functionalities:
  - User Service: membership, service announcement, security.
  - Bearer Service: session and transmission, proxy and transport, content synchronisation.
- MBMS GW functionalities:
  - Distributes MBMS user plane to eNBs.
  - Performs MBMS session control signalling via MME (control plane).

MBMS GW can be further split in MBMS GW for User Plane (UP) and MBMS GW for Control Plane (CP). Figure 2-14 and Figure 2-15 below represent the MBMS architecture with interfaces and protocol stack. Figure 2-15 also presents a short comparison with other network entities and interfaces from 3GPP normally used for unicast transmission (e.g., S-GW, P-GW). The new interfaces for

MBMS operation are M1 (between MBMS GW UP and eNB), M2 which is for optional use if Multi-cell/multicast Coordination Entity or MCE function is located outside eNB (M2 is between MCE and eNB for signalling), M3 which is between MCE and MME, Sm between MBMS GW and MME, SGi-mb and SGi.

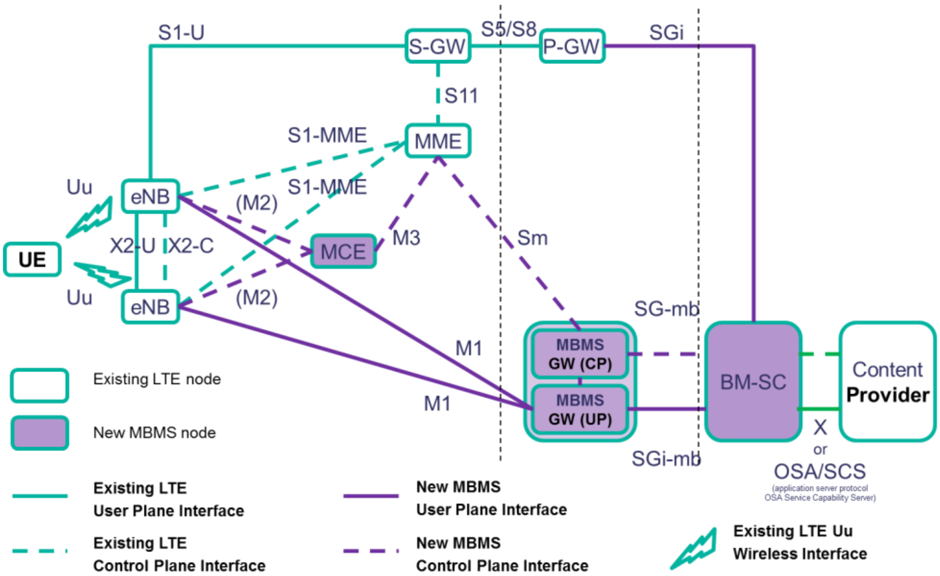

Figure 2-14 MBMS Architecture without Protocol Stack

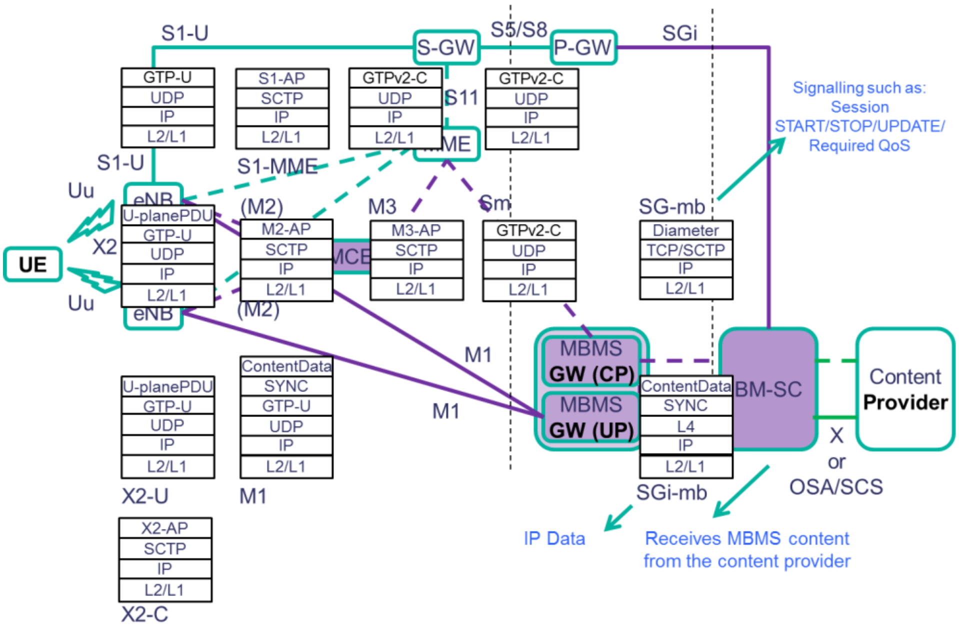

Figure 2-15 MBMS Architecture with Protocol Stack

With respect to MCE, its functionalities can be resumed below as follows:

- The MCE can be a separate entity or as part of eNB.
- Provides admission control:
  - Allocation of the time/frequency radio resource for eMBMS.
  - Deciding the radio configuration, e.g., Modulation and Coding Scheme (MCS)

2.3.1.4 Overview of 3GPP Roadmap to 5G

In order to better integrate future COHERENT work and to adapt it to current industrial needs, it is important to understand 3GPP tentative roadmap to 5G. This section aims at providing an overview of

3GPP roadmap to 5G. Please refer to Annex A.4 for a more detailed description on 3GPP Advancement with respect to 5G.

As shown in the figure below, we expect to talk about a complete 5G normative work not earlier than 2019 and a 5G deployment not earlier than 2020. 5G will most probably correspond to Release-15 and Release-16, as we will further explain and show. It is very important to explain these aspects, because they are closely related to COHERENT work on architecture. Since 3GPP work on 5G system architecture just started, COHERENT intends to invest significant efforts to influence the philosophy behind next 5G 3GPP architecture in WP7 activity.

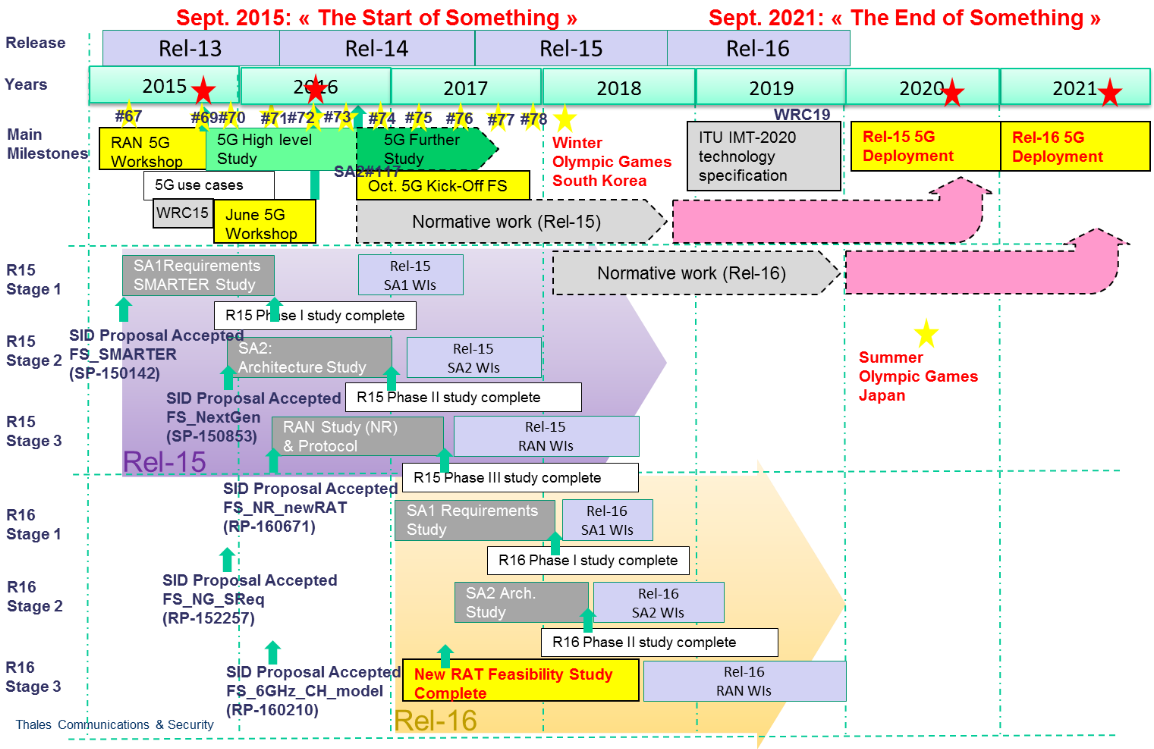

Figure 2-16 3GPP Advancements towards a 5G Architecture and Expected 5G Network Deployment

In 3GPP, the TSG (Technical Specification Group) RAN (Radio Access Networks) is responsible for the definition of the functions, requirements and interfaces of UTRA/E-UTRA network; while the TSG SA (Service & Systems Aspects) is responsible for the overall architecture and service capabilities of systems based on 3GPP specifications, but also has the responsibility for cross TSG coordination. As also shown in the figure describing current 3GPP advancements toward a 5G architecture, 3GPP has a method composed from 3 main steps:

- Service aspects and network requirements covered by Stage 1 (SA1),
- Architecture completion covered by Stage 2 (SA2) and
- Completion of the protocols covered by Stage 3 (RAN).

Then, there are two phases in each step, namely, study phase and Work Item Description (WID) phase. In the study phase, the best solution is chosen, and approved Work Item Descriptions (WIDs) in the 2nd phase lead to the definition of Technical Specification (TS) documents that are normative. One could foresee that a first 5G expected deployment may be in 2020, just in time for 2020 Summer Olympic games in Japan. Similar expectations are for 5G Release-16, but with a potential deployment by the end of 2021.

The first Study Item on 5G was first initiated by SA1 work which is responsible of new features, new services, charging requirements and new system and service capabilities. SA initiated the work with SA1 SMARTER (Study on New Services and Markets Technology Enablers) TR 22.89 [29], and the work propagated into SA2 (responsible of system architecture) with a new SID FS_NextGen and TR 23.799 [30].

With its SMARTER initiative, SA1 working group defined a total of 74 use cases with extremely different and sometimes conflicting requirements. As an example, some applications require increased spectral efficiency; some of them increased robustness and reliable synchronisation, and some of them asynchronous transmissions. Some of the considered use cases are totally new and never seen before, which is actually a huge step for 3GPP work seen now as "revolution" and not only as a simple "evolution". Some of the use cases refer to connectivity of drones (SMARTER UC12 & UC54), some of them to moving ambulance and telemedicine support (SMARTER UC65 and UC68), some of them to connectivity under high speed scenarios (SMARTER UC66) and some of them even to connectivity using satellites (SMARTER UC72). In COHERENT, some of these use cases are further discussed in section 6 of D2.2 (Support of Use Cases with COHERENT Architecture) and further investigated in D3.1 and D3.2.

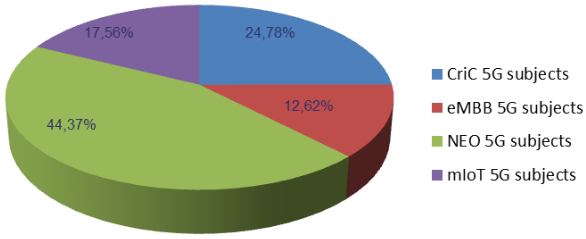

Figure 2-17 5G Current Composition

An initial estimation of 5G evolution, inspired from its use cases and new services shows that 5G is composed from (at least):

- 24,78% of Critical Communications (CriC) subjects;
- 17,56% of massive Internet of Things (mIOT) subjects;
- 12,62% of enhanced Mobile Broadband (eMBB) subjects;
- 44,37% of Network Operation (NEO) subjects.

There is more than 40% of estimated work on NEO, meaning that at the time being a lot of work has to be done on network side operation for 5G. This trend conforms to the design principles of COHERENT architecture:

- Elastic adaptation to individual network infrastructure,
- Dynamic adjustment to the network changes
- Flexible arrangement for spectrum usage.

2.3.2 Network Function Virtualisation (NFV)

At the standardisation level, the ETSI NFV ISG is defining concepts, architectures, and interfaces for delivery and management of VNFs and their service chains. In Figure 2-18, a high level representation of the ETSI MANagement and Orchestration (MANO) architecture is sketched. The following terminology is used:

- VNF: the virtualised network element like Router VNF, Switch VNF, Firewall etc.
- VNF Catalog: a repository of all usable VNF Descriptors (VNFD). VNFD describes a VNF in terms of its deployment and operational behaviour requirements.

- Network Services Catalog: catalog of the usable Network services. A deployment template in terms of VNFs and description of their connectivity through virtual links.
- NFVI Resources: a repository of NFVI resources utilised for the purpose of establishing NFV services.
- Virtualised Infrastructure Manager (VIM): manager of NFVI resources in one domain.
- VNF Manager (VNFM): managers of life cycle of VNFs. It creates, maintains and terminates VNF instances, installed on VMs which the VIM creates and manages.

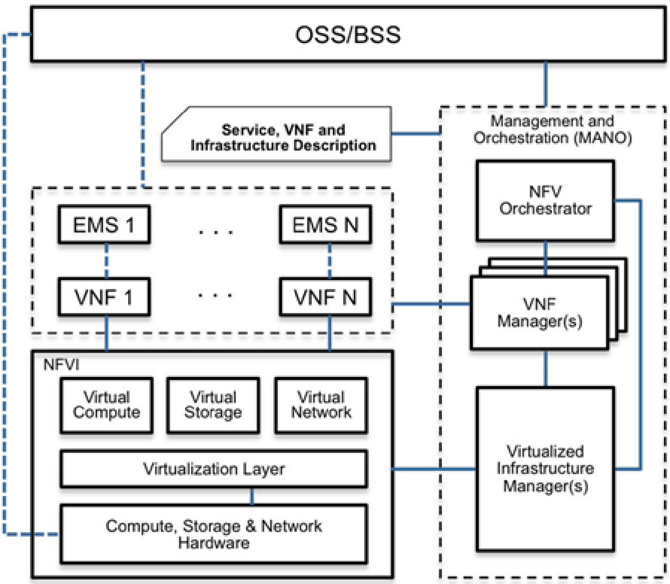

Figure 2-18 ETSI MANO architecture[31]

In ETSI NFV MANO, a network service is composed of a number of VNFs interconnected through a VNF Forwarding Graph (VNF-FG). In particular, the information related to the interconnectivity between the VNFs in order to deploy a network service, are described in a Network Service (NS) template, that is stored in the Network Services Catalog. The NFV Orchestrator, in combination with the VNF Manager, is in charge of deploying the network services over the physical infrastructure, configuring and operating (scale-in/scale-out) the VNFs as declared in the NS template, covering all the VNFs lifecycle. In terms of software, several open source projects are addressing platforms for NFV Infrastructures (NFVI) and NFV MANO tools. For example, Virtualised Infrastructure Manager (VIM) and NFVI are the current focus of the OPNFV [32] initiative, which has the goal to provide NFVI and VIM components, as well as their open APIs. Other projects are more focused on the management and orchestration functions of the NFV MANO architecture: for example, OpenBaton [33] and OpenMANO [34] provide open source software for NFVO and generic VNFMs. Moreover, a recent initiative from ETSI is developing an open source MANO software aligned with ETSI NFV, with a first demonstration performed at Mobile World Congress (MWC) 2016 in Barcelona.

COHERENT will exploit the state-of-the-art on issues like service chaining techniques using flow tagging or a routing header, or correct forwarding actions to packets that have to traverse the COHERENT network. Another relevant capability provided by COHERENT testbed will be the possibility to automatically deploy complex virtual environments with a rich set of VNF-based embedded tools, depending on the experimentation specific features extended in order to support the specification of further components related to the control, management and monitoring features. We highlight that the COHERENT partner, Eurecom, has already signed for Open Source MANO.

The COHERENT design allows for backward compatibility with existing cloud systems and inherently supports the SDN and NFV design principles. Furthermore, it is flexible enough to include other design paradigms like the concept of the Network Slicing. Actually work in parallel with systems will be able to realise the ETSI MANO architecture. As an example, the all the LTE components can be exposed through a NFV framework as a service, where the

COHERENT solution will work in parallel and will be responsible to support all the necessary real-time and non-real-time control functionalities required, especially in the RAN.

2.3.3 Software Defined Networking (SDN)

This section provides an overview of ongoing standardisation efforts in the framework of SDN. The Open Networking Foundation (ONF), working on promoting SDN and OpenFlow [35] technologies, defined an SDN architecture model as depicted in Figure 2-19.

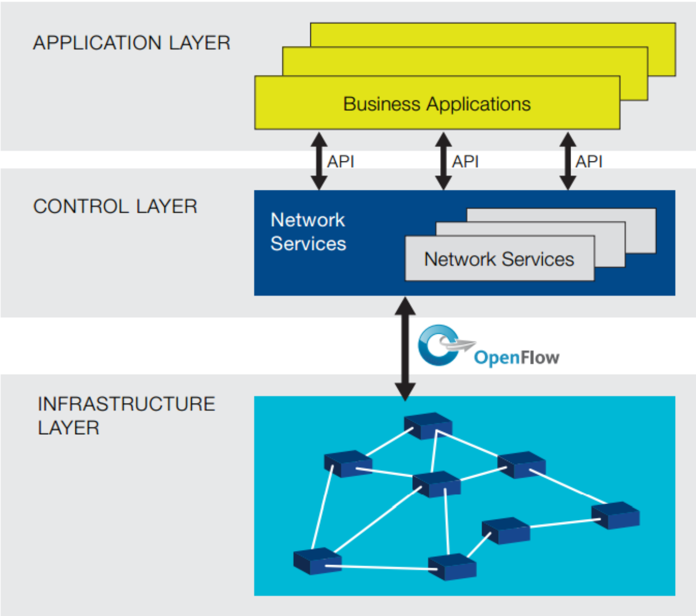

Figure 2-19 ONF-SDN architecture [36]

It has launched a Wireless and Mobile Working Group (WMWG) to collect use cases and determine the architectural and protocol requirements for extending ONF technologies (e.g., OpenFlow) for mobile and wireless networks. The ONF white paper [36] proposed two use case of OpenFlow-based SDN for mobile and wireless networks, namely inter-cell interference management and mobile traffic management. The Figure 2-20 shows an OpenFlow-enabled SDN architecture for inter-cell interference management, proposed in [36] by ONF. In Figure 2-20, SDN controller communicates with the base stations through the standard southbound interface (OpenFlow), any Radio Resource Management (RRM) upgrades can be achieved independently from the base station hardware. Compared to the distributed RRM, the OpenFlow-enabled SDN architecture shown in Figure 2-20 enables the centralised control of radio resource allocation with the global network view across various base stations (e.g., eNodeB 1, eNodeB 2 and eNodeB 3 in Figure 2-20).

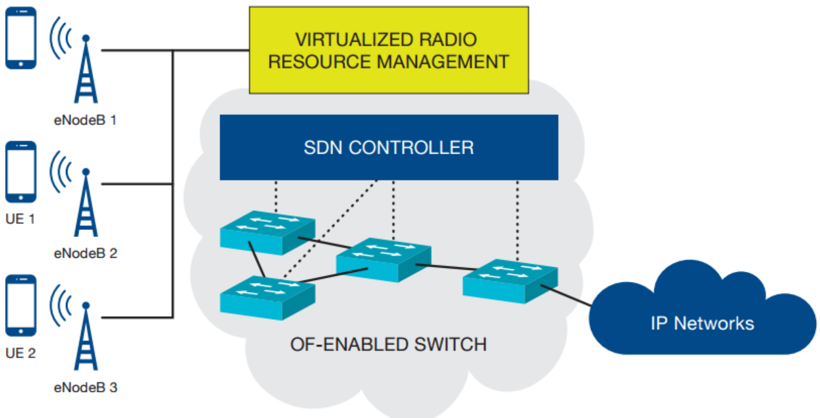

Figure 2-20 OpenFlow-enabled centralised base station control for interference management [36]

The Figure 2-21 the OpenFlow enabled SDN architecture for traffic management in mobile and wireless networks. As shown in Figure 2-21, the OpenFlow controller (OF controller) interacts with entities such as the Access Network Discovery and Selection Function (ANDSF) in order to discover available wireless networks close to the mobile user and perform traffic offloading between mobile networks and WiFi networks. Selection of the networks could be based on a Quality of Service (QoS) metric such as performance, signal strength, or distance, which are provided through the southbound interfaces (with the use of OpenFlow) between OF controllers and WiFi APs/base stations.

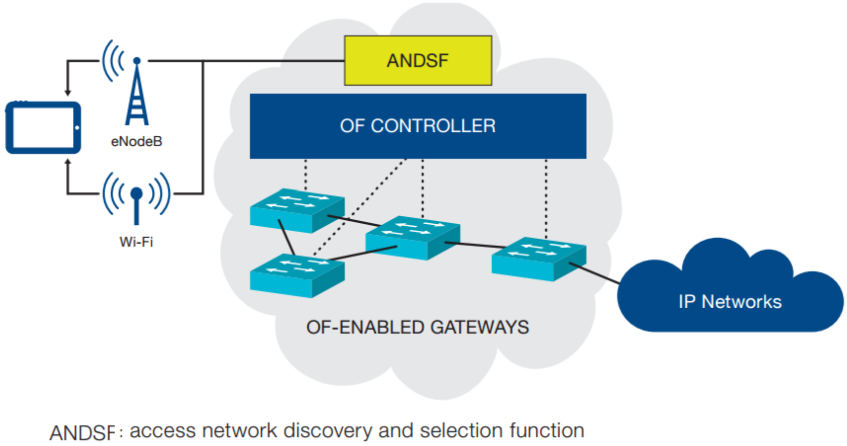

Figure 2-21 OpenFlow-based mobile offload [36]

The Internet Engineering Task Force (IETF) is also actively developing RFCs which are conceptually related to SDN, for example, the Forwarding and Control Element Separation (ForCES) working group. ForCES working group was established in 2001 and closed in 2014. The Figure 2-22 shows the ForCES architecture defined in [37]. There are two kinds of components inside a ForCES Network Element (NE): Control Element (CE) and Forwarding Element (FE). The framework allows multiple instances of CE and FE inside one NE. ForCES protocol provides a universal standardised control interface for FEs.

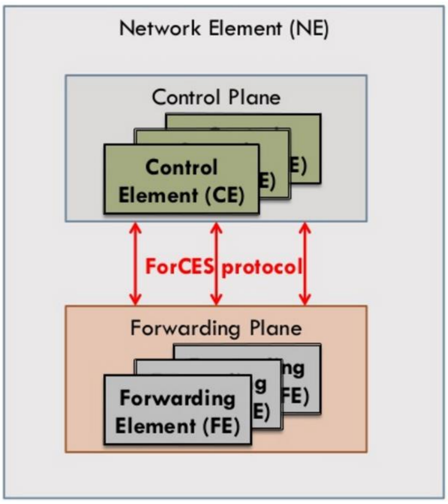

Figure 2-22 ForCES architecture[37]

Moreover, the Internet Research Task Force (IRTF) has a working group called the Software-Defined Networking Research Group (SDNRG) which aims at assisting other standardisation organisations by defining SDN architecture terminology and investigating open issues related to SDN. RFC 7426 [38] published by SDNRG provides a concise reference for SDN architecture and its terms for the SDN research community. Figure 2-23 depicts the SDN architecture abstractions in RFC7426. The architecture spans multiple planes, including forwarding plane, operation plane, control plane, management plane and application plane. The forwarding plane is responsible for handling packets in the data path based on the instructions from control plane. The operational plane is responsible for managing the operational states of the network device. While the control plane make packet forwarding decisions and sends the decisions to the network devices for executing the decisions, the management plane usually focuses on monitoring, configuring and maintaining network devices. Finally, the applications and services reside in the application plane. For more details please refer to RFC 7426 [38].

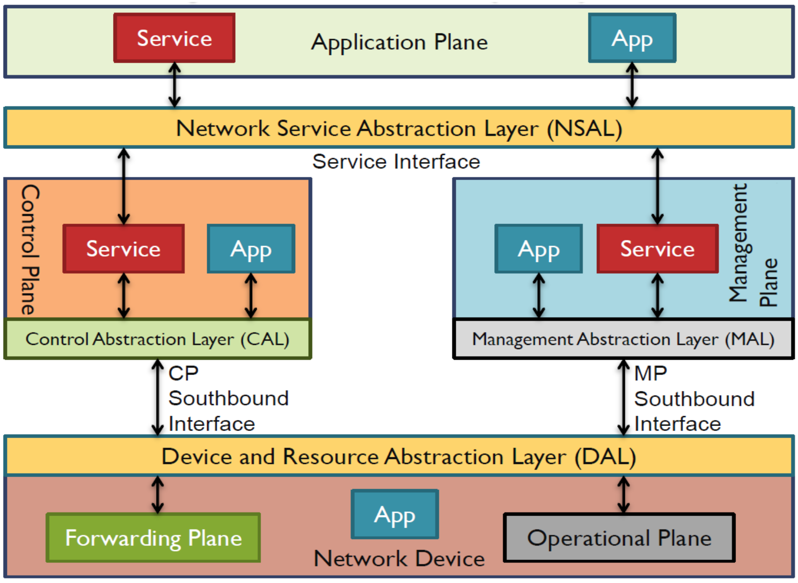

Figure 2-23 RFC 7426 SDN layer architecture [38]

2.4 Academic Literature Review

The ongoing research on software defined wireless network is briefly summarised in this section. The concepts and recent work on SDN are surveyed in [39], [40], [41]. [39] and [40], which presented the key building blocks of SDN infrastructure and the in-depth analysis of the hardware infrastructure, southbound and northbound APIs, network virtualisation layers, network operating systems (SDN controllers), network programming languages, and network applications. [41] introduces the concept of network virtualisation and explains the difference between service and resource based virtualisation.

Virtualisation and SDN control in the wireless domain have been extensively investigated in [42], [43], [44], [45], [46] and [47]. In [46] the architecture for software-defined RAN via virtualisation called OpenRAN is introduced. In [47] the CROWD solution is proposed for SDN control in DenseNets. In [48] the software defined architecture is developed for HetNets. The SDN and virtualisation in mobile networks taxonomy is summarized in [49]. The use of OpenFlow in LTE core network is investigated in [50]. IP-based routing in virtualisation-based LTE EPC architecture (vEPC) is considered in [51]. In [52] an OpenFlow SDN framework for Wi-Fi networks is proposed. An SDN-based plastic architecture for 5G networks consisting of unified control plane and a clean-slate forwarding plane is presented in [53].

One of the first work on programming abstractions for wireless and mobile networks can be found in [54] and [55], where the authors identify the invariances of network control in wireless networks and clearly separate network control from network management, acknowledging the latency requirements of the former as opposed to the soft real-time requirements of the latter. The work is extended to flexible functional splits and radio slicing in Enterprise WLANs in [56] and [57].

Note that RAN densification [47][58] and Cloud-RAN (C-RAN) [59][60][61] seem the most promising candidates to meet the extreme requirements of 5G communications and the exponential increase in user traffic. Multi-tier cellular systems are presented in [62], [63] and [44]. The role of SDN in the 5G evolution is present in [64].

From architecture point of view, the studies on software defined wireless networks can be grouped to three main directions: ideas derived from SDN for Internet, centralised solutions similar to C-RAN, and new approaches applying SDN design principles at the mobile edge.

SDN oriented approaches: The majority of the software defined wireless network research derives from the original SDN concept. The common features are the decoupling of the control and data plane, and the use of the logical centralised control.

OpenRoad is the very early study on this topic [65]. It is the mobile version of SDN, which use OpenFlow for control, FlowVisor for network slicing, and NOX as the network operation system to support the programmable control in WiFi and WiMAX networks. OpenRoad allows different control algorithms concurrently running in one network, and thus realises network slicing, one of the key features in SDN. Network slicing is extended to cellular networks in [66], where the network virtualisation substrate and CellSlice are proposed to virtualise wireless resources and allow virtual mobile network operators to coexist in a single physical network.

