Coherent

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

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

Deliverable D2.1
Use Cases and Architecture

Version:	1.0
Due date:	31.12.2015
Delivered date:	31.1.2016
Dissemination level:	PU

The project is co-funded by

European Commission

Authors

George Agapiou (Editor-in-Chief), Alexandros Kostopoulos, Panagiotis Matzoros, Pantelis Argyrokastritis (OTE); Aarne Mämmelä, Tao Chen (VTT); Kostas Pentikousis, Fang-Chun Kuo, Umar Toseef (EICT); Rebecca Steinert, Shaoteng Liu, Ian Marsh, Per Kreuger (SICS); Mariana Goldhamer (4GC); Antonio Cipriano, Dorin Panaitopol (TCS); Dimitri Marandin (CMA); Karol Kowalik (INEA); Navid Nikaein, Chia-Yu Chang, Konstantinos Alexandris (EUR); Adrian Kliks (PUT)

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
0.1	15.11.2015	ToC created.
0.2	20.11.2015	Contributions in chapter 2.
0.3	30.11.2015	Contributions in chapter 3.
0.4	5.12.2015	Contributions in chapter 4.
0.5	10.12.2015	Contributions in chapter 5.
0.6	15.11.2015	Contributions in Annex.
0.7	17.11.2015	Contributions in chapters 1, 6.
0.8	19.11.2015	Internal review.
0.9	21.1.2016	Addressing comments.
1.0	31.1.2016	Final version.

Executive summary

In this deliverable we define major scenarios and use cases for the COHERENT project. The target of the deliverable is to identify meaningful and realistic business cases, and build interesting use cases which will be considered for the evaluation of the COHERENT prototype. In order to do so, the involved business stakeholders and their interests must be first identified and the relationships between them need to be tackled. This analysis will provide useful input for defining requirements regarding COHERENT architecture.

Firstly, we briefly investigate 5G requirements, resources and KPIs, identify relevant international fora, provide an overview of the enabling technologies for 5G, and present research efforts in the area of management techniques for various infrastructures.

Additionally, this deliverable considers the various types of users to be served and the types of services and applications that are to be delivered. In our analysis, we describe the potential use cases for COHERENT project, as well as their challenges and benefits. The identified scenarios include network cooperation, spectrum management, critical communications, network slicing, massive IoT, as well as broadband communications in public transportation. All the aforementioned scenarios are candidates for further consideration during the project.

Moreover, the deliverable identifies initial requirements for the COHERENT system, which will be under continuous re-consideration. In particular, application-driven network requirements are derived based on the defined scenarios and use cases.

Finally, using the requirements analysis as input, we present high-level design aspects regarding COHERENT architecture. The proposed design leverages a centralized control layer where the global view of the network is gathered and exposed to the applications developers through a set of high-level programming APIs implementing the proposed abstractions.

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
5G PPP	5G Infrastructure Public Private Partnership
API	Application Programming Interface
ARPU	Average Revenue Per User
AS	Access Stratum
ASA	Authorised Spectrum Access
BBU	Baseband Unit
BNetzA	Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway in Germany
BS	Base Station
CAPEX	Capital Expenditure
CBRS	Citizens Broadband Radio Service
CC	Component Carrier
CDN	Content Delivery Network
CEPT	European Conference of Postal and Telecommunications Administrations
CMOS	Complementary Metal Oxide Semiconductor
CN	Core Network
CoMP	Coordinated Multipoint
C-RAN	Cloud Radio Access Network
CriC	Critical Communications
D2D	Device-to-Device
DAS	Distributed Antenna System
DMO	Direct Mode Operation
D-RAN	Distributed Radio Access Network
eICIC	enhanced Inter-Cell Interference Coordination
eMBB	enhanced Mobile Broadband
eMBMS	enhanced Multimedia Broadcast Multicast Service
eNB	evolved Node B
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
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
ICIC	Inter-Cell Interference Coordination
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
IMT	International Mobile Telecommunications
IoT	Internet of Things
IP	Internet Protocol
ISI	Integral Satcom Initiative
ITRS	International Technology Roadmap for Semiconductors
ITU	International Telecommunication Union
ITU-R	ITU – Radiocommunication Sector
IWF	Interworking Function
KPI	Key Performance Indicator
L1	Layer