Softcell is the first effort to extend the SDN concept to the mobile CN [67]. It applies SDN principles to redesign the control plane of CN. The centralised controller and the flow concept allow the previously centralised packet processing in CN to be distributed among separated packet processing middle-boxes, and thus improve scalability and flexibility. MobileFlow was proposed as another flow-based forward model to facilitate the deployment of new services and network features in the mobile CN [68]. An OpenFlow controller is introduced in [69], which allows the separation of control and user plane in CN of LTE networks and moves core control functions in CN to the cloud for reliability and scalability. A similar idea has been proposed in [70], where the mobile network SDN controller governs not only LTE, but also Universal Mobile Telecommunications System (UMTS), WiFi, and other wireless networks.

Further, a SDN-based plastic architecture is introduced for 5G networks [53], with the aim to support a heterogeneous set of services with flexibility. It introduces a clean-slate date plane design, and the SDN controllers at three levels, i.e. device, mobile edge, and CN, respectively. By this design, it avoids the use of tunnel protocols for mobility, and allows backward compatibility to 4G networks.

C-RAN oriented approaches: C-RAN oriented approaches centralise not only the control but also partial of the radio signal process in the network.

SoftRAN is one of the early proposals under this approach [42]. It virtualises the RAN into a single virtual BS, performing resource allocation, mobility, load balancing, and other control functions at a single place. The centralised control plane of the network takes advantage of full network knowledge for global optimisation. To solve the latency problem, time-critical controls remain at the local BS.

A recent design proposed by Arslan et al. combines the centralised signal process in C-RAN and the programmable feature at the fronthaul [71]. The Software-Defined Fronthauls (SDF) form a fronthaul network, where joint processed radio signals are forwarded to fronthauls by the centralised control. The control architecture is similar to SDN for Internet. The programmability in the fronthaul network allows practical fine timescale physical layer cooperation like CoMP. It provides potential for the fine-grained RAN optimisation in extremely dense wireless networks.

Mobile edge oriented approaches: Mobile edge oriented approaches apply the SDN design at RAN. The need of this approach comes from the adaption of the air interface as well as the fine-grained radio function coordination in dense wireless networks. To adapt air interface behaviours to network conditions, Bianchi proposed the MAClet concept, which allows the central controller dynamically change the MAC process in air interfaces, e.g., from the contention based medium access to time-division multiple access [72]. The SDF proposed in [71] is also a mobile edge solution, which brings the programmability to the radio fronthaul.

3. Requirements on Programmable 5G Architecture

An analysis of the relevant international 5G initiatives, investigation of 5G business models and requirements was presented in Coherent D2.1. In this section we present a set of fundamental requirements for the COHERENT Programmable 5G Architecture.

The goal of COHERENT is to provide a framework for control and coordination of heterogeneous wireless and mobile networks. Such framework must be able to support a number of cloud-based RAN services, to provide automation support in processes which involve different stakeholders, while at the same supporting existing and future wireless network technologies.

How can virtual networks be created? Which mechanisms are required to perform SDN control in all network segments and integrate the wireless LTE and Wi-Fi domains, in both the data and control planes of the architecture? How network slices are created and operate per virtual network operator? How the concept of RAN sharing fits into the picture? These are some of the basic challenges our solution addresses, exploiting a multi-domain architecture that is SDN-based, aligned with ONF and the rest of 5G-PPP activities, flexible enough for extreme service orientation.

By means of service requirements in order to enable programmable wireless networks, the SLAs we consider are multilevel. They provide a description for each stakeholder (like MVNOs and individual users) and define the services, the service duration and the service requirements that will be provided to him. The SLA considers for QoS parameters such as bandwidth, delay, jitter, while it should guarantee the delivery of real time services to the network providers, with a minimum/maximum QoS level offered. In terms of the integrated infrastructure requirements, our system is built over a heterogeneous infrastructure wireless access networks based on Wi-Fi and LTE technology. In addition, it supports isolation of virtual infrastructures and authorisation control of accessing them.

The upgrading and downgrading of already provided virtual infrastructures must be supported as well as the dynamic optimisation of the allocation of the virtual infrastructures over the physical resources. By means of integrated network service requirements, the integrated system provides end-to-end network connectivity and guarantees the efficient operation of the overall infrastructure across the different network domains. Furthermore, the COHERENT system supports the provisioning of per-user network services compliant with the profiles defined in the SLAs, provide mechanisms for network service monitoring in order to be able to verify the SLAs and provide procedures for end-to-end service resilience. End-user authentication and service access authorisation are achieved as well as the access to the full combination of cloud and network services.

Accurate observability of the network state represented in terms of network graph abstractions requires fast and resource-effective retrieval and processing of network metrics and monitoring information. This raises the need for real-time control which from our point of view imposes the strictest requirements in contrast to the wired SDN case where control functions can operate with less stringent deadlines.

Control functions operating in a centralised, decentralised or distributed manner and at different architectural levels generally require effective monitoring functions running close to the data source, capable of providing up-to-date state information while minimizing signalling overhead.

General requirements on what should be abstracted and how, need to be considered with respect to different control functions operating at different levels of the architecture. This is specifically important for time-critical RRM and RF control functions where the observation granularity needs to be consistent with the control-loop timing.

The COHERENT architecture is designed to meet several requirements related to timeliness, overhead and accurate representation of the network state at various time-scales, allowing for that certain elements of the network graph can be stored in a highly distributed manner in addition to centralised storage. Prominent requirements (as listed in D2.1) that need to be fulfilled by the COHERENT architecture are:

- R#8: "COHERENT shall provide well-defined open interfaces and protocols to support programmable control".
- R#28: "COHERENT shall allow for distributed monitoring, aggregation and analysis at various levels of the architecture, supporting self-organisation as well as distributed and centralised control/coordination".
- R#67: COHERENT shall be able to provide control and coordination interface between neighbouring cells / clusters.
- R#70: COHERENT shall be able to monitor and control network resources such that reliability, availability and network resilience are improved if requested.

The above general requirements have been formulated to enable the possibility to deploy and redeploy monitoring functions as needed for increased network observability (R#28, R#70). Up-to-date information about the network state can only be fulfilled by combining distributed and centralised processing of monitoring information and various data sources at the level of real-time control functions, regional controllers and the logically centralised controller and coordinator (R#28). For sophisticated control and autonomous network operation, exchange of monitoring information vertically and horizontally across the architecture need to be allowed through proper interfaces (R#8, R#67).

As described in Section 4.2 and Section 5.4, the architecture is designed to address the aforementioned requirements by being split into a logically centralised controller and coordinator implemented by multiple control instances and control functions for real-time networking operations. The split into real-time control functions and logically centralised control instances allow for representation of the network state at several levels and time-scales. The highly distributed real-time control functions can make direct use of local monitoring functions, which are also capable of providing information to network graphs managed by higher-level regional controllers.

4. COHERENT Concept

Future 5G heterogeneous radio access networks need a programmable *control and coordination* that offers fine-grain, real-time control without sacrificing scalability. The programmable control and coordination is driven by a key characteristic, namely *abstraction*. The control and coordination plane capitalises abstraction of low-layer resources in radio access networks. The abstracted resources should allow any RAN operations desired by the *network slices* while hiding the configuration details of RAN hardware. More specifically, abstracting RAN resources manages the complexity and greatly simplifies the implementation and deployment of advanced control and coordination functions in the RAN, while leveraging on a variety of physical layer technologies. Therefore, in this section, the key concepts of COHERENT architecture, namely control/coordination, abstraction and network slicing are introduced in sections 4.2, 4.3 and 4.4 respectively. Moreover, the terminology used in COHERENT is defined in section 4.1.

4.1 Definition

This section defines the terminology used in the COHERENT:

- **Radio Transmission Point (R-TP):** R-TP is a radio access point implementing full or partial RAN node functions while rest of functions are offloaded to and handled by the vRP. An R-TP may include control plane functions.

- **Virtual Radio Processing (vRP):** vRP is a computing platform allowing for centralised processing of full or partial RAN node functions (including the user plane and the control plane) offloaded from one R-TP or multiple R-TPs. A vRP may include control plane functions.

- **Radio Transceiver (RT):** RT is a logical radio access entity with full RAN node functions, which is the flexible combination of R-TP, vRP and RTC functions. A set of RTs is forming a radio access network which is coordinated and controlled by C3. There are multiple physical and virtual resources and components in one RT. Some examples of physical RTs include LTE eNBs in cellular networks or WiFi APs in the WLANs. An RT could be composed by one vRP (virtual device) and one or more R-TPs (physical devices). For example, in the Cloud-RAN architecture the R-TP coincides with the RRH, while the vRP coincides with the BBU Pool, however several other functional splits are considered in this project. In some particular case, e.g., D2D, RT could be an UE, being a relay node.

- **Transport Node (TN):** TN is the entity located between RTs and core network. A set of TNs is forming a backhaul/fronthaul network whose data plane can be configured by the C3. A network switch is an example of Transport Node.

- **Real-Time Controller (RTC):** A logical entity in charge of local or region-wide control, targeting at real-time control operations, e.g., MAC scheduling. It has local network view. It could run directly on one RT or on a virtualised platform and receives monitoring information gathered from one RT or multiple RTs. It can delegate control functionality to the RTC agent on the RTs. RTC communicates with an RTC agent/RTC agents on one RT or multiple RTs.

- **Central Controller and Coordinator (C3):** A logical centralised entity in charge of logical centralised network-wide control and coordination among entities in RAN based on centralised network view. C3 could be implemented with distributed physical control instances sharing network information with each other. Sharing network information among C3 instance creates the logically centralised network view and therefore achieves logical centralised control and coordination.

- **Slice:** A network slice is defined as a collection of specific network services and RAT configurations, which are aggregated together for some particular use cases or business applications. A network slice can span all domains of the network: software programs running on cloud nodes, specific configurations of the transport network, a dedicated radio access configuration, as well as settings of the 5G devices. Different network slices contain different network applications and configuration settings. Some application modules in network slices may be latency-critical. For such a slice, these modules are located in the RTC.

4.2 Control Separation to Network-Wide Control and Real-Time Control

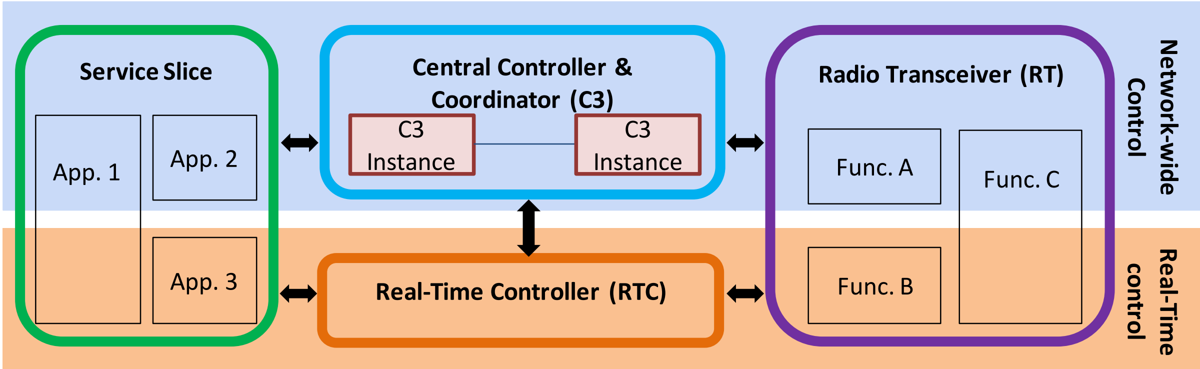

Figure 4-1 Network wide control and real-time control

The optimal resource utilisation in heterogeneous mobile networks (HMNs) is a hard problem. Recall that even the frequency assignment problem in Global System for Mobile Communications (GSM) networks is Non-deterministic Polynomial-time hard (NP-hard), not to mention the multi-dimension resource allocation in complex HMNs. Solutions with distributed control are scalable and flexible, but often yield sub-optimal results far away from the expectation. Therefore, COHERENT architecture adopts a centralised solution, SDN in RAN, which could achieve the global optimisation. Furthermore, applying SDN in 5G RAN enables the programmability. The benefit of global optimisation and programmability comes at the expense of scalability and latency. To address the scalability and latency issues, two control mechanisms are designed for achieving programmable 5G RAN, namely network-wide control and real-time control as shown in Figure 4-1.

As depicted in Figure 4-1, the Central Controller and Coordinator (C3) is a logically centralised entity[3], which provides network-wide control/coordination for the networks. Based on the centralised network view, the SDN principles are applied in the design of the C3. For overcoming scalability issue in a large and dense RAN deployment, or for performance/reliability reasons, the logically centralised C3 could be implemented with distributed physical control instances sharing network information with each other. Sharing network information among C3 instance creates the logically centralised network view and therefore achieves logical centralised control and coordination. The distribution of abstraction shields higher layer from state dissemination and collection, making the distributed control problem a logically centralised one.

The biggest challenge in creating such a software defined radio access network is the inherent delay between any centralised C3 and the individual radio elements. For example, RAN operations require for real-time control when it comes in Resource Block (RB) scheduling. Essentially what we want to affect is the RB allocation in the MAC using the existing resource elements definition under normal control plane mode. Note that, in 3GPP LTE, the radio frame has a length of 10 ms and each frame is divided into ten equally sized subframes of 1 ms in length and the scheduling is done on a subframe basis for both the downlink and uplink. The controller operations for such a network function (RB scheduling) must be extremely fast and supporting hard real-time applications where worst-case execution time is at least as important as average-case execution time.

[3] Note that defining C3 as a logically centralised entity neither prescribes nor precludes implementation details, e.g., the federation or hierarchical connection of multiple control instances.

Under such limitation of the inherent latency, we cannot simply expect the C3 to adapt perfectly to the rapidly varying channel conditions at the radio elements in certain cases. To overcome the latency challenge, the real-time controller (RTC) shown in Figure 4-1 is designed to offer real-time control.
RTC should be close to the physical radio elements so that RTC could adjust to rapidly varying wireless networks. Furthermore, for the sake of prompt control, RTCs in the RAN do not coordinate with each other and therefore, the network information is not shared between RTCs. Therefore, RTCs perform distributed control in the RAN.

By separating control functionalities between the C3 and the RTC, the C3 makes decisions that affect the logically centralised network states, while the RTC handles control decisions for latency-sensitive network functionalities in radio transceivers (RTs) without coordinating with other RTCs. Moreover, different network slices contain different network applications and configuration settings. Some application modules in network slices may be latency-sensitive. For such a slice, these modules are located in the RTC.

4.3 Abstraction

RAN coordination and programmability are central concepts in 5G that are aimed to improve service quality, resource usage, and management efficiency, while addressing the limitations of the current LTE and WLAN systems operating under highly distributed control [73]. A coherent representation of the network state and infrastructure resources is crucial for effective RAN coordination and control of programmable infrastructures and services. Moreover, programmable infrastructures require programmability constructs that provide means to observe and manipulate virtual and physical Network Functions (NFs) and their behaviour via high-level abstractions.

In COHERENT, abstractions encompass representations and models of time-frequency resources, spatial capabilities (i.e. number of transmit and receive antennas), as well as throughput per network slice or per allocated resources. In principle any data structure can be used for storing and accessing abstracted representation of the network state (e.g., CQI defined in LTE). However, for unified large-scale coordination of infrastructure resources, structuring network information into network graphs in a systematic way offers effective representation of physical and virtual infrastructures. These can be applied to RAN coordination and wide range infrastructure coordination and control operations. Essentially, network graphs enable the possibility to apply mathematical models and algorithms, which often leads to highly efficient solutions in terms of convergence and network performance [74][75]. Graph-based abstractions can for example model LTE resource allocation problems that are solved efficiently using constraint satisfaction and local search algorithms [74].

4.3.1 Conceptual overview

The concept of network graph abstractions maps horizontally and vertically at different levels of the proposed architecture. The elements of network graphs (i.e. the vertices and edges) are created from distributed data sources such as raw metrics provided by the infrastructure or more sophisticated monitoring functions operating in a decentralised and centralised manner. A network graph is created by collecting information accessible directly from infrastructure entities or stored in some dedicated storage (e.g., storage networks and databases). At the level of local real-time controller entities, a network graph may represent the network state relative to a certain RAT infrastructure and associated nodes (e.g., WiFi or LTE). Network graphs at the centralised coordination level represent the state of a defined part of the network, such as a smaller region or domain. High-level network graphs (e.g., for centralised coordination and control) can be aggregated based on selected sets of regional network graphs for the purpose of multi-domain coordination.

In the hierarchical architecture, controller instances implement capabilities for creating network graphs for performing control operations and for providing regional or logically centralised views of the infrastructure. The capabilities include functionality for: 1) gathering network information from distributed data sources for the purpose of creating a network graph; 2) aggregating existing network graphs; 3) processing of network graphs for the purpose of coordination and control operations; 4) disseminating network graphs and results to other controllers and network entities upon request or as part of a coordination and control operation, synchronously or asynchronously. Local processing of network information and network graphs is necessary for meeting the requirements on scalability and timeliness, meaning that for real-time controller functionality the data sources should remain as close as possible to the local controller. Centralised coordination and control involve data transactions and information exchanges between physical and logical controller instances (e.g., located at geo-distributed data centers), suitable for less time-critical orchestration and management of infrastructure resources.

4.4 Network Slicing and Slice-Specific Network View

The industry consensus is that by 2020 there will be the emergence of a new 5G radio access standard, where 5G network of the future will involve the integration of several cross-domain networks and the 5G systems will be built to enable logical network slices across multiple domains and technologies. This will enable operators to provide networks on an as-a-service basis and meet the wide range of use cases that the 2020 timeframe will demand[1][2]. In the same context we consider that a profound relationship exists between the concept of Network Slices and 5G integrated environments.

In the existing approaches, the infrastructure resources are shared among several parallel network slices. Nevertheless, every provider uses a specific control framework or/and a specific cloud management system (e.g., OpenStack), but actually all the configuration effort and fine-tuning of the components is left to the users. When it comes to complex scenarios, the configuration effort required in order to manage and fine-tune every single software or hardware component, actually supersedes the research effort and leaves absolutely no opportunities for a technological breakthrough.

The Network Slices envisioned in COHERENT, span the whole protocol stack from the underlying (virtualised) hardware resources up to network services and applications running on top of them. This approach is aligned with the industry and telecom perspective, towards 5G [1][2], in order to meet the demands of extremely diverse Use Cases. From our point of view, a *COHERENT Network Slice*, is a composition of adequately configured network functions, network applications, and the underlying cloud infrastructure (physical, virtual or even emulated resources, RAN resources etc.), that are bundled together to meet the requirements of a specific use case or business model. Note that advanced orchestration and automation is required in order to provide the necessary functionality and derive from the "silo" approach of independent functions to an integrated end-to-end solution (see Figure 4-2).

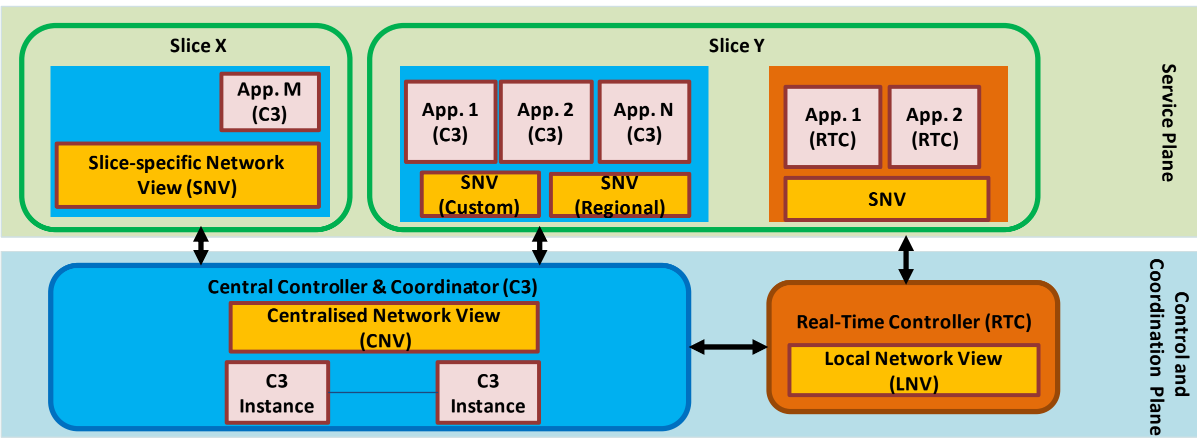

Figure 4-2 Coherent and Network Slicing

As it can be seen from the figure, each slice is composed of a set of control applications running on top of either the C3 or on top of the RTC. This is done in order to account for the fact that time-critical and non-time-critical aspects of a slice shall be handled transparently by the service layer. Each slice control logic can leverage on Slice-Specific Network Views (SNVs) which are derived from the centralised network view (CNV) and from the local network views (LNVs). SNVs can consist either in a subset of the information contained in the CNV and/or LNVs or they can be a slice-specific aggregation of these views.

The COHERENT approach to slicing is aligned with the current trends in NFV, and in particular with the ETSI MANO initiative. In particular, it is our standpoint that the COHERENT is flexible enough to interact with current and future NFV Management and Orchestration Platforms. More details on this aspect will be provided in Section 5.2 where the COHERENT approach to the data-plane functional split is described. Nevertheless, although a rich literature exists on the NFV management and orchestration, most of the works in literature focus on the problem of mapping an input virtual network request (often in the form of a VNF Forwarding Graph) onto a physical virtualised network substrate (often offering computational as well as networking resources). However, these works implicitly assume that once a VNF is mapped on a node, the virtualisation layer (i.e. the hypervisor) will take care of scheduling the various VNFs ensuring both logical isolation and an efficient use of the substrate resources. Such an assumption does not hold anymore if radio nodes are added to the set of virtualised resources available in the substrate network (alongside computational and networking resources). In this case, in fact, the amount of resources available at each substrate radio node is a stochastic quantity depending on both channel fluctuations and end–users distribution.

In [56] we study the VNF placement and scheduling problems in the Radio Access Network (RAN) domain. In the proposed problem formulation which will also be leveraged in COHERENT we expect MVNOs to specify their requests in terms of a VNF Forwarding Graph. Such VNFs can include functions such as load–balancing and firewall, as well as virtual radio nodes. Moreover, in order to satisfy the diverse requirements imposed by future applications and services, MVNOs are allowed to deploy custom resource allocation schemes within their network slice. At the same time, the underlying system shall both enforce strict performance isolation between MVNOs and ensure efficient resource utilisation across the network in spite of the non–deterministic nature of the wireless medium. Such goals are achieved through a dynamic SLA model which takes into account the instantaneous cell capacity for each slice. The preliminary algorithms and proof-of-concept evaluation will be accounted for in D5.1, more details can be found in [76].

5. COHERENT Architecture

In the previous section we described the basic concepts that lie at the foundation of the COHERENT Architecture, namely **control separation** (logically centralised control with C3 and real-time control with RTC), **network abstractions**, and **network slicing with SNV** (slice-specific network view). In this section we shall first see of those concepts fits inside the COHERENT architecture in Section 5.1. Moreover, we will further discuss the control and coordination plane and the user plane in Section 5.2 and Section 5.3, respectively. Finally, the functional blocks responsible for collecting and aggregating the network graph are described in Section 5.4.

5.1 Overview

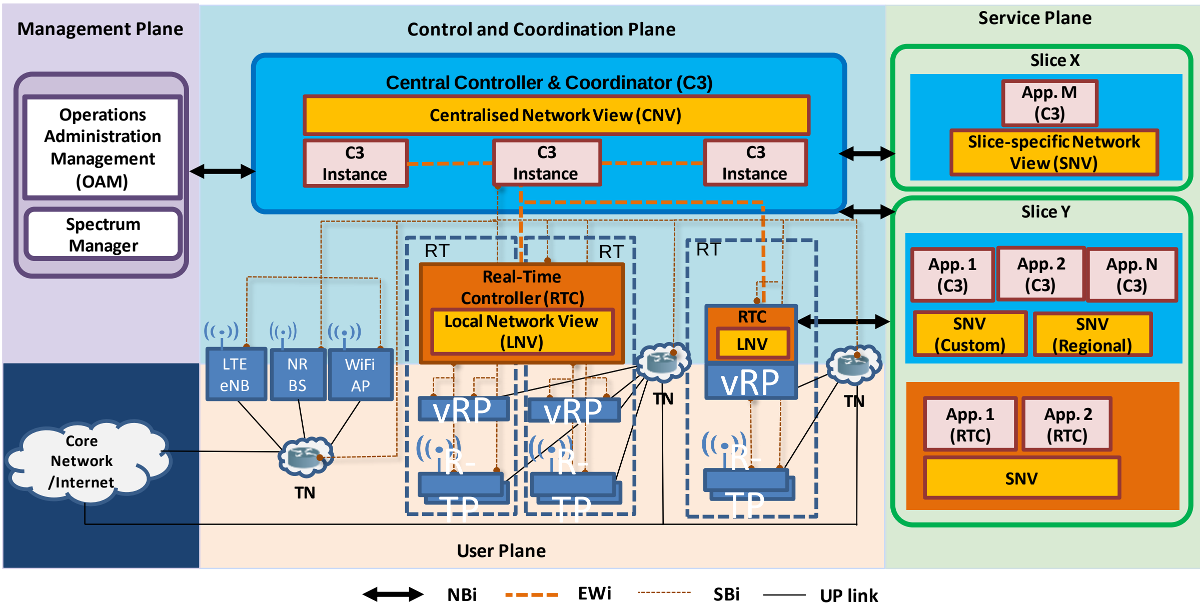

Figure 5-1 COHERENT Architecture

COHERENT architecture provides a programmable control and coordination that offers fine grain, real-time control without sacrificing scalability. ***Scalability*** and ***timeliness*** for control and coordination are achieved by introducing two control mechanisms, namely Central Controller and Coordinator (C3) as well as Real-Time Controller (RTC), as shown in Figure 5-1. Furthermore, the programmability is driven by a key characteristic, namely ***abstraction***. By receiving status reports from low layer entities, C3 maintains a centralised network view of the governed entities, e.g., transport nodes (TNs)[4] and Radio Transceivers (RTs) in the RAN. It is worth stressing that with the acronym RT we address a generic element in the RAN. For example, an RT could be a legacy LTE eNBs or a legacy WiFi AP or a New Radio [5]Base Station (NR BS). Similarly, an RT could be also a vRP with one or more R-TP. The specific applicability of the RT to any of those elements is purely an implementation choice. Nevertheless, Section 5.3 accounts for the most interesting configuration that will be pursued by the COHERENT project.

Based on the Centralised Network View (CNV), the SDN principles are applied in the design of the C3. For overcoming the delay limitation between the C3 and the individual access network elements, latency-sensitive control functionalities are offloaded from the C3 to RTCs. The network entities (TNs and RTs) connect to C3 and RTCs through the southbound interface (SBi). The possible southbound communication protocols are OpenFlow[35], BGP-LS[77], PCEP[78], NETCONF[79],

[4] Note that for simplicity, Figure 5-1 provides a purely logical view of the architecture. In other words, it does not show the deployment cases of TNs and RTs in the user plane. In the C-RAN case, the TNs can be deployed not only in the backhaul but also the fronthaul (between vRPs and R-TPs).

[5] New Radio (NR) is the 3GPP name for new 5G radio technology.

YANG[80], SNMP[81], LISP[82], OVSDB[83], CAPWAP [84] (for WiFi) and LWAPP[85] (for WiFi).

As mentioned before, C3 is a logically centralised control entity because the C3 instances share the network graphs with each other. The communication between controllers (RTC-C3, C3-C3) for sharing network graphs and offloading control functions is through east-west interface (EWi).

As defined in Section 4.1, a network slice in the service plane is defined as a collection of specific network applications and RAT configurations. Different network slices contain different network applications and configuration settings as shown in Figure 5-1. Through northbound interface (NBi), the C3 and/or RTCs provide the required network view, namely slice-specific network view (SNV), for the network service slices so that network service slices could express desired network behaviours (by programming) without being responsible themselves for implementing that behaviour (with hardware). The example of the NBi is REST[86].

Some application modules in network slices may be latency-sensitive. For such a slice, these modules are located in the RTC. Additionally, monitoring modules which are latency-sensitive may need to operate close to the data source for reducing overhead and observe the network at high information granularity. The need for such monitoring modules may be service specific or operation specific. The examples of latency-sensitive network applications are flexible RAN function splitting in Cloud-RAN, MAC scheduling (regular, CoMP, transmission mode selection, etc.), X2 HO decision, MAC/PHY (more generally cell) reconfiguration. In addition, most of the MEC application areas [87] are also relevant to RTC, e.g., localisation, augmented reality, low latency IP service, etc. Moreover, the slice-specific network view (SNV) provided by the C3 could differ from the centralised one in the C3 in terms of space (e.g., it could be limited to a region) but also in terms of aggregation of nodes and edges. In general, we could have different views inside the same slice, according to what the application wants to do.

While the COHERENT control and coordination plane make control decisions for RAN functions and sends the decisions to the network entities for executing the decisions, the management plane usually focuses on monitoring, configuring and maintaining the long-term decisions for network entities in the infrastructure, e.g., evaluation of the LSA rules between the operators, queries the databases managed by the National Regulation Agency (NRA) for the spectrum usage rules in the spectrum manager. The entities in management plane are connected to the C3 through NBi.

5.2 Control and Coordination plane

COHERENT control and coordination plane is shown in the functional architecture as in Figure 5-2. It is a plane proposed in COHERENT to control and coordinate the cooperation and joint resource allocation in heterogeneous RAN. It is an extension of RAN control plane with new functionalities to control and coordinate a group of RTs and TNs. The legacy RAN user and control plane are in RT and UE.

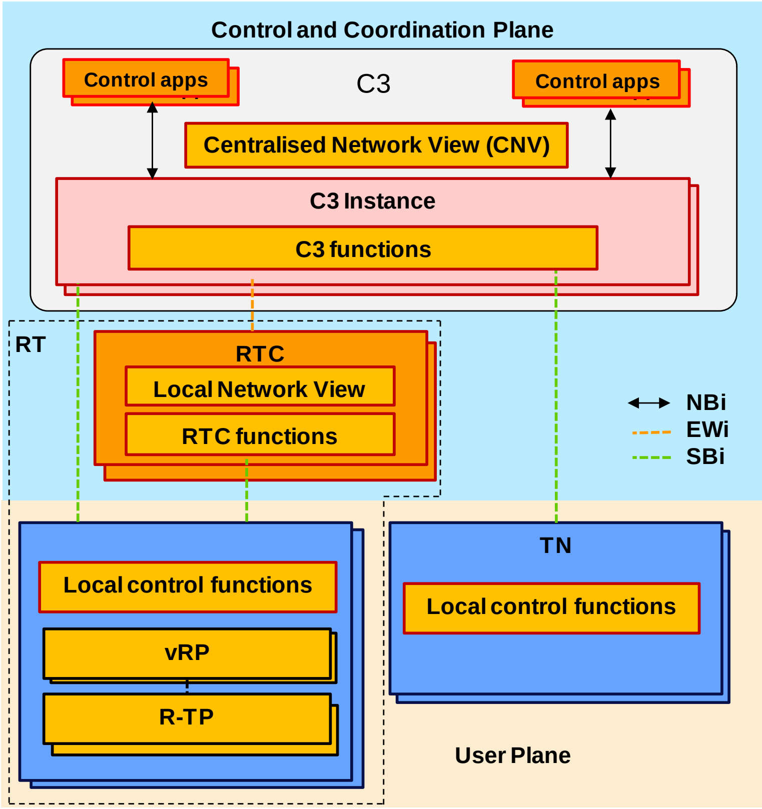

Figure 5-2 COHERENT functional architecture for 5G RAN

The control and coordination plane is comprised by C3 and RTC. C3 is a logical centralised entity. The main function is to orchestrate the behaviours of network entities (RTs and TNs) in the RAN so that network behaviours in the RAN are harmonised. By receiving status reports from low layer entities, it maintains a centralised network view (CNV) of the governed entities in the RAN. Based on the centralised network view, the SDN principles are applied in the design of the C3. The SDN principles in this context include the logical centralised view of network, the separation of control plane and user plane, and the abstraction of low layer information and status for generic, device-independent control.

RTC deals with the fast status update and control decision for R-TP and vRP with the response time in the order of milliseconds.

Inside the control entities, namely C3 and RTC, there are SBi and NBi interfaces. The SBi connects to different radio access technologies (RAT). Control applications are built upon the NBi interface, to perform the high level spectrum management, mobility management, traffic steering and network slicing functions. The C3 will guide the low layer control entities to implement, e.g., RAN sharing and network slicing at the RAN.

For the user plane, the functions of RT can be split into two building blocks: R-TP and vRP. As shown in Figure 5-3, the split is flexible so the R-TP can include only RF, or RF plus Physical layer, or even Physical to MAC layer. Furthermore, some control functions which are tightly bound to the data plane are kept in the RT and TN, namely local control functions. The local control functions could be located in vRP and/or R-TP, depending on the user plane function split between vRP and R-TP.

5.2.1 System Functionalities of RTCs and C3

The functionalities of RTC, C3 and southbound API are summarized as following:

- C3 Functionality/Functionalities:

- Gather information from RTC, network and infrastructure through APIs
- Control different parameters of different networks or infrastructure entities (Multi-RAT, switches) through APIs
- Upgrade/Create Network Graphs among control instances
- Control-delegation mechanisms between RTC and C3
- Provide slice-specific network graph to network applications through API.
- Collect/coordinate configurations/policies from network applications.

• RTC Functionality/Functionalities:

- Gather information from network and infrastructure through APIs
- Control different parameters of different network or infrastructure entities (Multi-RAT, switches) through APIs
- Provide local network graph to network applications through API.
- Collect/coordinate configurations/policies from network applications
- Control-delegation mechanisms between RTC and the network devices, e.g., eNB, vRP, R-TP.

• Southbound API Functionality/Functionality:

- To provide transmission of relevant parameters between the controller and relevant network and infrastructure entities;
- To provide configuration of relevant network and infrastructure entities.

5.3 User Plane

In a traditional mobile network, the radio and base–band processing units, which compose base stations, are placed in close proximity. Historically it is also good to recall that passing from UMTS (3G) to HSPDA (3.5G) cellular systems, dealing with L2 processing at the base station close to the antennas allowed, for instance to introduce fast hybrid ARQ responses, and fast link adaptation. Another reason for this close proximity between antenna and baseband processing is to mitigate the high signal losses associated with the RF cables that are typically used for their interconnection. In order to circumvent these limitations operators moved to the Distributed RAN architecture (D–RAN), where RF cables are replaced with optical fibres and a digital interface is used to carry the IQ (in–phase/quadrature) signals between the base band units (BBU) and the radio elements, named Remote Radio Head (RRH). Cloud RAN (C–RAN) has recently emerged as a solution capable of reducing the deployment and operational costs of mobile networks while at the same time enhancing network capacity, coverage and power consumption. C-RAN[60] achieves such goals by consolidating BBUs in large high–volume computing infrastructure, named BBU Pools, and by sharing them among multiple sites.

Nevertheless, although the link between RRHs and BBU Pool has been used as demarcation point in the C–RAN architecture, other RAN functional splits can be in principle defined as shown in Figure 5-3, each of them coming with different requirements.

In the RAN function split, the R-TP (Radio Transmission Point) is the actual point of attachment for wireless terminal. Full or partial RAN node functions are implemented in R-TP, while rest of functions are virtualised and offloaded to the vRP (Virtual Radio Processing). In the case of a C-RAN architecture R-TP coincides with RRH while vRP coincides with BBU Pools. Similarly, vRP can be implemented with software and deployed in high volume computing infrastructure and/or distributed micro-clouds. In general, the lower the RAN functional split is executed within the RAN protocol stack, the higher is the centralisation benefits. However, the fronthaul requirements become also more stringent. For example, C–RAN enables an MNOs to implement Coordinated Multi-Point (CoMP) transmission and reception, while a split above the Medium Access Control (MAC) layer cannot support CoMP only allowing higher–layers cooperation features (e.g., joint scheduling).

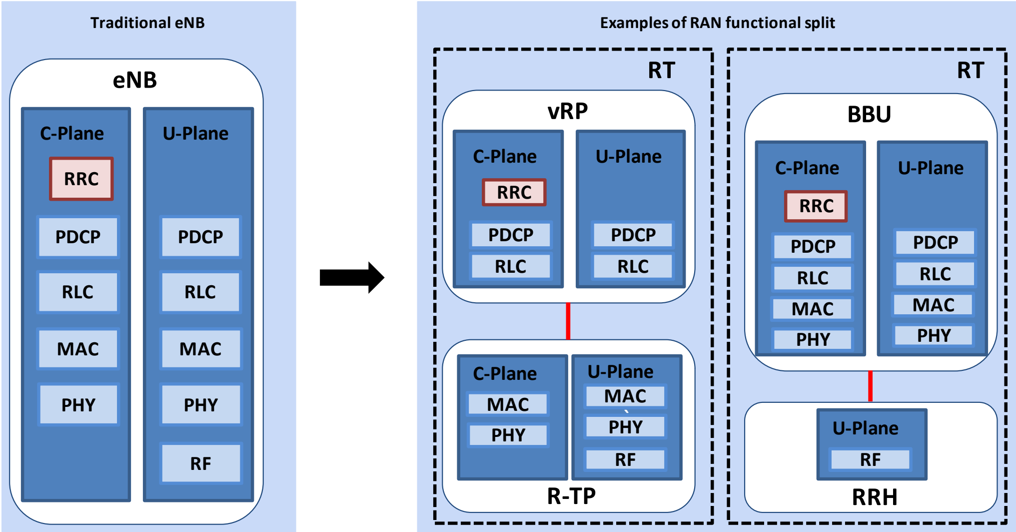

Figure 5-3 Possible RAN functional splits in LTE RAN without the separation of control and user plane

Figure 5-4 shows how RAN functional split is applied to the user plane of the COHERENT architecture. As seen in Figure 5-4, the control plane has been offloaded from RT to C3 and RTC as COHERENT adopts the concept of separating control and user planes in SDN. For the user plane, the functions of RT are split into two building blocks: The R-TP (implementing PHY and MAC layers) and vRP (implementing RLC and PDCP layers). When the vRP is centralised for multiple R-TPs, the C-RAN liked architecture and process are supported. So the vRP is the place for NFV in the user plane of the RAN. Otherwise if the vRP and R-TP are collocated, it is a standalone base station. Notice how one vRP is responsible for implementing the virtualised radio processing of multiple R-TP, thus allowing significant statistical multiplexing gains. Finally, Figure 5-4 also illustrates a particular configuration for the RTC: in particular one instance of the RTC is responsible for controlling the *low-latency functions* of multiple vRP as well as of multiple R-TP. Similar considerations can be made for the C3 entity, although for *non-real-time functions*.

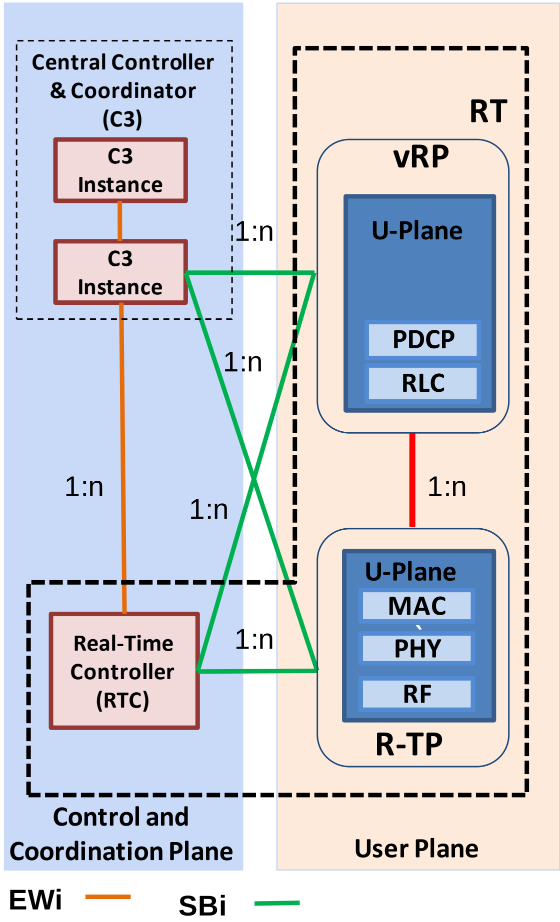

Figure 5-4 The functional splits in the user plane of COHERENT architecture (LTE case)

Similarly, in Figure 5-5 we depict the functional split addressed in COHERENT with regard to the WiFi radio access nodes. As it can be seen in this case part of the MAC as well as the entire PHY are maintained within the R-TP. This is due to the fact that it would make very little sense to introduce C-RAN like architecture in the case of a cheap and commodity technology like WiFi. In particular low-latency MAC features such as acknowledgments and rate adaptation are kept within the R-TP, while non latency sensitive features like wireless client scheduling and in general management frame handling are consolidated within the vRP. More details about the benefits of such choice are reported in D5.1 were preliminary evaluation results for the WiFi use case are reported.

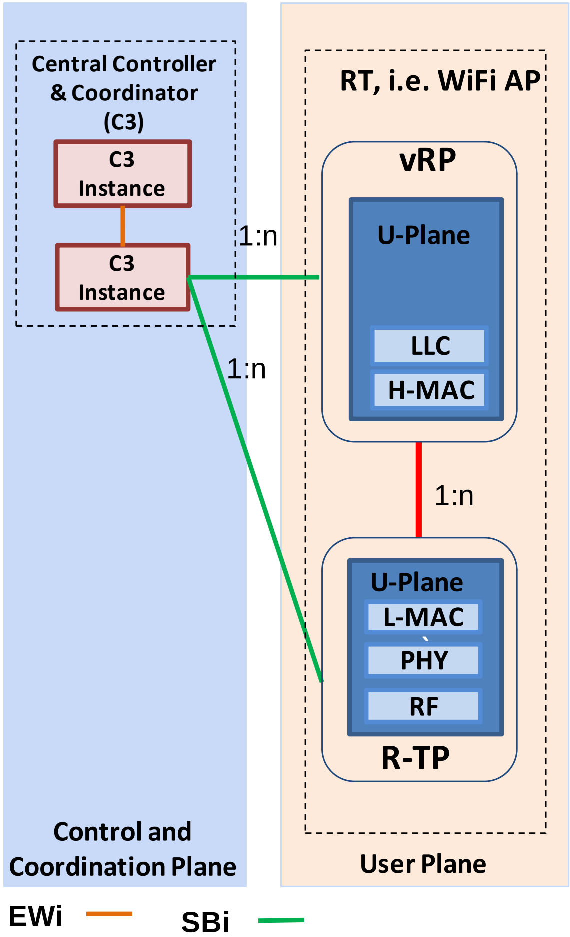

Figure 5-5 The functional splits in the user plane of COHERENT architecture (WiFi case)

Notice that the ones depicted in Figure 5-4 and Figure 5-5 are just some of the ways to split RAN functionalities in the user plane of COHERENT. However, COHERENT architecture could support flexible network function split. More specifically, any network function split cases can be applied to the COHERENT architecture.

5.4 Functional blocks for Processing Network Information

In this section we describe the functional blocks necessary for implementing the COHERENT network abstraction concept (as shown in Figure 5-6). Further, we describe the functionality of the blocks at infrastructure and control level. The architectural concept of network graphs and processing of network information targets mainly three design goals:

- Programmable network observability, meaning that the architecture should allow for dynamic deployment of software for retrieving information from a physical or logical infrastructure.
- Distributed/decentralised processing for the purpose of supporting controller and management actions at various time scales and with low processing and signalling overhead.
- Deployment of flexible, extendable and generic software for building network graphs representing the network state needed for different applications and RATs.

The functional blocks as described in Figure 5-7 are aimed to support monitoring, control and management of the infrastructure as well as services (or slices) deployed across RANs. In the former case, instances of the functional blocks can be instantiated for monitoring available resources and performance of the physical infrastructure for the purpose of supporting service (re-)deployment. In the latter case, instances of the functional blocks can be deployed as part of a service for the purpose of providing monitoring information specific to a slice.

In the COHERENT architecture, data is generally stored in a distributed or decentralised manner. For short-term usage, data is mainly stored close to the data source (i.e. in the nodes or dedicated storage), whereas data for long-term usage is maintained by appropriate distributed database management systems (e.g., CASSANDRA [88]). The creation and processing of a network graph is based on retrieving relevant information from existing short-term and long-term data storage using suitable tools (e.g., Neo4j [89]) chosen by the service provider or operator.

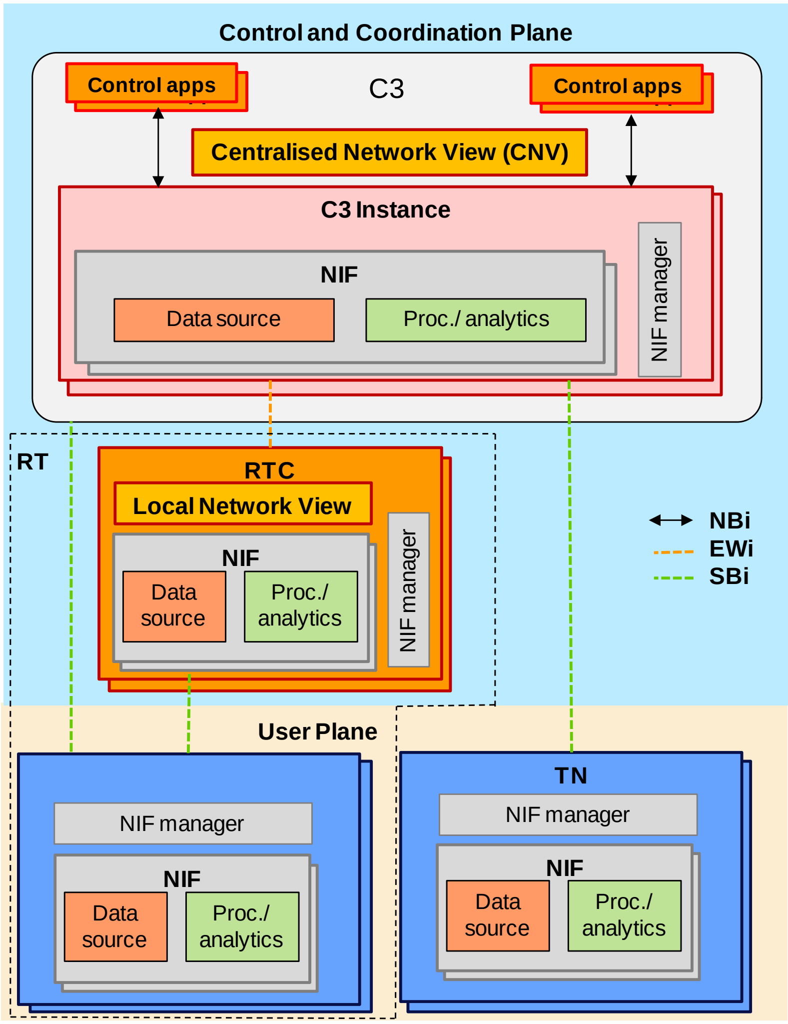

Figure 5-6 The functional blocks implementing the COHERENT network abstraction concept

Functional blocks:

- Network Information Function (NIF): A NIF implements functionality for accessing data sources and processing of data. A NIF can gather information from other NIFs, e.g., for the purpose of creating a local network view in terms of a network graph. A NIF can in its simplest form consist of one data source that corresponds to one metric (e.g., signal strength). A NIF require storage, either local (e.g., shared memory) or external (e.g., storage network).
- NIF manager: Provides information about which NIFs are running, together with managing configuration and forwarding of information to and from NIFs. The manager keeps a registry of available NIFs and also manages deployment of NIFs (supporting programmability). The NIF manager supports synchronisation and exchange of network information and states between entities in the infrastructure and controller planes.

- A NIF data source encompasses measurements, logs, counter readouts, neighbour network topology, vendor-specific information about the infrastructure as well as information retrieved from external data sources. A data source can be used by one or several NIFs.
- A NIF data processing block includes aggregation and light-weight data analytics or other types of processing for the purpose of deriving metrics and abstractions.

Infrastructure level:

- At infrastructure level NIF implements monitoring functionality and functions for accessing node-local performance and state information.
- An infrastructure node entity (e.g., eNB) implements a NIF manager along with relevant NIFs in order to maintain local data structures (e.g., a partial network graph, individual metrics, topological information etc.) reflecting the local network conditions as perceived by the node (e.g., signal strength from broadcasting neighbouring RTs).

Control level:

- RTCs implement functionality for accessing data sources as well as processing and analytics for the purpose of providing a regional and local network view. The data processing functional block includes functionality for creating, storing and maintaining network graphs but also other types of information (e.g., metrics, states, topology information at RTC level or requested from a certain RT). NIFs implement functionality for retrieving and processing information available in the infrastructure entities controlled by the RTC. The RTC can use NIFs for monitoring, measurements and information retrieval for creating, processing and maintaining network graphs and other types of information (e.g., logs, parameter configuration, states or metrics). The output from a NIF constitutes a Local Network View. The NIF manager keeps track of the NIF instances and manages information requests between the RTC, the C3 and the infrastructure entities.
- C3s implement functionality for accessing data sources readily available at the controller and provided by RTCs or neighbouring C3s. Each C3 instance maintains a local network view (via the NIF manager and deployed NIFs) based on the input from RTCs and infrastructure entities. In order to provide a centralised network view, a controller network is needed for exchange and synchronisation of network information and states.

The ways that information is handled in the RTC and C3 instances are different with respect to consistency and the scope of control. An RTC keeps information consistent within its regional and real-time scope of control, meaning that it has to synchronise information from various infrastructure sources to produce a complete local network view. At C3 level, synchronisation and information exchange includes signalling between C3-to-infrastructure, C3-to-RTCs and C3-to-C3.

Finally, information exchange (i.e. signalling, transmission of monitoring information, states, etc.) between network entities maintaining NIFs is done via: i) SBi as well as NBi, and ii) existing communication protocols between infrastructure entities (e.g., X2). The communication (retrieval of network information and control messages) between RTC/C3 and NIF manager in RT and TN is done through a logical interface SBi that is specific for a RAT. Moreover, access to data sources (i.e. data-path monitoring and measurements, metrics, state variables, etc.) on the nodes is implementation-specific (e.g., sockets, shared memory and processes).

6. Support of Use Cases with COHERENT Architecture

A detailed description of COHERENT's reference scenarios has been presented in D2.1 of COHERENT. In this section, we show how the COHERENT architecture can support a number of different use cases. Furthermore, the use cases of spectrum management using COHERENT architecture are presented in D4.1 of COHERENT in details.

The use cases discussed in this section are general and include multiple specific use cases described in D2.1. In more detail we provide a mapping of the D2.1 Use Cases that we plan to explore in the context of the project, to the following five categories:

- Multi-Tenancy
- Device-to-Device Communications
- Broadcast Operation
- UE-Relaying Operation for Public Safety Service
- Multi Connectivity

Note that towards 5G communications, mobility management mechanics are expected to be increasingly important, while both the control plane and data plane procedures are involved in various performance optimizations. Furthermore, mobility management is orthogonal to the Use Case categories we propose and can be actually part of each one of them. For example, we consider for mobility management issues both under Multi-tenancy but also under multi-connectivity scenarios.
Therefore, the mapping for the mobility management is not provided in the table. Furthermore, before we begin the description of each Use Case category mentioned above, we firstly provide a description on the COHERENT approach in mobility management in Section 6.1, focusing on the X2 handover.

Table 6-1 Mapping of use cases in D2.1 to the examples of general use cases which could be supported by COHERENT architecture

Use Cases for COHERENT Architecture in D2.2	Mapping to D2.1 use cases
Section 6.2 Multi-Tenancy	**Use Case 1.RS**: RAN sharing among heterogeneous mobile networks
	Use Case 1.SR: Supporting RAT sharing
	Use Case 1.CO: Cooperation among multi-operators
	Use case 3.FS: Flexible resource sharing for broadband PMR networks
	Use case 4.MV: Enhanced Mobile Virtual Network Operator (MVNO) for PMR services
	Use case 4.GD: User groups' differentiation in multimedia service provision
	Use case 4.SD: Service differentiation
Section 6.3 Device-to-Device Communications	**Use case 3.MN**: Coordination of rapidly deployable mesh networks
	Use case 3.CE: Coverage extension and support of out-of-coverage communications (D2D communications)
Section **Error! Reference source not found.** Broadcast Operation	**Use case 3.CE**: Coverage extension and support of out-of-coverage communications (D2D communications)
	Use case 4.MV: Enhanced Mobile Virtual Network Operator (MVNO) for PMR services
	Use case 4.eM: Dynamic eMBMS for public safety applications
Section 6.5 UE-Relaying Operation for Public Safety Service	**Use case 3.MN**: Coordination of rapidly deployable mesh networks
	Use case 3.CE: Coverage extension and support of out-of-coverage communications (D2D communications)
	Use case 4.MV: Enhanced Mobile Virtual Network Operator (MVNO) for PMR services
	Use case 4.GD: User groups' differentiation in multimedia service provision

	Use case 6.AG: Air-to-ground communications
Section 6.6 Multi Connectivity	**Use Case 1.CO**: Cooperation among multi-operators
	Use Case 2.MM: Massive MIMO / distributed antenna system in dense small cell deployments
	Use Case 2.FSA: Flexible spectrum access
	Use case 3.FS: Flexible resource sharing for broadband PMR networks
	Use case 4.MV: Enhanced Mobile Virtual Network Operator (MVNO) for PMR services
	Use case 6.MR: Delivery of services in public or private transportation in urban areas

6.1 Mobility Management

Efficient handoff mechanisms are essential for ensuring seamless connectivity and uninterrupted service delivery. Handover procedure in LTE/LTE-A has been radically evolved when compared to the previous 3GPP standards. In particular, X2 handover is introduced to allow neighbouring eNBs to handle the user mobility without the involvement of the core network. While most of the application could considerably benefit from the X2 handover performance improvement, delay breakdown and impact of parameters from the UE perspective are not well investigated.

In COHERENT we believe that the impact of user mobility on new 5G designs, small cell networks, Ultra dense Networks (UDN) and Heterogeneous Networks (HetNets) will be increasingly important.

The X2 topology as well as the X2- AP structure provide advantages related to the data forwarding operation.

Overview of the X2 Procedure

The X2 procedure can be described in five steps (see [90] for more details):

- **Before Handover:** UE is attached to the source eNB. The Dedicated Radio Bearers (DRBs) and Signalling Radio Bearers (SRBs) are established and UL/DL traffic is transmitted between the source eNB and the UE. The UE remains in the Radio Resource Control (RRC)-Connected, EMM-Registered, and ECM-connected states with respect to the source eNB, and keeps all the resources allocated by E-UTRAN and EPC)

- **Handover Preparation:** UE sends the periodical measurement report to the source eNB; this report contains information about the neighbouring cells. The source eNB triggers the handover (i.e., eNB decides that the handover is necessary) based on the reported measurement results, i.e., A1-A5/B1, B2 event (see [91]) and chooses the best reported target cell by the UE. Then, the source eNB sends a X2 handover request to the target eNB. This message contains the information needed to perform the handover (e.g., UE context information, Radio Access Bearer (RAB) context, Target Cell ID). Considering the QoS in the RAB context, the target eNB performs call admission control and if it is able to provide the requested resources for the new UE, it sends a handover (HO) request acknowledgment (ACK) to the source through the X2 direct tunnel setup (i.e., handover is eNB accepted). The source eNB receives this message that includes the configuration of the GTP-U tunnels per radio access data radio bearer as well as the RRC Connection Reconfiguration message in a transparent container that the source eNB has to forward to the UE. In the RRC message, L1/L2 parameters are provided to the UE in order to be synchronised with the target eNB. Finally, the source eNB sends the HO command message that encloses the RRC Connection Reconfiguration message to the UE. If the target eNB cannot accept the HO request (due to load or the required setup), it responds to the source eNB with an X2 failure message. During this step, the UE states remain unchanged.

- **Handover execution:** UE receives the RRC Connection Reconfiguration message and transits to the RRC idle state triggering the detachment from the source eNB. The source eNB sends the

Sequence Number (SN) status transfer message that contains the Packet Data Convergence Protocol (PDCP) sequence numbers to the target eNB through X2 interface. For UL the first missing data unit is included and for DL the next sequence number to be allocated. Then, UE is synchronised with the target based on the given parameters and send the HO Confirm message that encloses the RRC Connection Reconfiguration Complete to acknowledge the successful handover to the target eNB. As a result, the UE transits to the RRC connected state with respect to the target eNB. Concerning the UE synchronisation, if a dedicated random access preamble has been received in the RRC Connection Reconfiguration message, the UE does not need to perform the random access procedure, i.e., contention free Random Access Channel (RACH) process. If this is not the case, the UE performs the normal random access procedure described in [9] (contention based RACH).

- **Handover Completion:** The target eNB receives the RRC Connection Reconfiguration Complete message and the path switch procedure is initiated between the target eNB and the MME/S-GW. The target eNB starts to forward all the packets received from the X2 interface to the UE before any new ones coming from the Serving Gateway (S-GW) (i.e., target eNB receives the end-marker from the old path switch and starts transmitting packets from the new path switch). Afterwards, the source eNB UE context is released via receiving UE release context message from the target eNB. Finally, the S1 bearer that was initially established between source eNB and UE is also released.

- **After Handover:** UE is attached to the target eNB. The DRB and SRB are established and UL/DL traffic is transmitted as in the initial step

In general, X2 handover can be initiated by the eNB for several reasons: 1) Quality-based handover: The indicated QoS levels included in the measurement report by the UE are too low and the UE needs to switch to another eNB for enhancing its QoS metrics. 2) Coverage-based handover: UE is moving from the one cell to another. In general, it could be intra-LTE or from LTE to UMTS or Global System for Mobile communication (GSM), when the UE moves to an area that is not LTE-covered, i.e., inter-Radio Access Technology (interRAT). 3) Load-based handover: This is an optimisation case concerning the load among different eNBs. The required information is transferred through the X2 load indication message.

The COHERENT Approach

The idea is that the COHERENT framework can be used to facilitate the handover procedures and programmatically act reactively (or proactively) to user mobility events. As C3 instances create the centralised network view and therefore logical centralised control and coordination can be achieved, actually the handover decision can be more efficient based on knowledge retrieved by the Network Graphs. All the necessary information regarding the network configuration and operation state is available through fast and resource-effective retrieval and processing of network metrics and monitoring information.

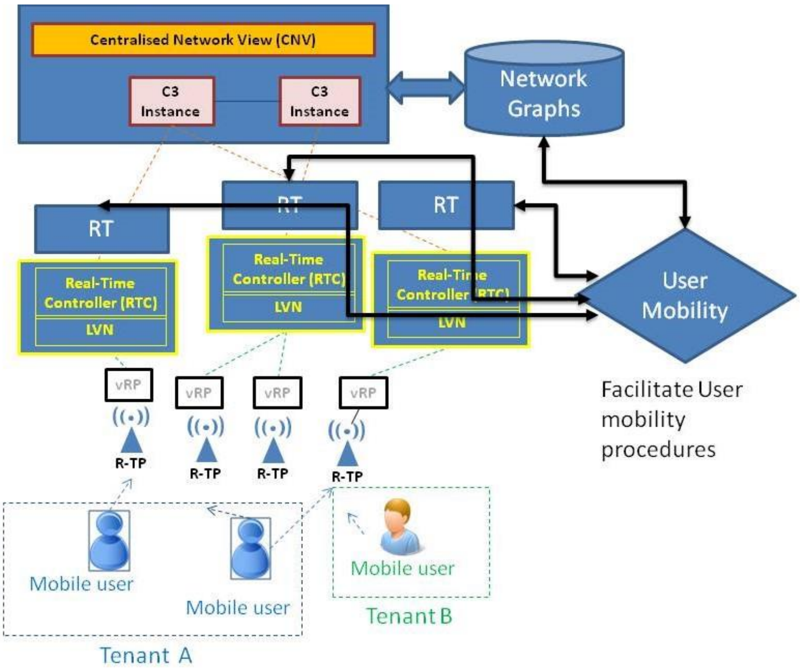

Figure 6-1 Coherent architecture and User Mobility

In COHERENT we study all the LTE related interfaces that are used for the mobility management using the X2 handover mechanism. Note some of functions in C3 are distributed in nature, and that the state/context is shared across all the instances.

Handover Criteria and Parameterisation

Various metrics are studied like delay. Note that handover delay can be classified into two different main categories: the protocol delay that captures the processing time and handover signalling delay and the transport delay that captures the transmission time through the physical medium of the X2 link (wired or wireless).

Another well-known handover criterion, commonly used in conventional HO decision algorithms for mobile communication systems (also applied in 3GPP LTE), is based on RSRPs comparison method in which hysteresis and handover offsets are included. According to the 3GPP standard we have the following definitions:

- **Received Signal Strength Indicator (RSSI):** E-UTRA Received Signal Strength Indicator (RSSI), comprises the linear average of the total received power (in [W]) observed only in the configured OFDM symbols and in the measurement bandwidth over N number of resource blocks, by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise etc.
- **Reference Signal Received Quality (RSRQ):** Reference Signal Received Quality (RSRQ) is defined as the ratio N×RSRP/(E-UTRA carrier RSSI), where N is the number of RB’s of the E-UTRA carrier RSSI measurement bandwidth. The measurements in the numerator and denominator shall be made over the same set of resource blocks.

Other representative handover algorithms include not only Received Signal Strength (RSS) criteria; a brief description of the non-RSS HO algorithms is given as follows:

- **Interference-based:** Interference-aware handover decision algorithms enable the shifting to femtocell communication paradigm in HetNets, where co-tier and cross-tier interference is taking into account based on interference level at the cell sites or Received Signal Quality (RSQ) at the UEs.

- **Speed-based:** Speed handover decision algorithms typically compare the UE speed with specific thresholds to mitigate the HO probability for high speed users (i.e., fast handover case) decreasing the overall handover signalling cost. Such algorithms can be combined with load/traffic-type criteria that are discussed below.

- **Load-based:** To this direction, load-aware handover decision algorithms can be developed considering the service delay that a user experiences from the network. In addition, an implementable framework based on Software Defined Networking (SDN) architecture can be included to support the algorithm, as suggested to be a key enabler for the realisation of 5G networks. This approach overcomes the shortcomings created by only considering Received Signal Strength (RSS) criteria in HO decision for HetNets.

All these handover criteria can be used to improve the handover decision based on various performance objectives. The efficiency of the approach will be demonstrated during WP3 activities related to user mobility.

6.2 Multi-Tenancy

Network slicing has evolved as a fundamental feature of the emerging 5G systems enabling dynamic multi-service support, multi-tenancy and the integration means for vertical market players. In high level the Network slicing concept is about operating virtual networks on top of physical infrastructures, with virtual resource isolation and virtual network performance guaranties. The separation of different functions by abstractions (e.g., radio resources from packet processing) simplifies the integration challenges especially for applications supporting vertical industries beyond telecommunications. Note that the notion of resources in 5G network slicing includes network, compute and storage capacity resources; virtualised network functions; shared physical resources; and radio resources.

In COHERENT Network slicing will be performed by using the abstraction mechanisms over different physical infrastructures into a logical network. This will contain shared resources, such as radio spectrum or dedicated core network equipment, and virtual network functions obtained by breaking down single physical equipment into multiple instances. The Network slice fully supports a particular communication service exploiting the principles of SDN and NFV in order to fulfil a number of business requirements for various stakeholders like MVNOs and OTT the players. Our goal in CHERENT is to a) Drastically transform the networking perspective by abstracting, isolating and separating logical network behaviours from the underlying physical network resources and b) Exploit network slicing for reducing CAPEX and OPEX for operators, allowing also programmability and innovation. The creation and management of network slicing is a challenging process that poses new problems in service instantiation and orchestration, resource allocation/sharing and assignment procedures as well as network virtualisation technologies.

In COHERENT we consider the following three aspects of the multi-tenancy with respect to the Network Slicing design paradigm [92][93]:

- **Infrastructure Sharing:** although cloud computing can be used to meet traditional challenges, like scalability concerns and provide for fast resource provisioning times, a multifaceted analysis is required when it comes in multi-operator environments with time-critical applications and services Towards building infrastructure sharing mechanisms for LTE the evolutionary approach applies the SDN concept into a part of the traditional core network architecture. This evolutionary approach analyses the traditional mobile network functions, such as PGW, SGW, MME, etc., and decides which functions should be implemented to the controller, and which should be implemented in the traditional dedicated hardware. Work in this direction focuses on slicing techniques to create multiple virtual core networks shared among multiple network operators.

- **Spectrum Sharing:** Dynamic spectrum access is a promising approach to increase spectrum efficiency and alleviate spectrum scarcity. Many works like [94] investigate issues like cooperative spectrum sharing under incomplete information. Other approaches like [95] are both incentive-compatible and individually rational, and are used to determine the assigned frequency bands and prices for them. The idea is to find policies that guarantee the largest expected profits by selling frequency bands jointly.

- **RAN Sharing:** This class relies on Hypervisor-based solutions to create the virtual eNB, which uses the physical infrastructure and resources of another eNB, depending on requests from the MNO.

One important use case, also in the context of COHERENT, is RAN sharing implemented by adding advanced control feature to the current 3GPP LTE. Before describing the RAN sharing case, few words will be spent just below to quickly introduce LTE PHY layer and its representation of time-frequency resources. This information is in fact mandatory to describe our use case on RAN sharing at layer 2.

RAN Sharing Use cases in LTE

One important use case, also in the context of COHERENT, is RAN sharing implemented by adding advanced control feature to the current 3GPP LTE. Before describing the RAN sharing case, few words will be spent just below to quickly introduce LTE PHY layer and its representation of time-frequency resources. This information is in fact mandatory to describe our use case on RAN sharing at layer 2.
LTE uses Orthogonal Frequency Division Multiplexing (OFDM) for the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) in the uplink, while it allows bandwidth ranging from 1.4 MHz up to 20 MHz and operates in both paired and unpaired spectrum by supporting both Frequency-Division Duplex (FDD) and Time-Division Duplex (TDD). OFDM divides the available bandwidth into multiple narrower sub-carries and transmits the data on these carries in parallel streams. Each subcarrier is modulated using different levels of modulation, e.g., QPSK, QAM, 64QAM. Note that an OFDM symbol is obtained by adding the modulated subcarrier signals.

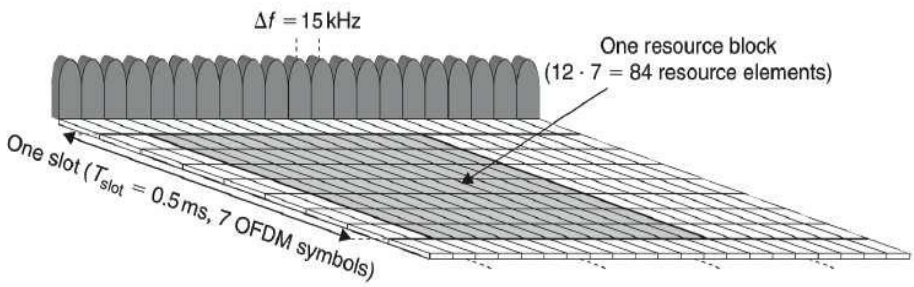

Figure 6-2 Each RB has 12x7 = 84 resource elements in the case of normal cyclic prefix and 12x6 = 72 resource elements in the case of extended cyclic prefix[96]

In LTE downlink transmissions are grouped in (radio) frame of length 10 ms, one radio frame is formed of 10 subframes of 1ms duration and there are ten subframes in the uplink and ten frames in the downlink. Each subframe is divided into two slots of 0.5 ms duration. Each slot counts 6 or 7 OFDM symbols for normal or extended cyclic prefix used. The Physical Resource Block (PRB) is the smallest element assigned by the base station scheduler. Transmission Time Interval (TTI) is the duration of a transmission on the radio link. The TTI is related to the size of the data blocks passed from the higher network layers to the radio link layer. A scheduler can determine to which user the shared resources (time and frequencies) for each TTI (1 ms) should be allocated.

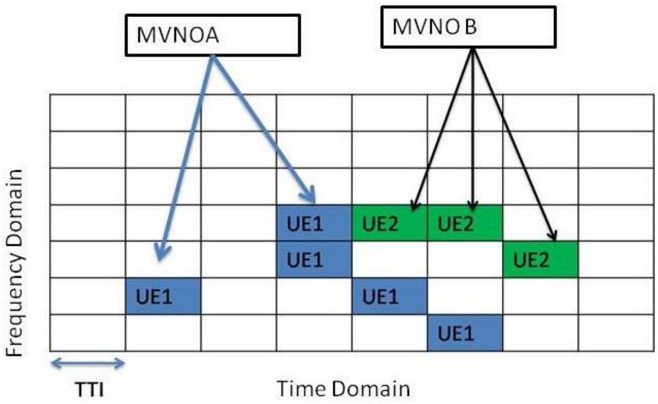

Figure 6-3 RB Scheduling

Our activities in the RAN sharing problem are related to the design and implementation of policies that are able to effectively schedule Resource Blocks effectively between different MVNOs with respect to specific differentiation objectives and with isolation guarantees.

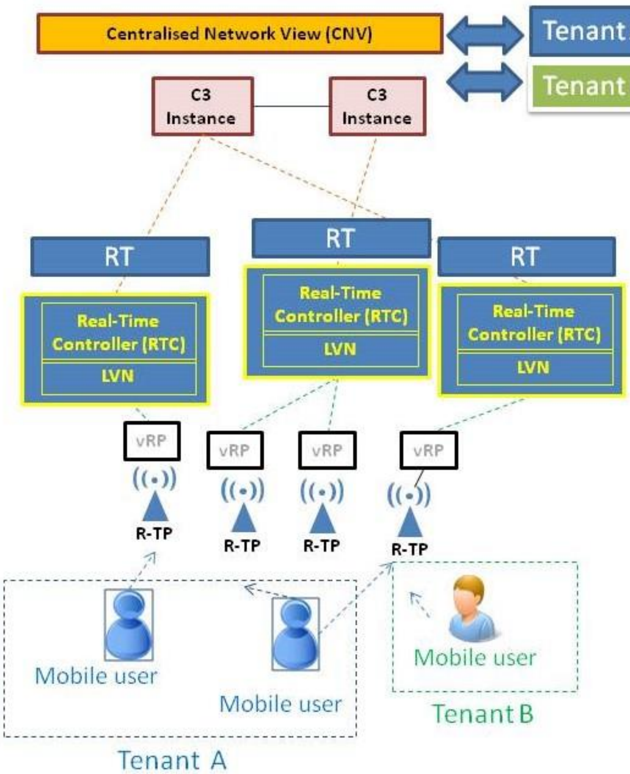

Figure 6-4 Multi-tenancy operation and RAN Sharing in COHERENT

Mapping of COHERENT architecture to use cases related to MVNOs operation under multi-tenancy and network sharing, will be realised in the form of algorithms built and implemented over the OAI infrastructure in the context of WP5 activities. As presented in Figure 6-4 a high level representation is presented of the COHERENT control plane interaction with different tenants, under the RAN sharing use case.

The RTC is the component of the COHERENT architecture that will be responsible to facilitate the efficient operation of policies that will be able to support the RAN sharing concepts. We remind that an RTC keeps information consistent within its regional and real-time scope of control, meaning that it has to synchronise information from various infrastructure sources to produce a complete local network view. Together with the C3 integration and a fully programmable underlay, in COHERENT we are able to support the Infrastructure Sharing, the Spectrum sharing and the RAN sharing concepts using a unified control and management solution.

6.3 Device-to-Device Communications

D2D communications is a promising technology to provide energy-efficient, high data rate and low latency services for end-users in the future 5G wireless networks. Moreover, D2D communications have a high potential for supporting new specific use cases, for instance linked to V2X or public safety applications, as we will briefly mention later on. COHERENT has identified a use case called UC3.CE in D2.1, which is directly related to D2D communication. Use case UC3.MN about coordination of rapidly deployable mesh networks can also benefit from D2D communications, as well as use cases related to supporting communications in a crisis scenario or during a natural disaster. We also define a set of requirements in Table 6-2, which need to be fulfilled by the COHERENT architecture to properly support these use cases in D2.1. D2D discovery and communication are the new features to be studied in 3GPP Rel-12 and Rel-13 under the Proximity Services (ProSe) study item. D2D technology will help to boost new kinds of applications and services for end users on top of the D2D-enabled wireless networks, e.g., proximity-aware social networking service (SNS), D2D advertisement dissemination, public safety network, V2X, Multimedia Broadcast Multicast Service (MBMS), etc. To support D2D discovery, D2D communication and these D2D applications, conventional cellular or WiFi network architectures need to be modified and optimized. To support D2D, 3GPP Rel-12 has introduced two new network elements (the ProSe Function and the ProSe App Server) and some new network interfaces (PC1 - PC6) into the LTE Evolved Packet System (EPS) [27], as shown in Figure 6-5. Some solutions for signal design, radio resource allocation for both D2D discovery and D2D transmission have also be identified in [97]. Traditional RRM functions in LTE cover the radio bearer control, radio admission control, radio mobility control, scheduling and dynamic allocation of resources to UEs for uplink and downlink. New RRM functions are required for D2D link in future networks. Those new RRM functions are critical for interference control for multiple D2D one-to-one and/or one-to-many communications among users located in the same or adjacent cells.

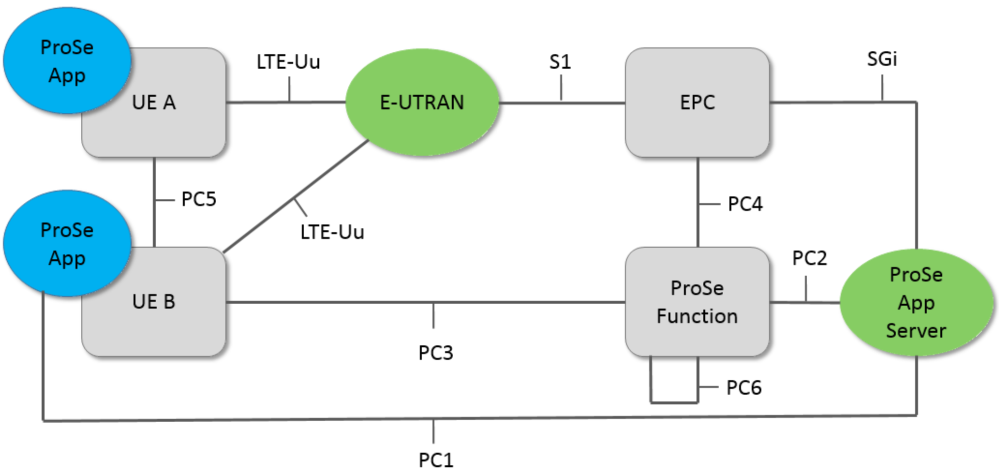

Figure 6-5 Network Interfaces and Components to support D2D in 3GPP Rel-12

Table 6-2 The D2D-related use case in D2.1 and some identified requirements

UC3.CE	Coverage extension and support of out-of-coverage communications (D2D communications)
R#60	COHERENT shall manage multiple-hop relaying in a transparent way for the final user
R#61	COHERENT shall guarantee high reliability and acceptably low latency also for multi-hop communications and D2D communications (e.g., for PMR services)
R#64	COHERENT shall allow a more intense reuse of the spectrum in dense areas and so increasing cell spectral efficiency and improving cell coverage

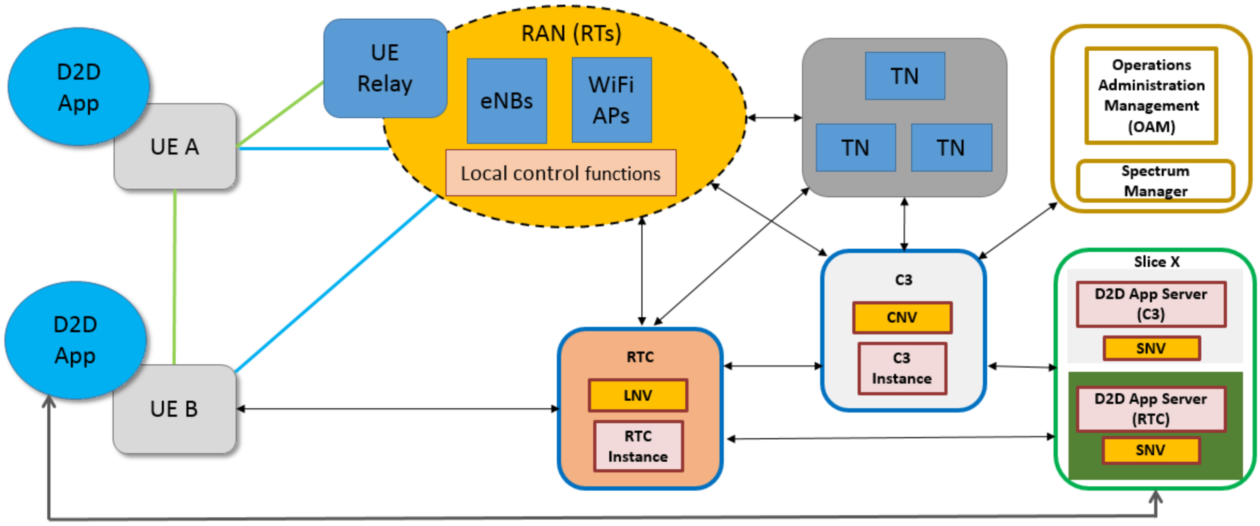

Figure 6-6 COHERENT Architecture to support D2D

6.3.1 Relation with the COHERENT Architecture

D2D communication and the related applications require a variety of control functions from the network architecture. Some D2D applications require low latency and are time-critical, thus they have high-priority QCIs. Different D2D applications should use different network slices according to their QCIs, as shown in Figure 6-6.

- **Critical D2D applications and services**: public safety communications, V2V communications, real-time games, virtual reality office etc.
- **Non-Critical applications and services**: proximity-based SNS, advertisement dissemination, D2D caching, local file sharing etc.

The control functions for D2D can be divided into two classes: Real-time control functions and Semi-static control functions. In COHERENT, real-time control functions are performed by RTC or the RTs. Some Semi-static control functions such as traffic offloading and balancing can be performed in C3.

- **Real-time control functions for D2D**: D2D channel measurement and reporting, interference coordination for multiple D2D links in the same region, radio resource allocation etc.
- **Semi-static control functions for D2D**: cell association and handover, D2D discovery, traffic offloading, traffic balancing using D2D relaying, routing and relay selection for multihop D2D etc.

The exchange of interference information is for interference coordination, and the frequency of exchange is rather high (in the order of tens of milliseconds) making D2D interference coordination a typical real-time task. When scheduling a D2D transmission, the information of interference edges is considered when performing resource allocation. The potential interference (interference spillage) is modelled as a function of interference edge weights, traffic load of transmission link. Using this function, we can attach a price for each D2D link, with higher price means larger interference spillage.

When doing D2D scheduling, only local graph information related to the D2D pair is needed, which reduces the controlling overhead. The RTC is responsible for construct the local interference graph by collecting the D2D measurements and conduct the interference coordination in a fast and distributed way.

6.4 Broadcast Operation

The broadcast scenarios in the section are mainly for broadcasting in the cellular networks (3GPP) as described by Section 2.3.1.3. Note that this does not exclude the broadcasting usage over other R-TPs (e.g., WiFi AP) in the COHERENT architecture.

6.4.1 Interworking of COHERENT Architecture and Legacy 3GPP Functions

In COHERENT, C3 & RTC could be physically co-located in certain the deployment scenarios such as relaying or broadcast operation with legacy 3GPP functions. In this situation, the southbound API functionalities do not change, but the combined functionalities will become as described below:

- Control different parameters from different network or infrastructure entities (Multi-RAN, switches, different layers) through APIs;
- Collect/coordinate configurations/policies from network applications;
- Gather information from network and infrastructure through APIs, and upgrade/create Network Graphs among control instances;
- Provide slice-specific and local network graph to network applications through API;
- Specify functional split of different entities, e.g., for eNB in a configuration vRP & R-TP, or other configuration such as BBU (Baseband Unit) & RRH (Remote Radio Head).

Possible COHERENT protocol stack using vRPs in a 3GPP view is depicted in Figure 6-7and Figure 6-8 as a proof on concept, showing that COHERENT architecture could be mapped to 3GPP systems.

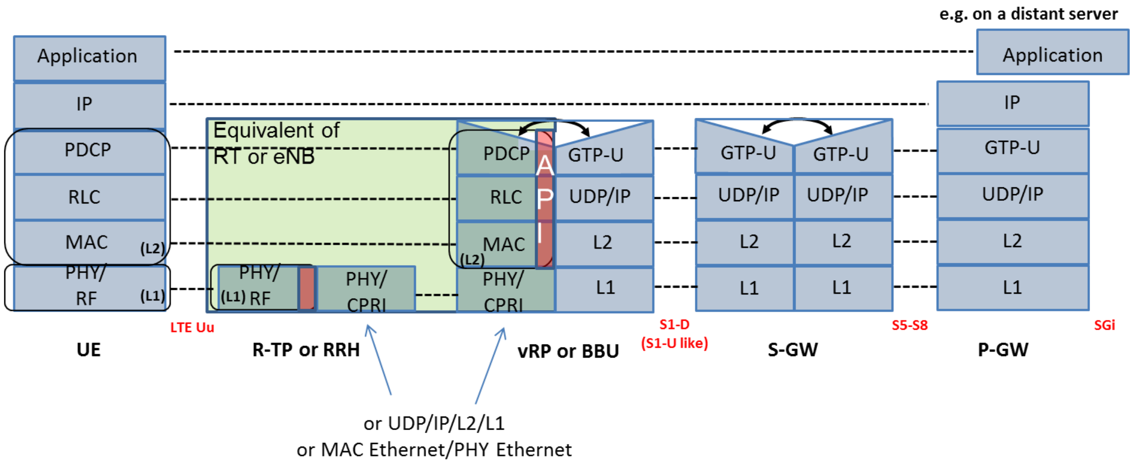

Figure 6-7 Possible User Plane COHERENT protocol stack using vRPs (3GPP view)

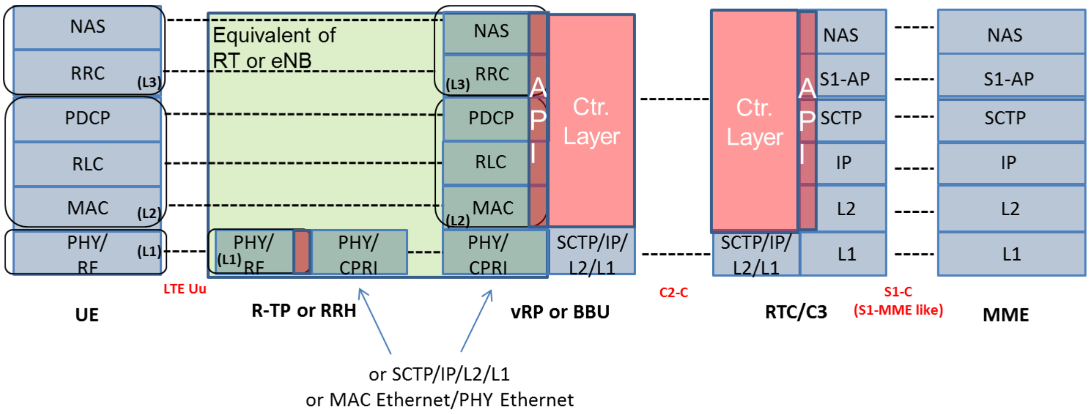

Figure 6-8 Possible Control Plane COHERENT protocol stack using vRPs (3GPP view)

For CP and UP split, not necessary the same virtual machine or vRP instance is used. Moreover, alternatives for I/Q data transmission between RRH & BBU are:

- CPRI (Common Public Radio Interface)[98][99]
- OBSAI (Open Base Station Architecture Initiative)
- ORI (Open Radio equipment Interface)
- Radio over Ethernet

CPRI has a strict latency control, is bidirectional and it has a constant bit rate. However, CPRI issue is that it requires a huge bandwidth. An adaptation/translation between CPRI and MAC PDUs could be used on the fronthaul side in order to send MAC packets instead of directly sending I/Qs (which would require larger bandwidth).

6.4.2 3GPP Broadcast Operation adapted to COHERENT Architecture

3GPP Broadcast operation is not very flexible or easily configurable (see Section 2.3.1.3 for details). The introduction of a controller for fast parameter management adapted for example to Public Safety situations would be necessary. Different broadcast mode situations can be imagined but in the following paragraphs we will describe only two of them:

- Protocol stack adapted to COHERENT architecture for broadcast system using eMBMS (User Plane and Control Plane):

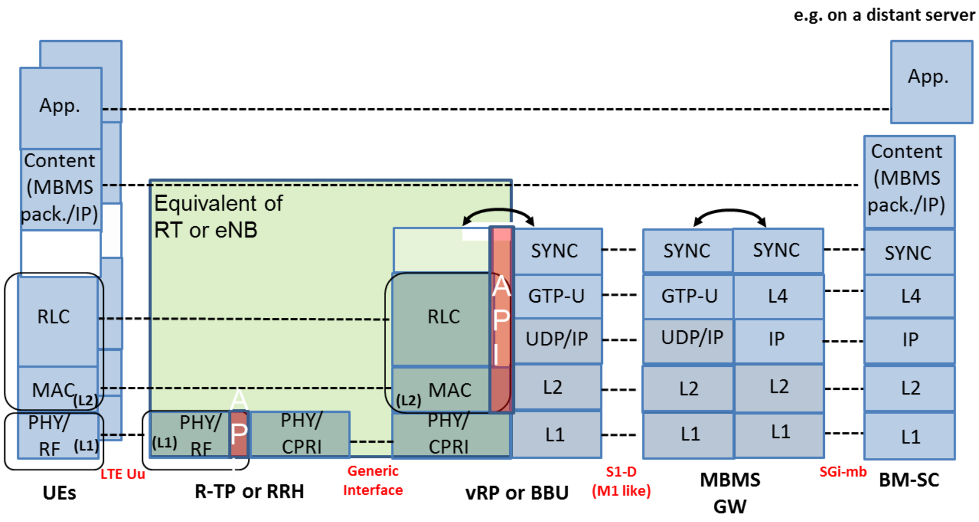

Figure 6-9 Possible User Plane for Broadcast Operation using vRPs & eMBMS service

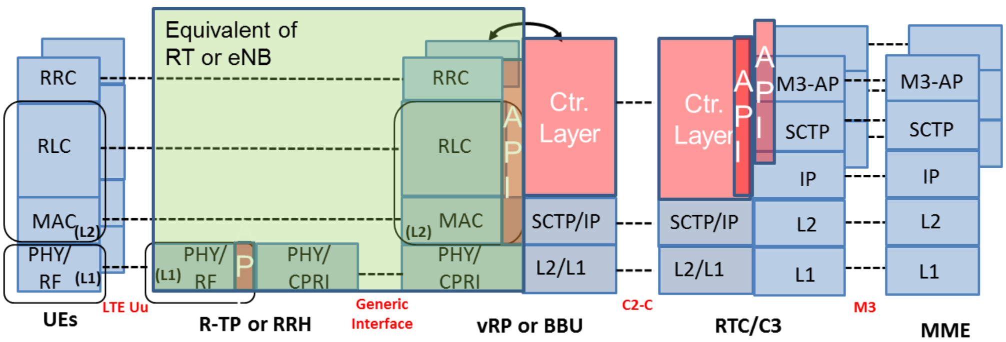

Figure 6-10 Possible Control Plane for Broadcast Operation using vRPs & eMBMS service

- Protocol stack adapted to COHERENT architecture for broadcast system using unicast bearer (User Plane and Control Plane):

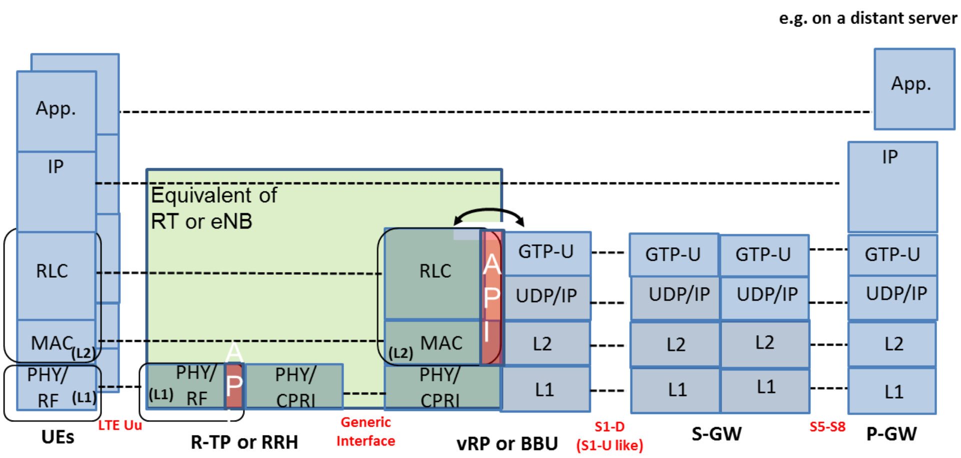

Figure 6-11 Possible User Plane for Broadcast Operation using vRPs & Unicast Bearer

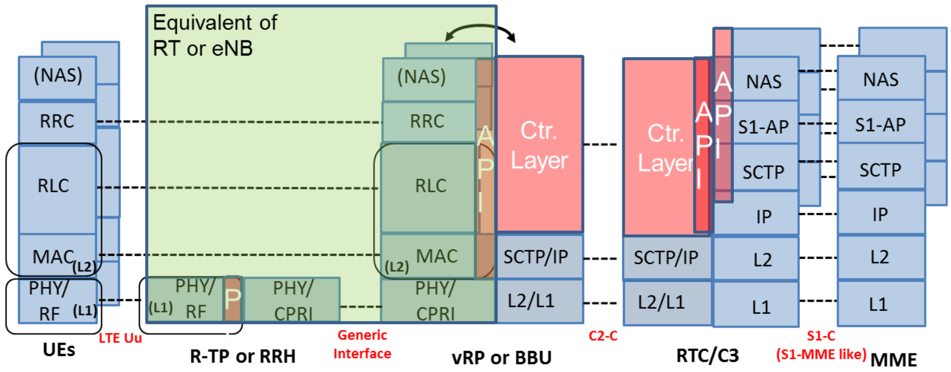

Figure 6-12 Possible Control Plane for Broadcast Operation using vRPs & Unicast Bearer

While in the first case an eMBMS-like architecture is used with classic eMBMS components, in the second case unicast architecture is used, with unicast bearer adapted for broadcast operation, which is a very new approach. Above figures clearly show that there is no impact on the current network entities. C3/RTC can be used with functionalities mentioned in Section 5.2.1.

6.5 UE-Relaying Operation for Public Safety Service

In this section we are providing a description of how COHERENT could improve Public Safety services by adding a UE-Relay operation under the control of the network operator. The UE-Relay operation is necessary for network extension but it could also improve network operator services over a given area. However, mobility management and required configuration are functions that currently do not exist for this type of operation and will represent extra added values. These functionalities are related to the use cases already presented in D2.1 (see Table 6-1 in Section 6)and will be further described in D3.1 and D3.2 (WP3).
Table 6-3

Similar to Section **Error! Reference source not found.**, the relaying scenarios in this section are mainly for UE-relaying in the cellular networks (3GPP). Therefore, the southbound API functionalities mentioned in Section 6.4.1 are applied to this section as well. Different layer abstraction models and network graphs can be used also for improving mobility and network management as illustrated in Figure 6-13.

The COHERENT architecture can therefore also be applicable to network coverage extension and mobility management is illustrated in Figure 6-13. Depending on technology-related parameters such as Layer 1 (PHY) and Layer 2 (MAC, RLC, PDCP), service parameters and requirements (such as required throughput, latency, priority over other services, or Public Safety operations), and system parameters such as mobility information (trajectory, speed etc.) and type of environment (urban, rural, etc.), the C3 and RTC can decide, for instance, how to improve mobility management, how to perform relay selection and how to configure the system for coverage extension in out-of-coverage situations.

Some specific usages of the C3/RTC (see Section 6.4.1) are therefore applicable to mobility management, to coverage extension and to optimisation of network performance when possible. The role of the C3/RTC is to configure different network entities and to make decisions with respect to the network state. The C3 has a global view and can also decide to virtualise different layer functions.

The COHERENT C3/RTC obtain inputs from a block responsible for network graph creation, physical or logical system entities (for different RATs) and possible other information from databases with rules and policies for different RATs. The network graph can be periodically updated with information coming from measurement reports and is connected to an abstraction toolbox responsible of low layer and high layer abstraction of different transmission modes such as Point-to-Point and Point-to-Multipoint transmissions, or different Layer 1 and Layer 2 configurations. Moreover, the abstraction toolbox implementation corresponds to one or several NIFs and NIF managers as described in Section 5.4.

The abstraction toolbox can be used in two ways: i) to identify significant PHY, MAC & RLC parameters, probably depending on the set-up and ii) to derive best metrics to be used depending on the use case and network optimisation goal. For instance, in the case of coverage extension the operator wants to reach users which are not normally reachable and therefore this can be seen as a new type of application with different requirements e.g., in terms of latency. In this case there may be for example unexpected service interruptions and traffic congestions. Different service priorities can be provided to relay firemen and policemen under network out-of-coverage situations in both urban and rural scenarios. The system shall find the best configuration such that the service requirements are satisfied. There are many applications and examples are varying from areas where normally there is no network coverage such as tunnels, in-door situations (buildings, basements), or even valleys and caves, to emergency situations such as life-saving emergency relief after a disaster.

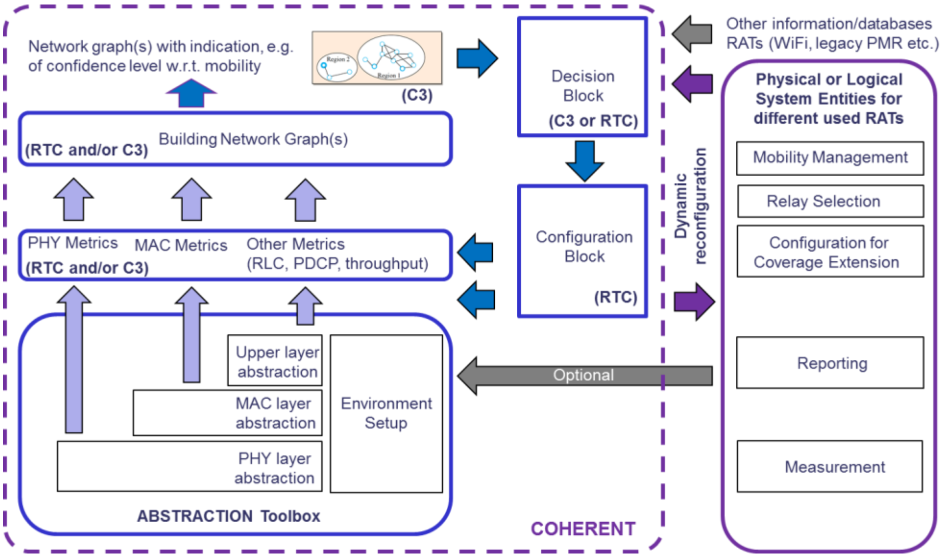

Figure 6-13 Application of COHERENT architecture for network coverage extension and mobility management

After acquiring significant network information such as BER, BLER, estimations of link qualities and speed, achievable data rates and latencies, the C3/RTC can reconfigure dynamically different system entities (MME, Gateways, eNodeB, Access Points, UE or other) for relay selection, coverage extension and inter-RAT or intra-RAT handover. C3/RTC may also be responsible of protocol configuration for measurement and reporting, which can be event-based or periodical.

Figure 6-14 explains some of the potential scenarios using the new network architecture. We have represented both control plane and user plane for legacy communication and for public safety scenarios. In Figure 6-14 the role of the COHERENT RTC/C3 is to configure communications towards distant (e.g., out-of-coverage) users such firemen or policemen who require the usage of the network infrastructure or to configure in-coverage or out-of-coverage direct mode operations (Device-to-Device or D2D). This kind of interfaces (see new control plane for Public Safety scenarios in Figure 6-14) have also been described in Section 2.3.1.2 and Annex A.3 for 3GPP architecture in Release-12 and Release-13, but only between UEs, as one may notice PC5 interface with a very light PC5 signalisation layer which is used for very basic operations and does not have any level of control from EUTRAN side. Moreover, if eNB is deployed using a flexible architecture e.g., using network virtualisation strategies or protocol stack split, it allows for a software design which is more attainable since each protocol (e.g., Layer 2 and above) can be seen as an instance or a process, allowing for a faster configuration and adaptability. The main advantage of this kind of architecture is that it still supports legacy communications and for new services it does not require any significant system upgrade or replacement of old equipment.

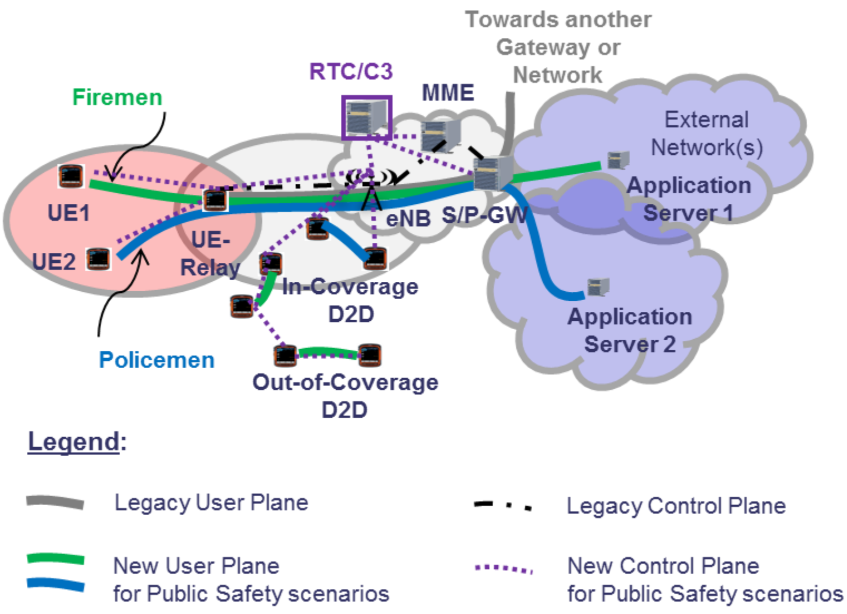

Figure 6-14 Example of Coverage Extension and Mobility Management

In the example represented in Figure 6-14 a UE1 used by firemen is relayed by a UE-Relay (a legacy UE acting as relay) towards an Application Server 1. The interface between UE1 and UE-Relay will be further discussed in WP3 and not clearly defined by standardisation bodies; it can be a ProSe interface PC5 (3GPP) or other such as WiFi or Bluetooth (IEEE). Similarly, policemen using UE2 can access as well the network infrastructure and can communicate with an Application Server 2. The UE-Relay having the relay functionality may have its own traffic to transmit or to receive, but is also responsible of relaying control plane and user plane traffic to UE1 and UE2. Nevertheless, different priorities can be defined. For the sake of purpose Figure 6-14 also presents other classical PMR-like operations as Direct-Mode Operations (DMO) performed by out-of-coverage and in-coverage policemen and firemen users which communicate directly without passing through the network infrastructure but using some of the operator resources. The devices can have some control from the network side but the user plane is directly between devices. However, 3GPP did not yet define a specific procedure for UE-relay operation, out-of-coverage situations or single-hop and multi-hop scenarios. Currently, as being shown in Subsection 2.3.1.2, relaying can be performed only at application level without any impact on Access Stratum which is expected for Release-14 or beyond. COHERENT control entities (C3 and RTC) can add new relaying features with minimum impact on existing equipment from a given Release, and can lead to better mobility management, better coordination, better interference management and better resource allocation without any significant change of the equipment. Similarities can be also found with D2D communications from Section 6.3, and the overall concepts will be further discussed in WP3.

6.6 Multi Connectivity

In principle, dual connectivity allows the UEs to receive data simultaneously from different eNBs (eNBs) in order to boost the performance in a heterogeneous network with dedicated carrier deployment. Dual connectivity techniques are expected to be increasingly important especially in environments consisting of multiple macro and small cells. LTE dual connectivity is a technology which extends carrier aggregation and coordinated multi-point (CoMP) to inter-eNB with non-ideal backhaul while LTE Release 12 focuses on inter-eNB Carrier Aggregation (CA) case for per UE throughput improvement and mobility robustness. As also presented in [100] the motivation for dual connectivity comes from the need to meet the following challenges:

- Efficient radio resource utilisation across eNBs
- Mobility robustness
- Increased signalling load
- UL/DL imbalance between macro and small cells
- Increase UE throughput especially for cell edge UEs

The way COHERENT architecture is designed in order to provide a robust yet extensible control framework, it can be used to facilitate the multi-connectivity procedures and programmatically act reactively (or proactively) to user mobility and association events.

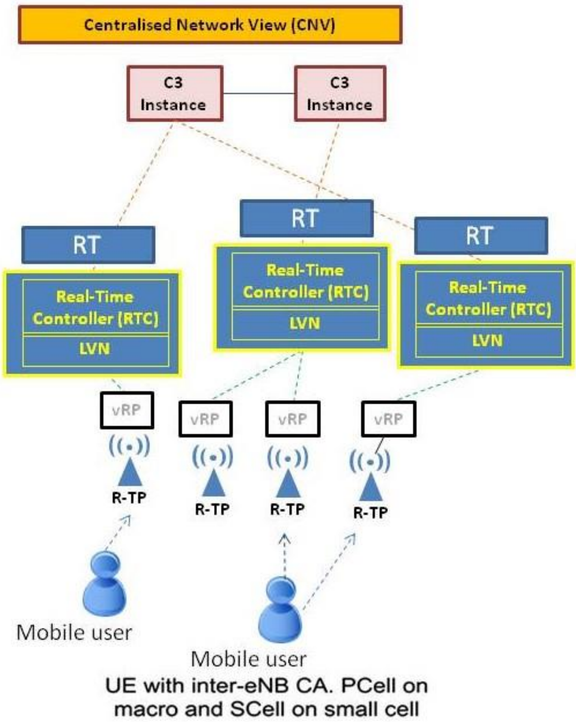

Figure 6-15 Multi-connectivity scenarios

In multi-connectivity scenarios the COHERENT solution adopts a SDN approach in the way core network data are handled, while also the RAN packet scheduling efficiency is promoted. We also highlight that the COHERENT integrated solution will provide the necessary testbed over which a number of research efforts like new flow control algorithms will be demonstrated and synchronisation issues between the eNBs will be considered, under the multi-connectivity concept. Currently the UEs send measurement report, while in future releases, the UE will be able to decide which eNB to connect with.

7. Mapping to the 3GPP Framework

Section 5 provides the generic COHERENT architecture. While the generic architecture covers basic concepts suitable for different RATs, including, but not limiting to, LTE and WiFi systems, it needs the necessary adaptation when it is applied to particular radio access technologies, like LTE. Particularly, the COHERENT project has invested efforts in 3GPP RAN3 (see contribution in [101]) in WP7 activities. This section describes the architecture, which COHERENT proposed to 3GPP, targeting to display enhancements to the already defined 3GPP architecture. We name it *3GPP oriented Base Station System (BSS) architecture*. Since 3GPP has its own draft on 5G architecture (including a certain terminology used already in NGMN and Small Cell Forum), some adaptations and modifications have been made within 3GPP oriented BSS architecture in order to fit the requirements in 3GPP TR 38.913 [102], and to facilitate the consensus in the 3GPP community. The mapping of the 3GPP oriented BSS architecture to the COHERENT architecture is examined in details in Section 7.2. Since the 5G design in 3GPP is an ongoing work, our 3GPP oriented BSS architecture will continue evolving to meet the needs and requirements from standardisation bodies.

7.1 3GPP Architecture for the New Radio

The up-to-date architecture and terminology are reflected in the picture reproduced below from [103].

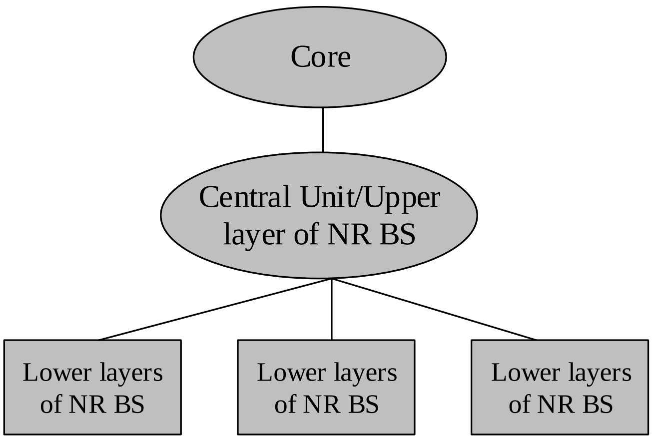

Figure 7-1 Centralised Deployment [103].

In the figure above the layers refer to the layers of the RAN protocol stack and NR (New Radio) is the 3GPP name for new 5G radio technology.

It should be noted that this draft does not reflects the separation of the Control Plane (CP) and User Plane and does not introduce the hierarchical control levels considered by 3GPP oriented BSS architecture.

7.2 Mapping of 3GPP oriented BSS Architecture to COHERENT Architecture

To follow the 3GPP common practice, several terms and building blocks in the 3GPP oriented BSS architecture are used differently from COHERENT architecture (shown in Figure 5-1). The 3GPP oriented BSS architecture is shown in Figure 7-2. The details of the architecture are explained in Section 7.2.

We refer to the future cellular system as Base Station System (BSS). BSS is defined as a multitude of logical nodes which include on the control plane multiple controllers and a single coordination function and on the user plane the processing of the user data as well as a virtual hybrid routing function forwarding the user plane (UP) higher layer protocol units to the selected R-TPs.

In the 3GPP community a distinction is made between the eNB-specific control functions and inter-eNB coordination functions. In the 3GPP oriented BSS architecture we address the coordination functions in a centralised mode (C3) and in the same time we keep the approach of the local control (in RTC).

In this section, unless otherwise stated, the *control* is defined as an activity to determine the behaviour of a network element. The *coordination* is defined as an activity to enable *the efficient operation* of the distributed control functions implemented in RTCs. In this section C3 will be renamed Central Coordinator (CCord) because we want to emphasize the functional split between control in the RTC and coordination in CCord. In addition, 3GPP oriented BSS architecture introduces the *virtual hybrid router* which is essential for mobility, multi-connectivity and Network MIMO operation of a 5G cellular network.

The mapping between the terminology used in COHERENT architecture (Figure 5-1) and the terminology used for 3GPP oriented BSS architecture (Figure 7-2) is shown in the following table:

Table 7-1 Mapping between COHERENT Terminology and 3GPP oriented BSS terminology

COHERENT Terminology	**3GPP oriented BSS Terminology**
Central Controller and Coordinator (C3)	Central Coordinator (CCord)
Virtual Radio Processing (vRP) on a computing platform	Central Unit / Upper layers of NR Base Station
Real-Time Controller (RTC)	Controller of the higher UP layers
Radio Transmission Point (R-TP)	Lower layers of NR Base Station

7.3 3GPP oriented Base Station System (BSS) Architecture

The 3GPP oriented BSS architecture shown in Figure 7-2 comprises the core network and the RAN. The core network includes P-GW, S-GW, Operation and Administration Server (O&M), MME and the Spectrum Manager. The Spectrum Manager enforces the high level policies regarding the spectrum sharing and communicates with the Central Coordinator (CCord).

COHERENT architecture supports a heterogeneous network, including virtualised and non-virtualised heterogeneous wireless and mobile nodes, centralised and non-centralised (including eNB for LTE and NR for 5G radio). In line with the COHERENT approach, all the radio stations, virtualised or not, and the R-TPs in a geographical area are coordinated by a Central Coordinator over a C1 interface. The entities on vRP include the Higher Layers of the UP and the associated Controller on the CP.

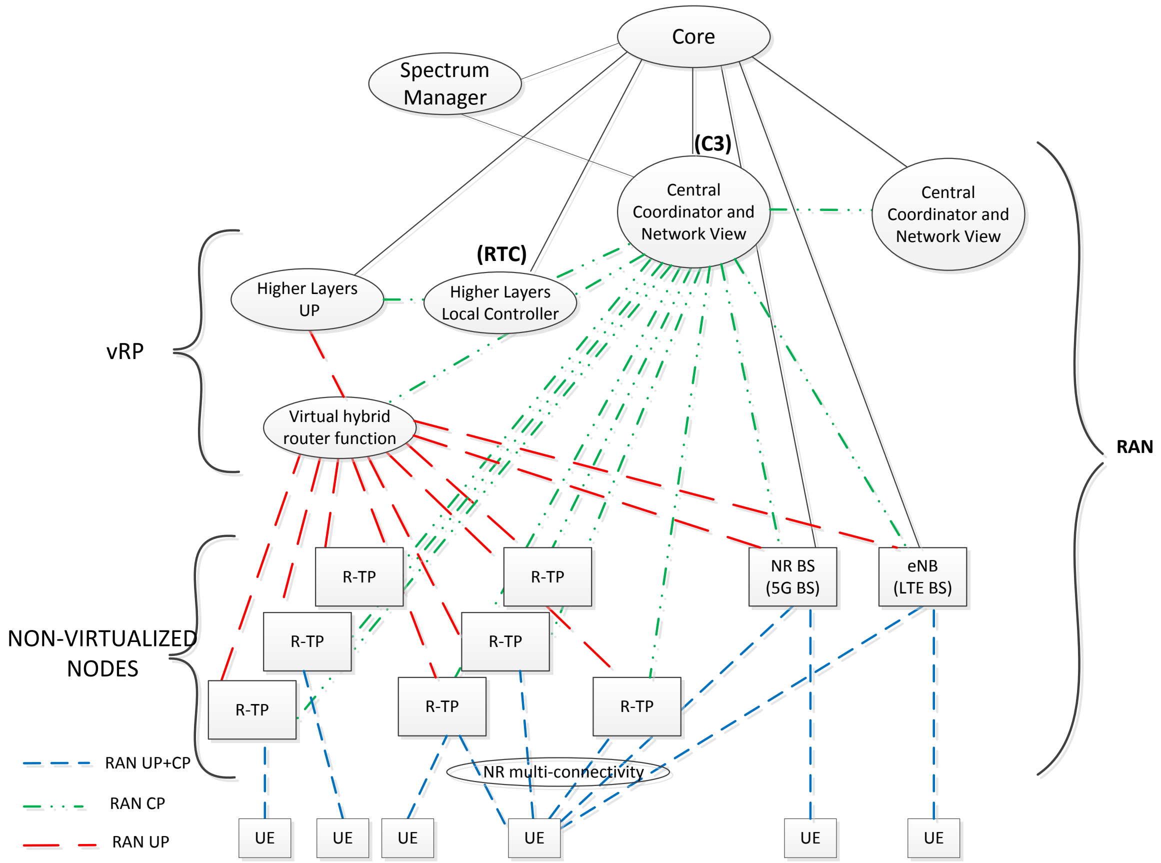

Figure 7-2 3GPP oriented BSS architecture

The difference between the vRP as presented in COHERENT architecture and the vRP in 3GPP oriented BSS architecture is that a *Virtual Hybrid Router* is introduced on the UP, between the vRP and R-TP. This new logical function will allow improving the hand-over performance and also will allow the operation in multi-connectivity, including Network MIMO. The *Virtual Hybrid Router* is controlled by the Central Coordinator (CCord).

The UEs can be served by the one or more base stations; in case of multi-connectivity (see also Section 6.6) the participating R-TPs are selected by CCord which considers the information in the network graphs included in the Network View.

Advantages of proposed architecture for cellular systems are summarized below:

- The transmission point selection and the coordination of resources will perform better as compared with the existing distributed solutions.
- The control is hierarchical, i.e. local control functions requiring low delay are separated from control functions with less stringent real-time requirements.
- The served area of the BSS is significantly increased, including a multitude of heterogeneous R-TPs, i.e. R-TPs with different transmission powers or using different RATs and also stand-alone NR and eNB base stations.
- For the served area of the BSS the mobility is supported in RAN. This approach allows for the reduction of the HO delay, reduction of the connectivity failures during HO, reduction of the Core overload.
- The Virtual Hybrid Router, under the control of the Central Coordinator, can be used in more complicated UP scenarios, such as multi-point transmission including Network MIMO with Joint Processing.

- The tight NR-LTE interworking is realised in RAN.
- The BSS also assures the compatibility between existing LTE Base Station and the New Core.

7.4 Description of 3GPP oriented BSS Functions and Interfaces Relevant to COHERENT

7.4.1 User Plane

7.4.1.1 VMs within the vRP

The user plane connects the UE with the P-GW, while passing through the S-GW. In RAN the user plane connects the UE with the VMs on the vRP. A VM hosts a number of UE-related data processing functions and possibly some of the control functions. The routing of PDUs (Protocol Data Units) is done at the level of virtualisation platform or per VM, to the destination R-TPs, identified by a TP-ID (similarly with BS-ID).

In this way, the VM hosts one or more of the higher layers related to one or more specific UE(s). As opposed to the existing Base Stations, the UP and CP are handled by different entities, applying in this way the SDN principles of UP and CP separation. We prefer this approach given the relative long time and the load balancing problems related to migration of an UE from a VM to another VM, as is the case in handover. The VM migration takes place only in special cases.

The traffic between S-GW and the serving VM, as well as the traffic between the VM and the R-TP should be tunnelled[104][105]; the TEID (Tunnelling End ID) could be provided by base station for downlink or by MME for uplink. In 3GPP oriented BSS architecture the TEID could be provided by Central Coordinator.

7.4.1.2 Virtual Hybrid Routing Function

The Virtual Hybrid Routing function of the VM or of the vRP sets a destination IP address including the port based on the TP-ID provided in general by the CCord as an IE (Information Element) in its messages.

However, the TEID (Tunnel End Identifier) is allocated today for uplink by MME; as in a network may be multiple MMEs and between MMEs could be applied load balancing, the Central Coordinator may become the entity which in a given area provides the TEID for each UE. This will simplify the UE-specific interactions with the core network.

If the UE is new in the coordinated area, the Central Coordinator authenticates the UE, allocates the TEID and sends messages to S-GW and surrounding R-TPs including both the TEID and the UE IP address which was allocated by the P-GW.

7.4.2 Control Plane

7.4.2.1 Hierarchical Control Concept

The LTE wireless access network coordinates itself in a distributed mode, through the existing X2 interface, with an API described in [106]. In the centralised approach the RAN control plane is split between:

- Local control entities, in charge with the local, low latency, control functionality
- The higher level coordination function, named Central Coordinator (CCord) in charge with the harmonisation of the operation of different controllers.

The Central Coordinator can be placed on a base station, on a server located at the network edge (routers, etc.) or on Internet. The Central Coordinator uses a new **C1 interface**, which can be in the future also part of the X2 interface or be a stand-alone interface.

7.4.2.2 C1 Interface Set-Up

The first operation is the set-up of the C1 interface. Assuming that the Central Coordinator IP address is know from an O&M configuration or can be discovered, the C1 set-up will be initialised by each control block in a VM and by each R-TP. At set-up each R-TP and VM should include its identifier and its capabilities in terms of supported UE ID (it should be an UE identifier at a level higher than C-RNTI, possibly X2 UE Identifier), layers and functionalities.

7.4.2.3 Central Coordinator

The Central Coordinator should assume the BSS coordination functions. The BSS coordination can address the following RAN functions:

- Allocation of resources for base station transmission
- Load balancing
- Mobility
- Power levels of DL or UL transmissions
- Time-frequency resource allocations for UE transmission, where time may refer to time-slots or subframes, while frequency may refer to a frequency channel, a group of frequency channels belonging or not to an operator, a sub-division of a channel (PRBs, subbands, etc.)
- Time-frequency resource allocations for reference signals
- Coordination of protected resources
- Operation in Network MIMO transmission modes
- Operation in multi-connectivity transmission mode
- Medium access in un-licensed bands
- Spectrum sub-licensing
- Inter-operator infrastructure and/or spectrum sharing
- Tight LTE – NR interworking
- Coordination of computing resource allocation on the virtualisation platform.

7.5 High-level Description of SB (south bound) API

The SB API includes the protocols and semantics related to the following functionalities:

- Construction of the network graph data-base resulting from the PHY/MAC abstractions defined in WP3 with M3.3 as first delivery target
    - The procedures and semantics for the CCord-RTC interface will be defined in D3.2

7.6 High-level Description of the Spectrum Manager API

- Enhancement of the network graph with the spectrum-related network graph, as defined in WP4
    - Semantics and procedures for Spectrum manager and CCord interaction will be defined in D4.2

7.7 High-level Description of NB (north bound) API

The NB API refers to:

- The semantics of the dedicated programming language which will be used by a programmer for interactions with the entities which interact based on the SB API and Spectrum API; to be defined for the cellular systems in T2.3.

- The semantics of the dedicated programming language which will be used by a programmer for interactions with the data-base including the Network View; to be defined for the cellular systems in T2.3
- The structure and semantics of the Network View.

8. Abstractions and Interfaces for Programmability in Heterogeneous Mobile Networks

8.1 Introduction

The goal of this section is to introduce the abstractions and interfaces for heterogeneous mobile networks programmability. We remind the reader that within COHERENT with the term “heterogeneous” we refer to radio access networks composed of, but not limiting to, WiFi and 3GPP radio network elements.

In this section we shall first introduce the basic concepts behind domain specific languages (DSLs) and the role they play in the definition of the COHERENT Abstractions and Interfaces. Then, we shall introduce the COHERENT semantic model. Finally, we will report on the northbound interface and the primitives which allow network programmers to manipulate the semantic model and thus access and change the state of the network.

It is worth noticing that, in this document, we shall account for the preliminary COHERENT semantic model and interfaces. Such model addresses the requirements of Enterprise WLANs build upon the 802.11 family of standards. Support for 3GPP networks shall be added within the lifespan of task T2.3 and will be reported in D2.4.

8.2 On Domain Specific Languages

Domain specific languages (DSLs) are computer languages that are targeted to a particular kind of problem, as opposed to a general purpose language that is aimed at any kind of application domain. Very popular examples of DSLs include CSS, regular expressions, SQL, etc.

DSLs are usually classified in two broad categories:

- **Internal DSLs.** In this case the DSL is embedded within a host language using some of its constructs and directives in order to have the appearance of an actual DSL. The host language is usually a general purpose language with dynamic features such as Python or Ruby. Internal DSLs are also referred to as embedded DSLs. Internal DSLs are a particular form of API in a host general purpose language.
- **External DSLs.** In this case the DSL has its how syntax and a full parser for its processing. Examples include the Makefile language.

Both types of languages have two fundamental components: syntax and semantics. The syntax captures which statement is valid in a certain program. In an Internal DSL the syntax is the one of the host language. On the other hand, the semantics of a language is what it means and what it should do when executed. The model is in this case what defines the semantic.

Defining a proper semantic model is the first step in the creation of a DSL and thus of an API (Application Program Interface) to be implemented in an SDK. Semantic model and DSL can evolve independently, for example new features can be added to the semantic model before having to decide how to expose them through the DSL. Conversely, from a particular idiom that proves to be particularly useful to some a domain specific problem, it is possible to go back to the semantic model and enrich it with the newly acquire knowledge.

8.3 Semantic Model

Programming wireless networks requires identifying how network resources are exposed (and represented) to software modules written by developers and how software modules can affect the network state. Although, OpenFlow provides a practical forwarding abstraction for packet-switched networks, it has been argued that programming current networks using OpenFlow is equivalent to programming applications in assembler, i.e. the interface is too low-level and exposes the programmers with too many implementation details. As a result, we have witnessed a plethora of efforts in recent years aimed at providing developers with higher level interfaces to their SDN, e.g., Pyretic, Nettle, Frenetic, Procera, etc.

The works above, however, aim at enabling programmability in wired networks and typically rely on OpenFlow as the data-plane control API. In contrast, our aim is to investigate which kind of abstractions can be effectively used to implement control and coordination tasks in wireless networks. In fact, a straightforward extension of current programming techniques would fail to capture the peculiarities of the radio environment. In particular, the flow abstraction on which OpenFlow relies does not account for: (i) the stochastic nature of wireless links (which are not equivalent to ports in Ethernet switches); (ii) the resource allocation granularity (the flow abstraction is too coarse for wireless networks); and (iii) the significant heterogeneity in the link and radio layer technologies (state management for network elements can differ significantly even across currently deployed technologies).

8.4 Enterprise WLANs

In this section we shall report on the COHERENT primitives and abstractions for Enterprise WLANs programmability. This document extends the original submission [55] by refining both the semantic model and the northbound interface. The primitives and abstractions presented in this section address four control aspects of wireless networks programmability, namely: state management, resource allocation, network monitoring, and network reconfiguration:

- State management. The abstractions shall allow programmers to describe the desired behaviour of the network leaving to the underlying run-time system the task of implementing it by configuring the individual network elements. Programmers shall not be exposed with technology-dependent details such as how to implement handover for a particular link-layer technology.
- Resource allocation. The abstractions shall allow developers to leverage a global view of the network resources. The programming abstractions shall be general enough to accommodate for resource allocation models ranging from random access to scheduled access.
- Network monitoring. The abstractions shall allow network developers to gather the status of the network using high-level querying primitives. Such information shall include network statistics and topology changes. The network programmer shall not be exposed to the system-level details of polling network elements, aggregating statistics, and detecting network events.
- Network reconfiguration. The programming abstractions shall go in the direction of separating fast timescale operations running at the edges of the network, i.e. near the air interface, from tasks running at the C3.

Table 8-1 introduces the terminology used in this section mapping it to the general COHERENT nomenclature and architecture. We use the term Wireless Termination Points (WTPs) to refer to the physical devices that form the Enterprise WLAN providing clients with wireless connectivity. WTPs basically coincide with Access Points (APs). A secure channel connects the WTPs to the remote Central Controller and Coordinator, or C3, (details about the southbound interface are provided in a separate specification document). Network Apps run on top of the C3 in their own sandbox and exploit the programming primitives through either a REST API or a native Python API (this is the COHERENT northbound interface). No real-time controller (RTC) is defined for the Enterprise WLAN scenario.

Table 8-1 Mapping between COHERENT Terminology and Enterprise WLANs Terminology

COHERENT Terminology	Enterprise WLAN Terminology
Radio Transceiver (RT)	Wireless Termination Point (WTP) or WiFi Access Point (AP)
Wireless Client	Light Virtual Access Point (LVAP) or WiFi Stations

In the next subsections we will introduce four key abstractions for wireless networks, namely the Light Virtual Access Point (LVAP) abstraction, the Resource Pool abstraction, the Channel Quality Map abstraction, and the Port abstraction. Figure 1 depicts the relationship between the abstractions using an UML class-diagram.

Here, a Resource Block represents the minimum chunk of wireless resources that can be assigned to a wireless client (i.e. a WiFi Channel) while the LVAP represents the state of a wireless client scheduled on a set of Resource Blocks. Both LVAPs and WTPs support a set of Resource Blocks named Resource Pool. The Port abstraction models the dynamic and reconfigurable characteristics of the link between WTPs and LVAPs. Link transmission statistics are associated to each Port. A relationship exists between LVAPs and WTPs and between WTPs modelling the link quality between the two entities. These relationships are captured by the Channel Quality Maps (CQMs). The diagram in Fig. 1 depicts the COHERENT semantic model for Enterprise WLANs using an UML Class Diagram.

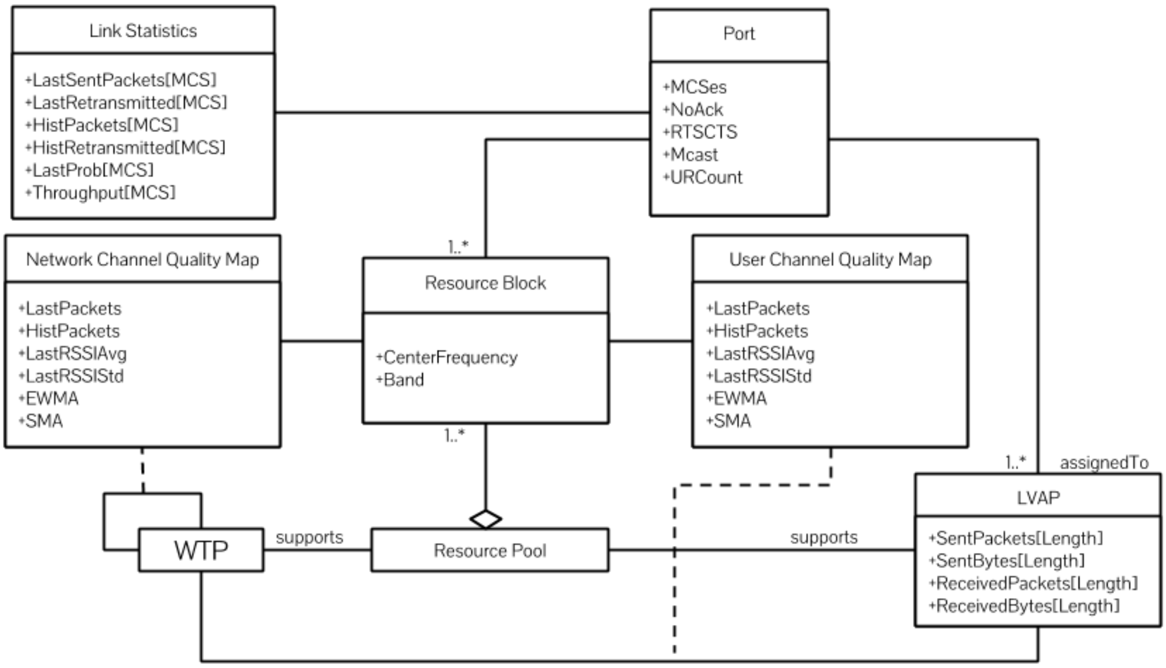

Figure 8-1 The COHERENT Semantic Model for Enterprise WLANs

8.4.1 Light Virtual Access Point

Different link layer technologies, or as a matter of fact even different releases of the same technology, can differ significantly in how a client's state is handled. For example, QoS and handover management changed significantly over the lifespan of the IEEE 802.11 family of standards.

Nevertheless, exposing the programmer with the implementation details of the technology being used would severely limit the adoption of a certain solution. On the contrary what we want to achieve is true Network Apps portability from one RAN to another, e.g., from a WiFi network to a mixed WiFi/LTE deployment.

The LVAP abstraction [55] provides a high--level interface for wireless client state management. The implementation of such an interface handles all the technology - dependent details such as association, authentication, handover, and resource scheduling. A client attempting to join the network will trigger the creation of a new LVAP. Specifically, in WiFi, an LVAP is a per-client virtual access point which simplifies network management and introduces seamless mobility support, i.e., the LVAP abstraction captures the complexities of the IEEE 802.11 protocol stack such as handling of client associations, authentication, and handovers.

More specifically, every newly received probe request frame from a client at a WTP is forwarded to the C3 which may trigger the generation of a probe response frame through the creation of an LVAP at the requesting WTP. The LVAP will thus become a potential candidate AP for the client to perform an association. The C3 can also decide whether the network has sufficient resources left to handle the new client and might suppress the generation of the LVAP. Similarly, each WTP will host as many LVAPs as the number of wireless clients that are currently under its control. Such LVAP has an ID that is specific to the newly associated client (in a WiFi network the LVAP can be thought as a Virtual AP with its own BSSID). Removing a LVAP from a WTP and instantiating it on another WTP effectively results in a handover.

8.4.2 Resource Pool

Broadly speaking, there are two main families of strategies to allocate resources in a wireless network: scheduled access and random access. In the former case, resources for a wireless link are allocated in the time, frequency, and space domains. In the latter case, a common random access scheme for medium access is used by all participating wireless clients in order to reduce collisions. LTE belongs to the former family and uses OFDMA as (scheduled) medium access scheme in the DL and SC-FDMA in the UL, while WiFi belongs to the latter family and exploits CSMA/CA as (random) medium access scheme. Please notice that LTE has already one channel working with random access, e.g., PRACH (Physical Random Access CHannel). However, this channel is used mostly during the UE attach phase. Moreover, future releases of the LTE standard dealing with M2M communications are expected to introduce random access also for data.

The minimum allocation unit in a WiFi system is the channel identified by a frequency, a band type, and the AP at which it is available. In order to have a consistent naming with LTE system, we decide to rename the WiFi Channel/Band combination as Resource Block. Each Resource Block is fully described by a 2–tuple <c, b> c and b are, respectively, the center frequency and the band type.

For example, the Resource Pool made available by an 802.11n AP tuned on channel 36 and supporting 40 MHz–wide channels is represented by the tuple (36, HT 40). The prefix HT is used to indicate that this band supports the High Throughput MCS. Resource Blocks can also be blacklisted preventing the C3 from using them.

Each WiFi AP/Client in the network can support as many Resource Blocks as the number of its WiFi interfaces. For example, a dual band WiFi AP would expose to the C3 the following list of Resource Blocks: (6, HT 20) and (36, HT 20). This models the fact that the AP can support at the same time two operating bands.

Notice that the proposed model does not forbid the same Resource Block to be assigned to multiple LVAPs in that this could in general result in a valid resource allocation scheme. In fact, wireless networks using random access protocols, such as CSMA/CA, effectively schedule multiple transmissions on the same resources in frequency and time with the aid of suitable back–off and retransmission schemes to handle collisions.

Notice how the same formulation can be used also to model the channels and bands supported by a wireless client. For example, a resource request could be represented by the following tuple (1, 20). This allows us to express resource allocation problems as an intersection between the Resource Blocks available in the network, and the Resource Blocks supported/requested by a client. Information on the link quality experienced by the requesting client on the matching Resource Blocks can be used to further filter the set of candidate Resource Blocks according to application–level parameters. A non–empty intersection set of Resource Blocks signifies that a valid solution for the resource allocation problem has been found.

Notice that, the final set could be composed of multiple Resource Blocks possibly scheduled at different APs and on different frequency bands. The support for such scenario depends on the actual implementation of the client radio interface. This model also effectively decouples uplink and downlink allowing clients to be scheduled at different APs on the uplink and on the downlink directions (if the feature is supported by the link–layer technology).

An example of a simple resource allocation scenario for a WiFi WLAN is shown in Table 8-2. Here P_N are the network Resource Blocks and P_L are the Resource Blocks supported by a client. It is easy to see that the intersection $P_N \cap P_L$ produces a non–empty set composed of two Resource Blocks scheduled at two different APs (W1, W2). Due to the fact that WiFi does not allow scheduling one client on more than one AP, the final resource allocation decision will be a single Resource Block selected using criteria such as the channel quality experienced by the client on the matching Resource Blocks and/or the specific application–level requirements.

Table 8-2 Resource Provisioning Model

Network	Client	Intersection
W 1 (6, HT 20) W 1 (36, HT 20) W 2 (1, HT 20) W 2 (36, HT 20) W 3 (11, 20) W 3 (36, 20)	L 1 (36, HT 20) L 1 (48, HT 20) L 1 (54, HT 20)	W 1 (36, HT 20) W 2 (36, HT 20)

8.4.3 Channel Quality Map

From the perspective of a client, it is not important to which WTP it is attached but what communication QoS it can obtain. Such information translates, at the physical layer, into the transmission efficiency of the radio channel linking interference with parameters such as packet error rate. The Channel Quality and Interference Map allow the control logic to reason about the channel quality and interference experienced by clients and to assign resources accordingly.

The Channel Quality Map abstraction provides network programmers with a full view of the network state in terms of channel quality between clients and APs over the available Resource Blocks. Let $G = (V, E)$ be a directed graph, where $V = V_{AP} \cup V_{STA}$ is the set of APs and wireless clients in the network, and E is the set of edges or links. An edge $e_{n,m,i} \in E$ with $n, m \in V$ exists if m is within the communication range of n over the Resource Block $i \in P$. A weight $q(e_{n,m,i})$ is assigned to each link representing the channel quality between the two nodes.

Notice how the Channel Quality Map is the WiFi-specific instantiation of the Centralised Network Views (CNV) described in Section 5. RSSI is taken as an estimate of the channel quality between two nodes in the wireless network, i.e. between two APs or between wireless stations and APs. The specific Network Information Functions (NIFs) for the Enterprise WLANs scenario are described in the following sub-sections. Notice that these are still preliminary NIFs.

8.4.4 Port

Links in a wired network, e.g., a switched Ethernet LAN, are essentially deterministic and the status of a port in a switch is binary, i.e., active or not active. While some Ethernet switches can select the transmission rate (10, 100, 1000 Mb/s), this feature is aimed at reducing power consumption when the traffic load is low and not as a mechanism for coping with fluctuations in the channel quality. In contrast, links in a wireless network are stochastic and, as a result, the physical layer parameters that characterize the radio link between an LVAP and a WTP, such as transmission power, modulation and coding schemes, and MIMO configuration must be adapted according to the actual channel conditions.

Such level of adaptation requires real-time coordination between LVAPs and WTPs and can only be implemented near the air interface. The Port abstraction allows the C3 to reconfigure or replace a certain transmission control policy if its optimal operating conditions are not met. For each destination address a Transmission control policy (TX Policy) defines the following parameters:

- List of MCS that can be used by the rate control algorithm. Notice that how the MCS are selected within the list is out of the scope of this document. Specifying a single MCS effectively disable rate control.
- The RTS/CTS threshold.
- The ACK policy, i.e. if the WTP shall wait for ACK from the destination station.
- Multicast mode. Applies only to multicast addresses and it is used to specify which type of multicast technique should be applied.
- The number of unsolicited retries for multicast address when operating in UR mode.

Finally, since a Port specifies the configuration of the link between a WTP and a LVAP, a WTP will have at least as many TX Policies as the number of LVAPs it is currently hosting. However, the number could be higher since in there could be also a transmission policy for multicast address and in principle also for broadcast addresses.

8.5 North-bound Interface

The semantic model discussed in the previous section is essentially an object model while the COHERENT DSL presented in this section is a way to configure, i.e. to populate, the semantic model.

The comprehensive list of Python primitives can be found in Table 8-3. Primitives can operate in polling or trigger mode. In the former mode (polling) the C3 periodically polls one or more WTPs for specific information, e.g., the number of packets received by and LVAP. In the latter mode (trigger) a thread is created at one or more WTPs. Such thread is identified by a firing condition, e.g., the RSSI of one LVAP going below a certain threshold. When such condition is verified a message is generated by the WTP. A termination condition can also be specified. All primitives are non-blocking and, as such, they immediately return control to the calling Network App. An optional callback method can be provided specifying a method to be executed when the primitive returns a result. A Python dictionary, whose structure depends on the actual primitive, is passed as parameter to the callback. Moreover, all primitives require at least one parameter named ssid specifying the name of the slice on which they shall operate.

Notice how all primitives in the COHERENT DSL are effectively available as a Python library, i.e. a module in Python terminology, or through the C3 REST API. Both version of the DSL are idempotent. This section will describe the COHERENT DSL from a technology agnostic standpoint leaving the implementation details to deliverable D2.3.

Table 8-3 Programming primitives in the Python-based SDK

Primitive	Parameters	Mode	Description
Packet/bytes counter	lvap, bins, every	Polling	Aggregates TX/RX packets/bytes in the specified bins.
Link statistics	lvap, every	Polling	Fetch ink transmission statistics.
User/Network Channel Quality Map	addrs, block, every	Polling	Fetch user/network channel quality map map.
Summary Trigger	addrs, keep, limit, rates, every	Trigger	Get fine grained link layer events.
RSSI Trigger	addrs, relation, value	Trigger	Callback when a condition is verified.

8.5.1 Light Virtual Access Point

The LVAP is exposed to the programmer through a Python object mapping properties to operations. These properties are: i) the Resource Block(s) on which the LVAP is currently scheduled, ii) the list of Resource Blocks supported by the LVAP and iii) the counters tracking the incoming/outgoing traffic. Such an interface allows programmers to fetch the Resource Block(s) a certain LVAP is currently scheduled at, by accessing the assigned_to property field of an LVAP object. Similarly, performing a handover is as simple as assigning a new list of Resource Blocks to the same field. It is worth stressing that, each Resource Block contains implicitly the information of the WTP at which it is available.

For example, consider the problem of performing a handover in an Enterprise WLANs. In this scenario we have the set of available access points in the network: W_1, W_2, W_N. Each access point supports a set of channels $\omega(W_n)$. Moreover, we have the client L_1 with its set of supported channels $\omega(L_n)$. Finally, let us introduce the function $U_C(l, W_n)$ which returns the signal quality of the client l at the access point W_n on channel c. The handover procedure in such a scenario is implemented by the DSL in Table 8-4. Note that t is a threshold on the RSSI quality (in dB for example).

Table 8-4 Comparison between Internal and External DSLs

<table>
<tr><th>External DSL</th><th>Internal DSL (Python)</th></tr>
<tr><td>input: a set of APs {W_1, ..., W_N},
a Station L_1,
the rssi threshold t

procedure handover ({W_1, ..., W_N}, L_1, t):

Initialise the Resource Pool
$P_N = \omega(W_1) \cup \omega(W_2) \cup \ldots \cup \omega(W_N)$

Select matching elements
$M = P_N \cap \omega(\mathbf{L_1})$

Filter elements by RSSI
$M' = \{b \in M : UCQM(L, b) > t\}$

Perform handover
$\mathbf{L_1} \leftarrow rand(M')$</td><td>def handover(wtps, lvap, t):

    # Initialise the Resource Pool
    pool = ResourcePool()
    for wtp in wtps:
        pool = pool | wtp.supports

    # Select matching elements
    matches = pool & lvap.supports

    # Filter elements by RSSI
    valid = [blk for blk in matches if
             blk.ucqm[lvap]['ewma'] >= t]

    # Perform handover
    blk = valid.pop() if valid else None
    lvap.assigned_to = blk</td></tr>
</table>

The DSL on the left essentially computes the union of all available channels in the network, then the intersection between this set and the set of channel supported by the client L is computed. The intersection is then filtered in order to remove the channel with low RSSI. Finally, one random valid channel is picked from the set.

On the right column of the same type it is reported the same DSL embedded into a host general purpose language, in this case Python. As it can be see the internal DSL follows quite closely the idiom of the internal DSL. Notice that, for the sake of simplicity, in either version of the handover routine error handling has been omitted. For example, if the valid Resource Block set is empty it is up to the application to decide to either lower the RSSI threshold or to handover the LVAP to best available WTP regardless of the link quality.

8.5.2 Channel Quality Map

In our implementation of the CQM we use the RSSI measured at each WTP as an approximation of the channel quality. As defined in Section 5.2, a Network Information Function (NIF) implements functionality for accessing data sources and processing of data. In the WiFi scenario, a NIF is created at every WTP in the network in order to collect such information. The NIF's data source consists of a monitor interface created on top of each physical radio available at a WTP. Such monitor interface is used to extract the signal strength field present in the radiotap header of every decoded WiFi frame.

For each neighbour within the decoding range, the NIF computes the average of the RSSI over a configurable window (e.g., every 500 ms). Moreover, the NIF computes also an N-points smoothed moving average (SMA) of such value. We remind the reader that an N-point SMA is the average of the last N-points of a time series. Standard aging techniques are used when no RSSI samples are collected in the last observation window. The SMA filter has been chosen because it can reduce the noise while being responsive to RSSI changes. Such property is useful when dealing with fast-fading affected RSSI signals. At the same time, fast response allows to promptly react to changes in the channel conditions.

Figure 8-2 sketches a sample channel quality map. Notice how in this case the RSSI has been taken as a measure of channel quality. The map is built using as an input the Wireless LAN radio measurements reports presented in Deliverable D3.1 Section 4.1.2. 3.

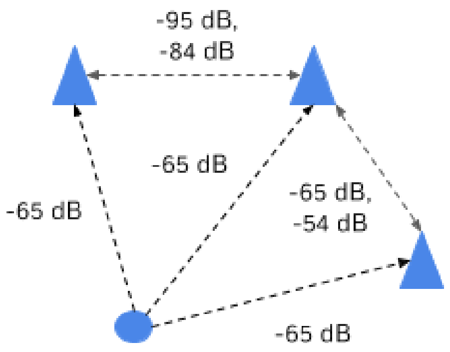

Figure 8-2 An example of RSSI Map

Table 8-5 summarizes the User/Network Channel Quality Map request (the details of the encoding will be reported in D2.3).

Table 8-5 User/Network Channel Quality Map Request

Parameter	Type	Description	Example
Addrs	Ethernet Address	Neighbouring stations/Aps to track	'FF:FF:FF:FF:FF:FF'
block	Resource Block	The channel to monitor	('04:f0:21:09:f9:96', 36, L20)
every	Int	Polling interval (ms)	2000
callback	Function (can be also a remote callback)	Method to be invoked when the response is available.	

In the above example specifying ff:ff:ff:ff:ff:ff will return the RSSI of any station within the decoding range of WTP 04:F0:21:09:F9:96 on channel 36. It is worth noticing that, the Channel Quality Map tracks the RSSI level of any active WiFi device including the ones belonging to networks that are not under the administrative domain of the C3. This includes also wireless clients that are not associated to any network but have their wireless interface active (WiFi clients periodically broadcast Probe Request messages in order to discover available networks).

A sample output of the primitive is reported below as a JSON document. In this case the station a0:d3:c1:a8:e4:c3 is a neighbour of the WTP 04:f0:21:09:f9:96 on the 802.11a channel 36. The report includes, besides the previously described averages, also the total number of frames received since the query was created (hist_packets) together with the average (last_rssi_avg), the standard deviation (last_rssi_std), and the number (last_packets) of RSSI measurements taken during the last observation window.


```
{ "a0:d3:c1:a8:e4:c3": {
              "hist_packets": 15810,
              "last_packets": 10,
              "last_rssi_avg": -79,
              "last_rssi_std": 7,
              "sma_rssi": -82,
              }
}
```


The Channel Quality Map can be accessed at the frame–level granularity providing the network programmer with a real–time picture of all link–layer events. Each WTP tracks the following meta–data associated to link–layer events:

- Transmitter Address. The MAC address of the transmitter.
- TSFT. The 802.11 MAC's 64-bit Time Synchronisation Function Timer. Each frame received by the radio interface is timestamped with a 1μsec resolution clock by the 802.11 driver.
- Sequence, the 802.11 MAC's 16-bit sequence number. This counter is incremented by the transmitter after a successful transmission.
- The frame RSSI (in dB), Rate (in Mb/s), Length (in bytes), and Duration (in μsec). The collected traces are then periodically delivered to the C3 where they are synchronised to a common time reference.

Notice that, only successful unique frames transmissions are recorded, i.e. frames with incorrect checksum and/or PLCP header as well as retransmitted frames are ignored. The collected meta–data can be exploited for several purposes.

Table 8-6 Transmission Summary Request

Parameter	Type	Description	Example
addrs	Ethernet Address	Neighbouring stations/Aps to track	'FF:FF:FF:FF:FF:FF'
limit	Int	Number of transmission reports to be generated	-1
every	Int	Reporting interval (in ms)	2000
block	Resource Block	The channel to monitor	('04:f0:21:09:f9:96', 36, L20)
rates	MCS set	Report only frames transmitted at the specified MCS.	[6, 12, 54]
callback	Function (can be also a remote callback)	Method to be invoked when the response is available.	

The statement above generates a periodic callback when a traffic trace has been received from a WiFi AP. The traffic traces include all the link–layer events within the decoding range of the WiFi AP. The limit of parameters instructs the WiFi AP to send only a specified number of reports after which the operation is stopped and all the allocated data structures on the WiFi AP are freed. Specifying −1 results in the WiFi AP sending traffic traces forever. Specifying ff:ff:ff:ff:ff:ff as addrs will generate transmission summaries for all stations in the network. Notice that the entire data structure containing a frame meta–information is 18–bytes long. A saturated 54–Mb/s channel can deliver up to 2336 frames/s[6]. In such scenario the system would generate 42048 bytes of meta–data per second per radio interface which correspond to a signalling bandwidth between the WTP and C3 of about 336 kb/s which is negligible considering the widespread adoption of Gigabit Ethernet in enterprise networks.

Finally, the rssi primitive allows the programmer to trigger a callback the first time the RSSI of a station verifies a certain condition at any WiFi in network.

Table 8-7 RSSI Trigger Request

Parameter	Type	Description	Example
addrs	Ethernet Address	Neighbouring stations/Aps to track	‘FF:FF:FF:FF:FF:FF’
relation	Enum(‘LT’, ‘GT’, ‘EQ’)	Condition to be verified	‘LT’
value	Int	The RSSI level	-70
block	Resource Block	The channel to monitor	(’04:f0:21:09:f9:96’, 36, L20)
callback	Function (can be also a remote callback)	Method to be invoked when the response is available.	

After the trigger has fired the first time and as long as the RSSI remains below −70dBm, the callback method is not called again by the same WiFi AP, however the same callback may be triggered by other WiFi APs. Specifying ff:ff:ff:ff:ff:ff as lvaps will trigger the callback when the RSSI of any client at any WiFi AP is below −70 dBm.

8.5.3 Link statistics map

Figure 8-3 sketches a sample Link Statistics Maps. This is a variant of the channel quality map where edges are annotated with the link delivery statistics. For each supported MCS (Modulation and coding scheme), the map report the delivery probability, and the link throughput.

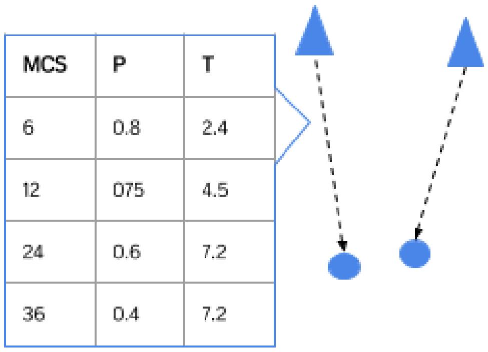

Figure 8-3 An example of Link Statistics Map

For each supported MCS the average delivery probability and the expected throughput in the last observation window are reported. Moreover, also the total number of successful and failed transmissions for the wireless clients is reported.

Table 8-8 Southbound Interface (Link Statistics Map Request)

[6] At 54 Mb/s an 802.11a radio encodes 216 bits/symbol. The error correction encoding then adds 6 more bits at the end of the frame, so a maximum length frame of 1536 bytes becomes a string of 12288 bits plus 6 trailing bits. The total 12294 bits can be encoded with 57 symbols each requiring 6µsec for transmission. As a result, ignoring backoff, the minimum time required for transmitting this frame is: DIFS (34µsec) + DATA (248µsec) + ACK (24µsec) = 322µsec, which corresponds to 2336 frames/second.

Parameter	Type	Description	Example
addrs	Ethernet Address	Neighbouring stations to track	'04:11:af:4f:34:0d'

8.5.4 Traffic matrix

The traffic matrix abstraction allows programmers to track the traffic exchanged by a certain wireless client and to use binning in order to aggregate such information by frame length (useful in wireless networks due to the fact that short packets incur higher transmission overheads). An example for traffic matrix graph is reported in the figure below.

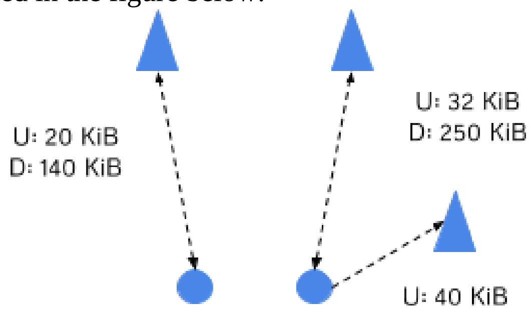

Figure 8-4 An example of Traffic Matrix Map

Table 8-9 Traffic Matrix

Parameter	Type	Description	Example
Addrs	Ethernet Address	Neighbouring stations to track	'04:11:af:4f:34:0d'
Bins	list	Bins to be used for classification.	[512, 1472, 8192]

8.6 Conclusions and Overlook

In this section we introduced the COHERENT Semantic Model for Enterprise WLANs. Moreover, we also provided: (i) a detailed description of the DSL created to access and manipulate such model, and (ii) a description of the northbound interface exposing such DSL.

We consider the COHERENT semantic model relatively stable and we expect that only minor changes will be done to it during the rest of the project. As future work we plan to focus on scalable collection and aggregation of network statistics and to the definition of Slice-specific Network Views on top of the currently Defined CNV.

On the other hand, significant work is currently undergoing for the definition of a semantic model for LTE-based mobile networks and for the associated DSL and northbound interface. We expect the outcome of such work to the reported in the future deliverables.

A Python based SDK mapping the abstractions introduced in the previous section into Python Constructs is also made available to network programmer. The details about the SDK are out of the scope of this document and are instead reported in D2.3. Online resources are available at the official EmPOWER website [107], while the source code (release under an APACHE 2.0 license) can be downloaded from github [108].

9. Conclusions

This report provided a comprehensive and concise overview of the 5G COHERENT architecture and abstractions. We first review main contributions that have been conducted by projects, standardisation bodies and academic literature recently on the topics of emerging technology enablers (e.g., SDN and NFV) and mobile architectures. A set of fundamental requirements for designing programmable 5G architecture was further introduced.

The COHERENT architecture focuses on the control and coordination of radio access functions in the 5G Radio Access Network (RAN). Three key concepts are proposed as the foundations of COHERENT architecture:

- **Control Separation:** By separating the control functionalities of centralised control and real-time control, COHERENT architecture supports dynamic adaptation to rapidly varying wireless networks while getting the benefit of performance gain introduced by centralised control.

- **Network Abstractions:** Abstraction which is a coherent representation of the network state and infrastructure resources is the key to drive programmability in COHERENT control and coordination for 5G RAN systems.

- **Network Slicing**: In the context of the Network-as-a-Service (NaaS) business model, the network slicing in COHERENT architecture enables the possibility of multi-tenancy operations on a common physical infrastructure by further abstracting the physical network into network specific slices possibly operated by different MVNOs, which can be considered as independent virtualised end-to-end networks.

We showed that a variety of use cases, including (but not limited to) multi-tenancy, Device-to-Device (D2D) communications, broadcast operation, UE-relaying operation for public safety service, multi-connectivity and mobility management could be applied to the COHERENT architecture. Moreover, since COHERENT intends to synchronise with current 3GPP advancements for 5G, we have provided in this report the contribution of COHERENT to the 3GPP framework based on COHERENT principles. Finally, we have devised the network abstractions, semantic models and northbound interfaces in the COHERENT system for programmable heterogeneous mobile networks.

Bibliography

[1] Ericsson Technical White paper, “5G systems – enabling industry and society transformation,” 2015.

[2] *5G White Paper*, white paper, NGMN Alliance, 2015.

[3] “CROWD project website,” [Online]. Available: http://www.ict-crowd.eu/.

[4] “FED4FIIRE project website,” [Online]. Available: http://www.fed4fire.eu/.

[5] “FLEX project website,” [Online]. Available: http://www.flex-project.eu/.

[6] “VITAL project website,” [Online]. Available: http://www.ict-vital.eu/.

[7] “ADEL project website,” [Online]. Available: http://www.fp7-adel.eu/.

[8] 5G PPP Architecture WG, “5G PPP View on 5G Architecture,” 2016. [Online]. Available: https://5g-ppp.eu/white-papers. [Accessed June 2016].

[9] 5G-Xhaul website, [Online]. Available: http://www.5g-xhaul-project.eu/.

[10] “5G-Xhaul website,” [Online]. Available: http://www.5g-xhaul-project.eu/.

[11] 5G-Crosshaul project, [Online]. Available: http://5g-crosshaul.eu/.

[12] Josep Mangues Bafalluy, “Millimetre wave in 5G Crosshaul,” in *Workshop on millimetre wave Technology for High speed Broadband Wireless Networks*, Valencia, 2015.

[13] CHARISMA website, [Online]. Available: http://www.charisma5g.eu/.

[14] E. Escalona, “Converged Heterogeneous Advanced 5G Cloud-RAN Architecture for Intelligent and Secure Media Access,” in *EuCNC*, 2015.

[15] SELFNET website, [Online]. Available: https://selfnet-5g.eu/.

[16] Pedro Neves, et.al, “The SELFNET Approach for Autonomic Management in an NFV/SDN Networking Paradigm,” in *International Journal of Distributed Sensor Networks*, 2016.

[17] 5G-NORMA website, [Online]. Available: https://5gnorma.5g-ppp.eu.

[18] H2020-ICT-2014-2 5G NORMA, “Deliverable D3.1 Functional Network Architecture and Security Requirements,” 2015.

[19] “Sesame website,” [Online]. Available: http://www.sesame-h2020-5g-ppp.eu/.

[20] Ioannis Giannoulakis, George Xylouris, Emmanouil Kafetzakisy, Michail Angelos Kourtis, Jose Oscar Fajardoz, Pouria Sayyad Khodashenasx, Antonino Albanese, Haralambos Mouratidis and Vassilios Vassilakis , “System architecture and aspects of SESAME: Small cEllS coordinAtion for Multi-tenancy and Edge services,” in *IFIP/IEEE Soft5G*, Seoul, 2016.

[21] “5GEX website,” [Online]. Available: http://www.5gex.eu/.

[22] A. 5. W. Paper, “5GEx Multi-domain Service Creation - from 90 days to 90 minutes,” 2016.

[23] Carlos J. Bernardos, Olivier Dugeon, Alex Galis, Donal Morris, Csaba Simon, Robert Szab, “5G Exchange (5GEx) – Multi-domain Orchestration for Software Defined Infrastructures,” in *EuCNC*, 2015.

[24] “Fanastic website,” [Online]. Available: http://fantastic5g.eu/.

[25] Frank Schaich, Berna Sayrac, Martin Schubert, Hao Lin, Klaus Pedersen, Musbah Shaat, Gerhard Wunder and Andreas Georgakopoulos, “FANTASTIC-5G: 5G-PPP Project on 5G Air Interface Below 6 GH,” in *EuCNC*, 2015.

[26] 3GPP TS 36.300 (v10.8.0), *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10)*, 2012.

[27] 3GPP TS 23.303, *Proximity-based services (ProSe); Stage 2*, 2016.

[28] 3GPP TR 23.703, “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP),” 2016.

[29] 3GPP TR 22.891, “Study on New Services and Markets Technology Enablers,” 2016.

[30] 3GPP TR 23.799, “Study on Architecture for Next Generation System,” [Online]. Available: http://www.3gpp.org/DynaReport/23799.htm.

[31] ETSI, G. 001 (MANO), V1.1.1, “Network Functions Virtualisation (NFV); Management and Orchestration,” 2014.

[32] “OPNFV,” [Online]. Available: https://www.opnfv.org/.

[33] “OpenBaton,” [Online]. Available: http://openbaton.github.io/.

[34] “OpenMANO,” [Online]. Available: https://github.com/nfvlabs/openmano.

[35] “OpenFlow,” [Online]. Available: https://www.opennetworking.org/sdn-resources/openflow.

[36] C. Kolias, “OpenFlow-Enabled Mobile and Wireless Networks,” Open Networking Foundation, 2013.

[37] L. Yang, R. Dantu, T. Anderson and R. Gopal, “RFC3746: Forwarding and Control Element Separation (ForCES) Framework,” IETF, 2004.

[38] Evangelos Haleplidis, Kostas Pentikousis,Spyros Denazis and Odysseas G. Koufopavlou, “RFC 7426: Software-Defined Networking (SDN): Layers and Architecture Terminology,” IETF, 2015.

[39] D. Kreutz, F. M. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky and S. Uhlig, “Software-defined networking: A comprehensive survey,” in *IEEE*, 2015.

[40] B. A. Nunes, M. Mendonca, X. N. Nguyen, K. Obraczka and T. Turletti, “A survey of software-defined networking: Past, present, and future of programmable networks,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617-1634, 2014.

[41] N. M. K. Chowdhury and R. Boutaba, “A survey of network virtualisation,” *Computer Networks*, pp. 862-876, 2010.

[42] A. Gudipati, D. Perry, L. E. Li and S. Katti , “SoftRAN: Software defined radio access network,” in *the second ACM SIGCOMM workshop on Hot topics in software defined networking*, 2013.

[43] C. Liang and F. R. Yu, “Wireless network virtualisation: A survey, some research issues and challenges,” *IEEE Communications Surveys & Tutorials*, pp. 358-380, 2015.

[44] G. Hampel, M. Steiner and T. Bu, “Applying software-defined networking to the telecom domain,” in *IEEE Conference on Computer Communications Workshops*, 2013.

[45] J. Costa-Requena , “SDN integration in LTE mobile backhaul networks,” in *Information Networking (ICOIN)*, 2014.

[46] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng, “OpenRAN: a software-defined ran architecture via virtualisation,” *ACM SIGCOMM computer communication review*, vol. 43, no. 4, pp. 549-550, 2013.

[47] H. Ali-Ahmad, “CROWD: an SDN approach for DenseNets,” in *Second European Workshop on Software Defined Networks (EWSDN)*, 2013.

[48] V. Sagar, R. Chandramouli, and K. P. Subbalakshmi, “Software defined access for HetNets,” *IEEE Communications Magazine*, 2016.

[49] V. G. Nguyen, T. X. Do, and Y. Kim, “SDN and virtualisation-based LTE mobile network architectures: A comprehensive survey,” *Wireless Personal Communications*, 2016.

[50] M. R. Sama, S. Ben Hadj Said, K. Guillouard, and L. Suciu, “Enabling network programmability in lte/epc architecture using openflow.,” in *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2014.

[51] R. Wakikawa and S. Matsushima, “Stateless user-plane architecture for virtualised EPC,” IETF draft, 2015.

[52] A. Patro and S. Banerjee, “COAP: A software-defined approach for home WLAN management through an open API,” in *ACM SIGMOBILE Mobile Computing and Communications Review*, 2015.

[53] R. Trivisonno, R. Guerzoni, I. Vaishnavi and D. Soldani, “SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility,” *Transactions on Emerging*

Telecommunications Technologies, pp. 82-92, 2015.

[54] R. Riggio, K. M. Gomez, T. Rasheed, J. Schulz-Zander, S. Kuklinski and M. K. Marina, "Programming Software-Defined wireless networks," in *10th International Conference on Network and Service Management (CNSM) and Workshop*, 2014.

[55] R. Riggio, M. K. Marina, J. Schulz-Zander, S. Kuklinski and T. Rasheed, "Programming Abstractions for Software-Defined Wireless Networks," *IEEE Transactions on Network and Service Management,* vol. 12, no. 2, pp. 146-162, 2015.

[56] R. Riggio, A. Bradai, D. Harutyunyan, T. Rasheed and T. Ahmed , "Scheduling Wireless Virtual Networks Functions," *IEEE Transactions on Network and Service Management,* vol. 1, no. 99, pp. 1-1, 2016.

[57] R. Riggio, A. Bradai, T. Rasheed, J. Schulz-Zander, S. Kuklinski and T. Ahmed, "Virtual network functions orchestration in wireless networks," in *11th International Conference on Network and Service Management (CNSM),* 2015.

[58] M. M. Rahman, C. Despins and S. Affes, "HetNet Cloud: Leveraging SDN & Cloud Computing for Wireless Access Virtualisation," in *IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB),* 2015.

[59] Liu, Cheng, et al., "The case for re-configurable backhaul in cloud-RAN based small cell networks.," in *IEEE INFOCOM*, 2013.

[60] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M.S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—a technology overview," *IEEE Communications Surveys & Tutorials,* pp. 405-426, 2015.

[61] Chia-Yu Chang, Ruggero Schiavi, Navid Nikaein, Thrasyvoulos Spyropoulos and Christian Bonnet, "Impact of packetization and functional split on C-RAN fronthaul performance," in *IEEE International Conference on Communications*, 2016.

[62] Hwang, Insoo, Bongyong Song, and Samir S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks.," *IEEE Communications Magazine,* vol. 50, no. 6, pp. 20-27, 2013.

[63] R. Vilalta , et al., "The need for a control orchestration protocol in research projects on optical networking.," in *European Conference on Networks and Communications (EuCNC),* 2015.

[64] Mamta Agiwal, Abhishek Roy and Navrati Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials,* 2016.

[65] K.-K. Yap, M. Kobayashi, R. Sherwood, T.-Y. Huang, M. Chan, N. Handigol, and N. McKeown, "OpenRoads: Empowering Research in Mobile Networks," *ACM SIGCOMM Computer Communication Review,* vol. 40, no. 1, p. 125–126, 2010.

[66] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "CellSlice: Cellular Wireless Resource Slicing for Active RAN Sharing," in *Communication Systems and Networks (COMSNETS),* 2013.

[67] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and Flexible Cellular Core Network Architecture," in *9th ACM conference on Emerging Networking Experiments and Technologies*, 2013.

[68] K. Pentikousis, Y. Wang, and W. Hu, "MobileFlow: Toward Software-defined Mobile Networks," *IEEE Communications Magazine,* vol. 51, no. 7, p. 44–53, 2013.

[69] J. Kempf, B. Johansson, S. Pettersson, H. Luning, and T. Nilsson, "Moving the Mobile Evolved Packet Core to the Cloud," in *Wireless and Mobile Computing, Networking and Communications (WiMob) 2012,* 2012.

[70] C. Bernardos, A. La Oliva, P. Serrano, A. Banchs, L. M. Contreras, H. Jin, and J. C. Zuniga, "An Architecture for Software Defined Wireless Networking," *IEEE Wireless Communications,* vol. 21, no. 3, pp. 52-61, 2014.

[71] M. Arslan, K. Sundaresan, and S. Rangarajan, "Software-defined Networking in Cellular Radio Access Networks: Potential and Challenges," *IEEE Communications Magazine,* vol. 53, no. 1, pp. 150-156, 2015.

[72] G. Bianchi, P. Gallo, D. Garlisi, F. Giuliano, F. Gringoli, and I. Tinnirello, "MAClets: Active MAC Protocols over Hard-coded Devices," in *8th international conference on Emerging networking experiments and technologies,* 2012.

[73] V. Nguyen et al., "SDN and virtualisation-based LTE mobile network architectures: A comprehensive survey," *Wireless Personal Communications,* vol. 86, no. 3, pp. 1401-1438, 2016.

[74] F. Ahmed et al., "Distributed Graph Coloring for Self-Organization in LTE Networks," *Journal of Electrical and Computer Engineering,* 2010.

[75] P. Cardieri, "Modeling interference in wireless ad hoc networks," *IEEE Communication Surveys & Tutorials,* vol. 12, no. 4, p. 551–572, 2010.

[76] R. Riggio, A. Bradai, D. Harutyunyan, T. Rasheed and T. Ahmed, "Scheduling Wireless Virtual Networks Functions," *IEEE Transactions on Network and Service Management,* vol. 1, no. 99, pp. 1-1, 2016.

[77] IETF RFC 7752, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP," 2016.

[78] IETF RFC 5440, "Path Computation Element (PCE) Communication Protocol (PCEP)," 2009.

[79] IETF RFC 6241, "Network Configuration Protocol (NETCONF)," 2011.

[80] IETF RFC 6020, "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)," 2010.

[81] IETF RFC 1157, "A Simple Network Management Protocol (SNMP)," 1990.

[82] IETF RFC 6830, "The Locator/ID Separation Protocol (LISP)," 2013.

[83] IETF RFC 7047, "The Open vSwitch Database Management Protocol," 2013.

[84] IETF RFC 5415, "Control And Provisioning of Wireless Access Points (CAPWAP) Protocol Specification," 2009.

[85] IETF RFC 5412, "Lightweight Access Point Protocol," 2010.

[86] Roy T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures: Chap 5 Representational State Transfer (REST)," 2010.

[87] Patel et al., "Mobile-Edge Computing – Introductory Technical White Paper," *ETSI,* 2014.

[88] "Apache Cassandra database," [Online]. Available: http://cassandra.apache.org/.

[89] "Neo4j information from Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Neo4j.

[90] K. Alexandris, N. Nikaein, R. Knopp and C. Bonnet, "Analyzing X2 handover in LTE/LTE-A," in *Wireless Network Measurements and Experimentation (WINMEE),* 2016.

[91] 3GPP TS 36.331 v10.6.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," 2012.

[92] Navid Nikaein, et al., "Network store: Exploring slicing in future 5g networks," in *Proceedings of the 10th International Workshop on Mobility in the Evolving Internet Architecture, ACM,* 2015.

[93] Katsalis et al., "Architectural Design Patterns for the RAN," in *IEEE ICC, 3rd International Workshop on 5G Architecture,* 2016.

[94] L. Duan, L. Gao and J. Huang, "Cooperative spectrum sharing: a contract-based approach," *IEEE Transactions on Mobile Computing,* vol. 13, no. 1, pp. 174-187, 2014.

[95] Chun, Sung Hyun, and Richard J. La., "Secondary spectrum trading: auction-based framework for spectrum allocation and profit sharing," *IEEE/ACM Transactions on Networking (TON),* vol. 21, no. 1, pp. 176-189, 2013.

[96] 3GPP TR25.892 v2.0.0, "Feasibility study for Orthogonal Frequency Division Multiplexing (OFDM) for UTRAN enhancement," 2004.

[97] 3GPP TR 36.843, "Study on LTE device to device proximity services; Radio aspects," 2014.

[98] CPRI Specification V6.0 (2013-08-30), "Common Public Radio Interface (CPRI); Interface Specification," 2013.

[99] "CPRI Specification Overview," [Online]. Available: http://www.cpri.info/spec.html.

[100] Anna Zakrzewska, et al., "Dual connectivity in LTE HetNets with split control-and user-plane," in *IEEE Globecom Workshops (GC Wkshps)*, 2013.

[101] R3-161120, "5G access architecture with UP/CP separation," 3GPP RAN, 2016.

[102] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies," 2016.

[103] 3GPP TR 38.801, v0.2.0, "Study on New Radio Access Technology: Radio Access Architecture and Interfaces," 2016.

[104] 3GPP TS 29.281, "General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)," 2016.

[105] 3GPP TS 36.413 , "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP)," 2016.

[106] 3GPP TS 36.423, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP)," 2015.

[107] "EmPOWER website," [Online]. Available: http://empower.create-net.org/.

[108] "Source code of EmPOWER," [Online]. Available: https://github.com/5g-empower/.

[109] "Charisma website," [Online]. Available: http://www.charisma5g.eu/.

[110] K.-K. Yap, M. Kobayashi, R. Sherwood, T.-Y. Huang, M. Chan, N. Handigol, and N. McKeown, "OpenRoads: Empowering Research in Mobile Networks," *ACM SIGCOMM Computer Communication Review,* vol. 40, no. 1, p. 125–126, 2010.

[111] 3GPP TS 36.300, v13.3.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," 2016.

[112] 3GPP, "3GPP TS 36.423, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP)," 3GPP, Release 13," 2015.

Annex A. Supplement Materials

A.1 LTE Radio Protocol Stack[26]

The content including the figures below is excerpted from [26] about LTE Radio Protocol Stack:

Figure A-1 shows the protocol stack for the user-plane, where PDCP, RLC and MAC sublayers (terminated in eNB on the network side) perform the header compression, ciphering, scheduling, ARQ and HARQ.

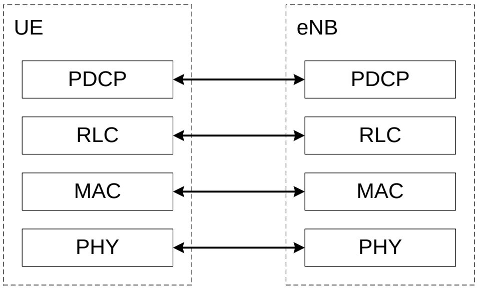

Figure A-1 User-plane protocol stack [26]

Figure A-2 shows the protocol stack for the control-plane, where:

- PDCP sublayer (terminated in eNB on the network side) performs the functions of ciphering and integrity protection;
- RLC and MAC sublayers (terminated in eNB on the network side) **perform the same functions as for the user plane;**
- RRC (terminated in eNB on the network side) performs the following functions:
  - Broadcast;
  - Paging;
  - RRC connection management;
  - RB control;
  - Mobility functions;
  - UE measurement reporting and control.
- NAS control protocol (terminated in MME on the network side) performs among other things:
  - EPS bearer management;
  - Authentication;
  - ECM-IDLE mobility handling;
  - Paging origination in ECM-IDLE;
  - Security control.

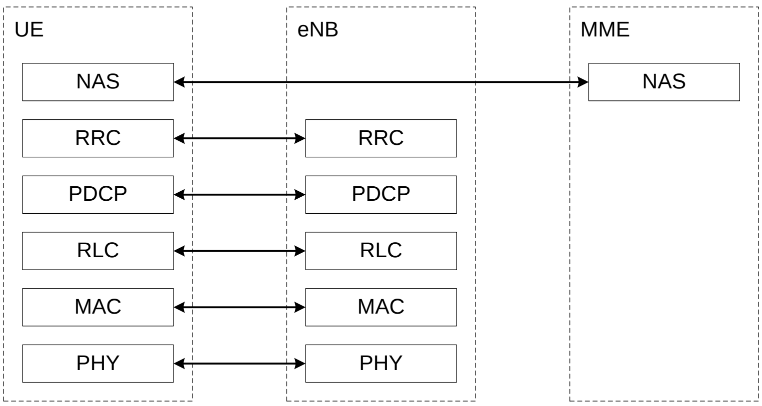

Figure A-2 Control-plane protocol stack [26]

A.2 Functional Split between Radio Access and Core Networks[26]

The content including the figure below is excerpted from [26] about Functional Split between radio access core networks:

The functional split is summarised in the figure below where yellow boxes depict the logical nodes, white boxes depict the functional entities of the control plane and blue boxes depict the radio protocol layers.

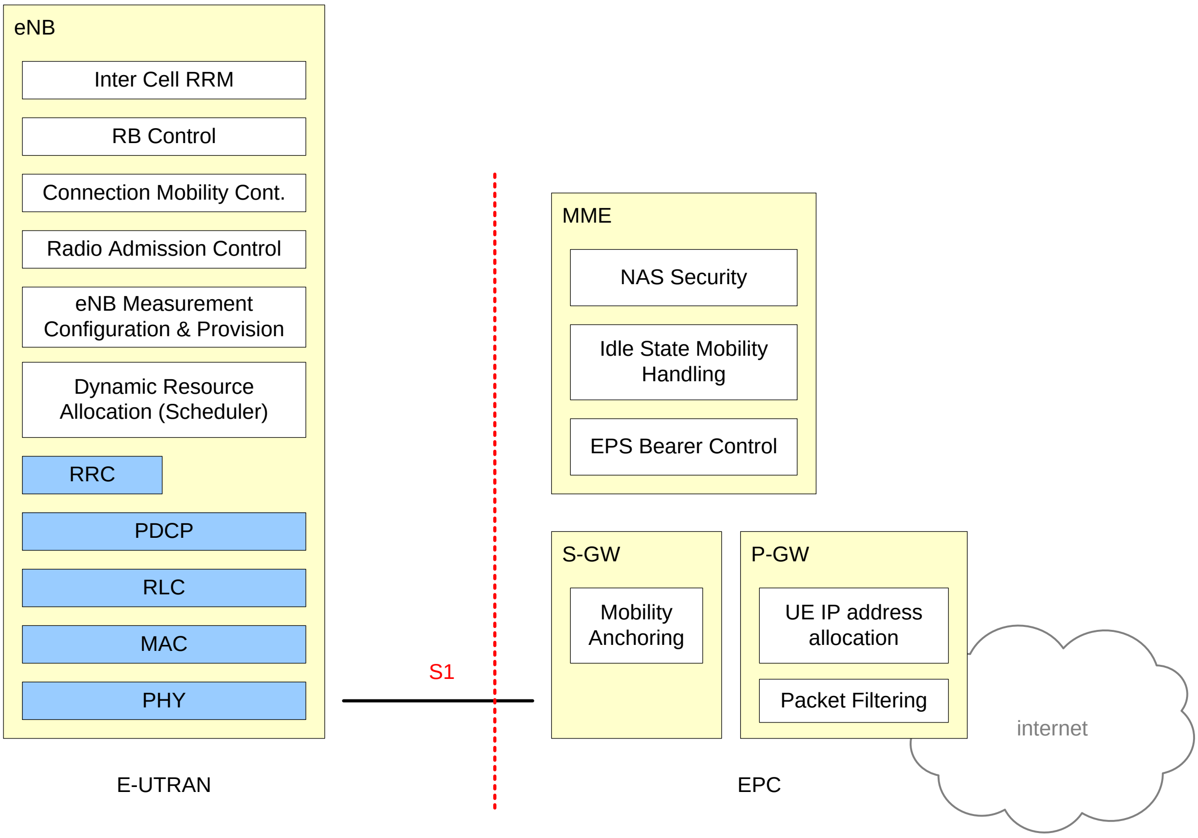

Figure A-3 Functional Split between E-UTRAN and EPC [26]

A.3 3GPP Radio Protocol Stack for D2D and UE-to-Network Relay

In TS 23.303 we can find the protocol stack for Prose UEs and Remote UEs:

A) Control Plane (CP) Protocol Stack;

B) User Plane (UP) Protocol Stack.

The control plane stack A) consists of protocols for control and support of the user plane functions:

- controlling the configuration of the ProSe-enabled UE;
- controlling ProSe Direct Discovery;
- controlling the set-up of the connection between the Remote UE and the ProSe UE-to-Network Relay; For example, according to the Model A "I am here", or Model B "who is there?"/"are you there?", and the Discovery Announcement/Response messages, a Relay UE keeps performing the measurement of the signal strength and can select a UE-to-Network Relay for relay reselection as defined in [26].
- controlling the attributes of an established network access connection, such as activation of an IP address.

Moreover, with respect to CP & UP protocol stack A) and B):

The control can be provided through two different interfaces: PC5 for Signalling Protocol between two ProSe UEs, and PC3 for Signalling Protocol and related control between a ProSe UE and the ProSe Function, as represented below in Figure A-4 and Figure A-5.

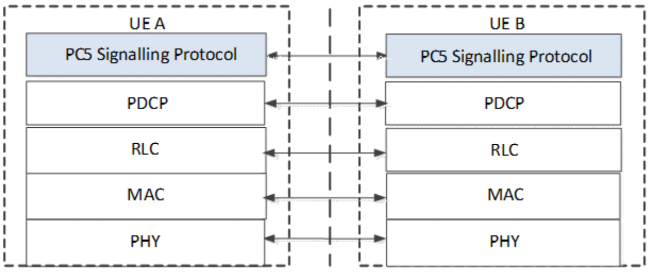

Figure A-4 Prose UEs PC5 Signalling Protocol

PC5 Signalling Protocol is used for control plane signalling over PC5 (e.g., establishment, maintenance and release of secure layer-2 link over PC5, TMGI monitoring requests, Cell ID announcement requests etc.). Some examples of PC5 signalling messages are described in TS 24.334 (CT1) namely: DIRECT_COMMUNICATION_REQUEST/REJECT/KEEPALIVE/RELEASE, TMGI_MONITORING, DIRECT_SECURITY_MODE_COMMAND, and others.

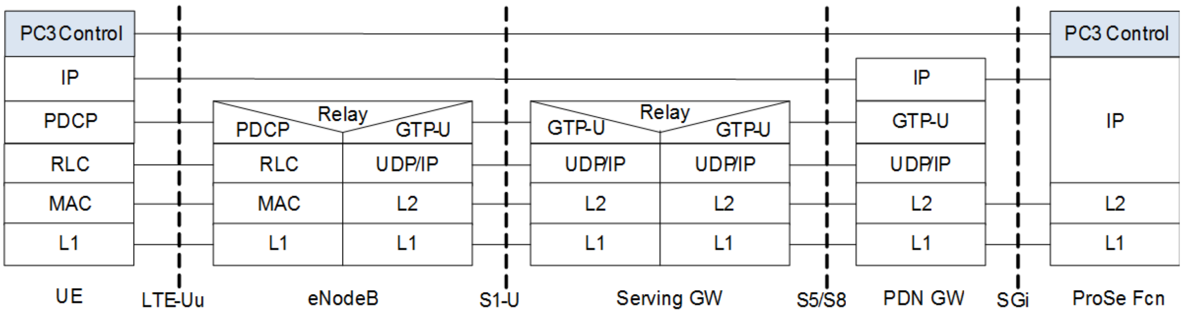

Figure A-5 Prose UE PC3 Signalling Protocol

ProSe Control Signalling between UE and ProSe Function is carried over the user plane and is specified in TS 24.334 (CT1) and it describes specific procedures for service announcements, discovery and security (which is also treated by SA3 TS 33.303, with respect to exact procedures and security flows). PC3 may be realised with one or more protocols.

The user plane can be provided through PC5-U for both UE-to-UE communication and Remote UE to Network communication as represented below in Figure A-6 and Figure A-7.

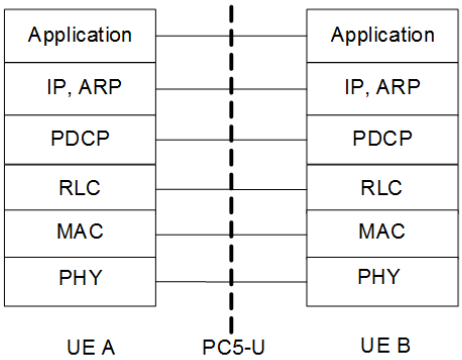

Figure A-6 User Plane for (two) Prose UEs

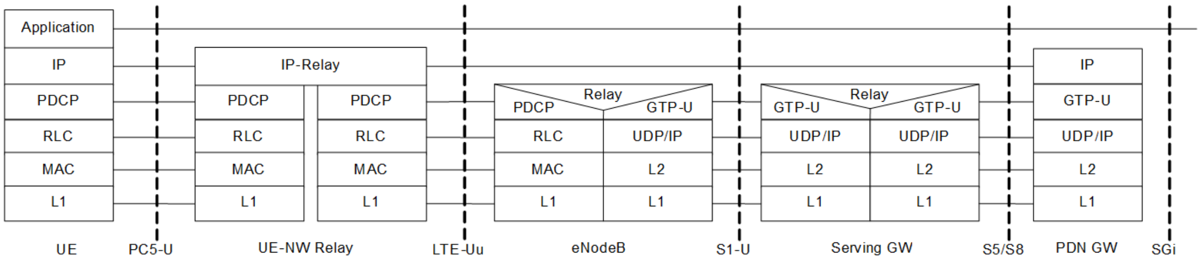

Figure A-7 User Plane for a Remote UE connected to the Network through a UE-to-Network Relay

The application level of Remote UE is transparent for UE-to-Network Relay, which is connected to the base station through the classic interface Uu.

A.4 3GPP Advancement with Respect to 5G

In order to better integrate future COHERENT work and to adapt it to current industrial needs, it is important to understand 3GPP working method and current advancements. Since there were a lot of advancement at 3GPP level with respect to new services and new requirements, this section tries to illustrate the latest 5G technology focus and a tentative 5G roadmap.

One of the responsibilities of 3GPP is to produce Technical Reports (TRs) and Technical Specifications (TSs) on evolution of existing radio technologies or new radio technologies defined by 3GPP, such as HSPA/HSPA+, LTE and LTE-Advanced (LTE-A). 3GPP specifications are grouped into "Releases", and a mobile system can be implemented based on the set of all specifications which comprise a given Release. Whenever a new feature is required by the market, it is proposed to be specified in a given release. Currently 3GPP is working on LTE-A Rel-14, intended to be functionally frozen in June 2017.

However, as seen as it is now, we expect to talk about a complete 5G normative work not earlier than 2019 and a 5G deployment not earlier than 2020. 5G will most probably correspond to Release-15 and Release-16, as we will further explain and show. It is very important to explain these aspects, because they are closely related to COHERENT work on architecture. Since 3GPP work on system architecture just started, we believe that COHERENT architecture work is synchronised with current 3GPP advancements for 5G and we believe that we can provide significant inputs to 3GPP, as a result of collaboration between partners. But first, we need to understand what 3GPP working methods are and how they are related to COHERENT.

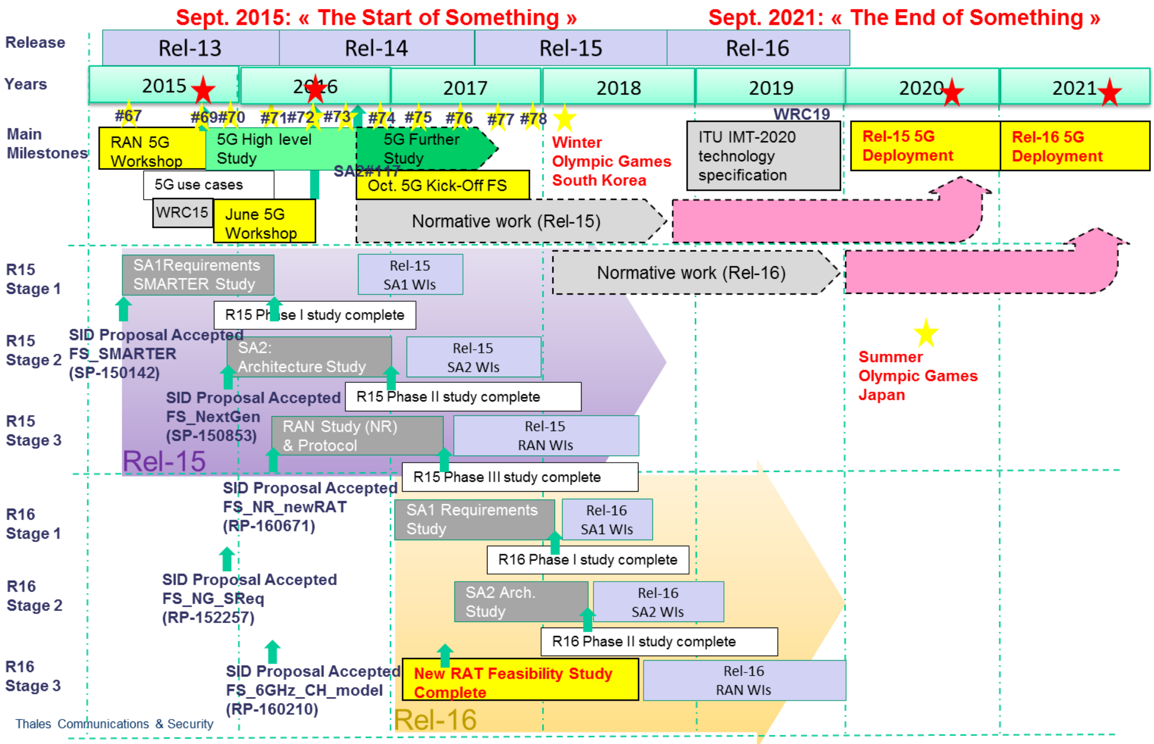

Figure A-8 3GPP Advancements towards a 5G Architecture and Expected 5G Network Deployment

First, at the plenary meeting #67 in 2015 it has been approved a Study Item Description (or SID) for SA1 requirements, which led to the creation of the Technical Report (TR) 22.891 on Study on New Services and Markets Technology Enablers - usually called SA1 SMARTER[7]. In order to better justify the need and system evolution towards 5G deployment (and since the beginning of 5G-related work), equipment providers and operators searched for new enablers, new services and new markets justifying 5G. As a result of this huge “brainstorm”, TR 22.891 now provides a complete list of use cases and services which is more or less surprising. As presented in Sept. 2015 as “the start of something”, several use cases were emphasized at 3GPP RAN 5G workshop and priority was given especially to:

- enhanced mobile broadband;
- massive machine type communications;
- ultra-reliable and low latency communications.

With its SMARTER initiative, SA1 working group defined a total of 74 use cases with extremely different and sometimes conflicting requirements. As an example, some applications require increased spectral efficiency; some of them increased robustness and reliable synchronisation, and some of them asynchronous transmissions.
An initial estimation of 5G evolution, inspired from its usecases and new services shows that 5G is composed from (at least) :

- 24,78% of Critical Communications (CriC) subjects;
- 17,56% of massive Internet of Things (mIOT) subjects;
- 12,62% of enhanced Mobile Broadband (eMBB) subjects;
- 44,37% of Network Operation (NEO) subjects.

We could have expected the first 3 main components, as it was explained during plenary meeting #69 in September 2015, but NEO percentage is somewhat surprising leading to more than 40% of estimated work. It therefore seems that at the time being a lot of work has to be done on network side operation for 5G, which justifies COHERENT principles in order to e.g., achieve adaptation to

[7] http://www.3gpp.org/DynaReport/22891.htm

heterogeneous network infrastructure, dynamic adjustment to network changes, and flexible spectrum usage. This can be explained by several ways: firstly, due to the important activities in the radio access area related to CriC, mIOT and eMBB, 3GPP TSG RAN will propose to reduce the scope of New Radio (NR) Waveforms and Next Generation (NG) New RAT to OFDM-like and SC-FDMA improved waveforms; secondly, there is a huge willingness to evolve towards a solution with backward compatibility, which means that several systems (LTE, LTE-A, 3G, 3G+, 2G, WiFi) should coexist together in the same network (including new 5G and old 4G/3G/2G services), which for SA and CT work groups it means a huge overload situation.

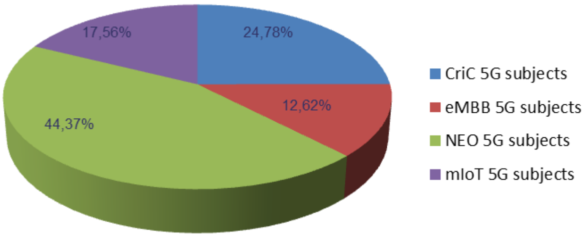

Figure A-9 5G Current Composition

As also shown in the figure describing current 3GPP advancements toward a 5G architecture, 3GPP has a method composed from 3 main steps: service aspects and network requirements covered by Stage-1, architecture completion covered by Stage-2 and completion of the protocols covered by Stage-3. For Release-15 the 1st Stage study is now complete, but the completion of the 2nd Stage with the architecture study would then be finalised by December 2016 and the completion of the 3rd Stage with protocols study should be normally finished by March 2017 for both Core Network (CN) and RAN-related aspects. But this covers only the work on the study items, since the normative work will probably begin by the end of 2016 with Work Items (WIs) SA1 to be completed by June 2017, WIs SA2 to be completed by December 2017 and WIs RAN to be completed by June 2018. As usual for defining complex new features, a two-step approach will be performed. In a first step, as a result of approved Study Items Descriptions (SIDs), a Technical Report (TR) document proposing different solutions would be defined. Then, in a second step, the best solution would be chosen, and as a result of approved Work Item Descriptions (WIDs) would lead to the definition of Technical Specification (TS) documents that will become in time the standards to be used by equipment providers and network operators. For all above mentioned reasons it is more and more clear that a first 5G expected deployment will be in 2020, just in time for 2020 Summer Olympic games in Japan. Similar expectations are for 5G Release-16, but with a potential deployment by the end of 2021.

It is important to mention that, currently SA1 work activated other work groups from SA such SA2 and SA3 and some others from RAN such as RAN1 and RAN2, which clearly shows that 5G has just started and is in line with expected above mentioned roadmap. The current 5G active work groups are represented below:

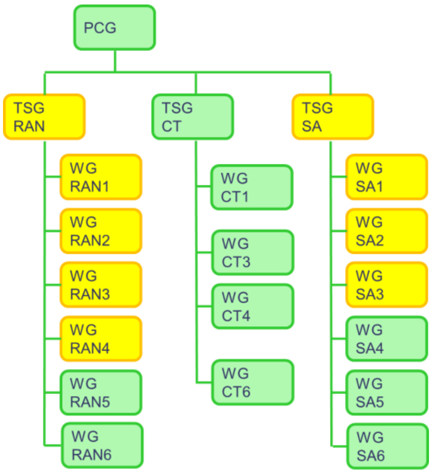

Figure A-10 Structure of 3GPP and Active 5G Groups as in May 2016

As being explained in Figure A-10, the highest decision making group in 3GPP is the Project Coordination Group (PCG) being for example responsible for final adoption of 3GPP Work Items (WIs), ratifying election results and resources committed to 3GPP. Before 2016 there were 4 Technical Specification Groups (TSGs) namely GERAN (GSM EDGE Radio Access Networks), RAN (Radio Access Networks), CT (Core Network & Terminals) & SA (Service & Systems Aspects) but since March 2016 it has been decided to regroup GERAN activities into RAN (GERAN1 and GERAN2 in RAN6, and GERAN3new in RAN5), which is a very good news since in this way 3GPP reorganises and therefore is able to liberate resources for e.g., 5G related work.

Without entering too much into details, TSG RAN is responsible for the definition of the functions, requirements and interfaces of UTRA/E-UTRA network; TSG CT is responsible for specifying terminal interfaces (logical and physical) and terminal capabilities (such as execution environments) and the core network part of 3GPP systems; and TSG SA is responsible for the overall architecture and service capabilities of systems based on 3GPP specifications, but also has the responsibility for cross TSG coordination. Therefore, it is clear why the first Study Item on 5G was first initiated by SA TSG and more precisely with SA1 work which is responsible of new features, new services, charging requirements and new system and service capabilities.

As previously mentioned, SA initiated the work with SA1 SMARTER TR 22.891, and the work propagated into other Work Groups from SA namely SA2 (responsible of system architecture) with a new SID FS_NextGen and TR 23.799 « Study on Architecture for Next Generation System" [8], and SA3 (responsible of security) with a new SID FS_NSA and TR 33.899 « Study on the security aspects of the next generation system" [9]. Moreover, SA1 work led to the creation of 4 other Technical Reports:

- TR 22.862 (as a result of FS_SMARTER-CRIC) on «FS_SMARTER - Critical Communications»[10]
- TR 22.861 (as a result of FS_SMARTER-mIoT) on «FS_SMARTER - massive Internet of Things»[11]
- TR 22.863 (as a result of FS_SMARTER-eMBB) on «FS_SMARTER - enhanced Mobile Broadband» [12]
- TR 22.864 (as a result of FS_SMARTER-NEO) on «FS_SMARTER - Network Operation»[13]

[8] http://www.3gpp.org/DynaReport/23799.htm
[9] http://www.3gpp.org/DynaReport/33899.htm
[10] http://www.3gpp.org/DynaReport/22862.htm
[11] http://www.3gpp.org/DynaReport/22861.htm
[12] http://www.3gpp.org/DynaReport/22863.htm
[13] http://www.3gpp.org/DynaReport/22864.htm

Currently there is no TR on eV2X (enhanced Vehicle-to-X communication) but the use cases included here can be partially integrated in the others SA1 TRs. However, the documentation of security considerations related to them is for Further Study (FS).

As a result of the work in TSG SA, several working groups are also conducting a feasibility studies, namely:

- RAN with TR 38.913 [14] on "Study on Scenarios and Requirements for Next Generation Access Technologies" (as a result of SID FS_NG_SReq), and TR 38.912[15], "Study on New Radio (NR) Access Technology" (as a result of SID FS_NR_newRAT) – for the later a number is reserved but no document is available so far;
- RAN1 being responsible of Layer 1 function, with TR 38.900[16] on "Study on channel model for frequency spectrum above 6 GHz" (as a result of SID FS_6GHz_CH_model) and TR 38.802[17] on "Study on New Radio Access Technology Physical Layer Aspects";
- RAN2 being responsible of Layer 2 and Layer 3 functions, with TR 38.804[18] or "TR for Study on New Radio Access Technology Radio Interface Protocol Aspects" a number is reserved but no document is available so far;
- RAN3 being responsible of S1 and X2 interfaces, with TR 38.801[19] on "Study on New Radio Access Technology: Radio Access Architecture and Interfaces";
- RAN4 being responsible of radio performance and protocol, with TR 38.803[20] on "TR for Study on New Radio Access Technology: RF and co-existence aspects".

A lot of Technical Reports are on-going, but no Technical Specifications (TS) are available so far. Seeing the current 3GPP available work it is expected that 5G normative work will begin in October-November 2016, first with SA1 approved Work Item Descriptions (WIDs) and normative related work and then with SA2 and RAN normative work. The normative Release-15 5G work is therefore expected to finish before end of 2018 and 5G deployment should be therefore available before the Summer Olympic games from Japan in 2020. The timing is therefore perfect for COHERENT in order to provide significant 3GPP contributions.

Some of the considered use cases are totally new and never seen before, which is actually a huge step for 3GPP work seen now as "revolution" and not only as a simple "evolution". Some of the use cases refer to connectivity of drones (SMARTER UC12 & UC54), some of them to moving ambulance and telemedicine support (SMARTER UC65 & UC68), some of them to connectivity under high speed scenarios (SMARTER UC66) and some of them even to connectivity using satellites (SMARTER UC72). This has never seen before for legacy cellular communication and therefore it is perceived as a system revolution. Here are some relevant use cases extracted from TR 22.861, TR 22.862, TR 22.863, TR 22.864:

- SMARTER UC1 Ultra Reliable Communication;
- SMARTER UC2 Network Slicing;
- SMARTER UC3 Lifeline Communications / Natural Disaster;
- SMARTER UC12 Connectivity for Drones;
- SMARTER UC31 Temporary Service for Users of Other Operators in Emergency Case;
- SMARTER UC54 Local Unmanned Aerial Vehicles (UAV) Collaboration;
- SMARTER UC65 Moving Ambulance and Bio-connectivity;
- SMARTER UC66 Broadband Direct Air to Ground Communications (i.e. DA2GC);
- SMARTER UC68 Telemedicine Support;
- SMARTER UC72 5G Connectivity Using Satellite.

[14] http://www.3gpp.org/DynaReport/38913.htm
[15] http://www.3gpp.org/DynaReport/38912.htm
[16] http://www.3gpp.org/DynaReport/38900.htm
[17] http://www.3gpp.org/DynaReport/38802.htm
[18] http://www.3gpp.org/DynaReport/38804.htm
[19] http://www.3gpp.org/DynaReport/38801.htm
[20] http://www.3gpp.org/DynaReport/38803.htm

Some of these use cases are further discussed in Section 6 of D2.2 and further in contributions from D3.1 and D3.2.

Moreover, here are some examples of "extreme" requirements of the above mentioned use cases:

- The 5G system shall support very large coverage areas (e.g., to enable very long range beyond 100 km).
- The 5G system shall support service continuity of 0 ms interruption (which can be also obtained through multi-connectivity):
  - the target for mobility interruption time should be 0 ms;
  - the target for control plane latency should be 10 ms.
- The 5G system shall support very low latency (~1 ms)
  - the target for reliability should be $1\text{-}10^{-5}$ within 1ms.
- The 5G system shall support an air-interface with latency of up to 275 ms when satellite connection is involved.
- The 5G system shall support speeds between 500 km/h and up to 900 km/h at different altitudes between 4000 m and 10000 m.

The opportunity of designing a complete new technology such as the next generation of wireless can be compared to Halley's Comet: it comes around only once or twice in a person's career.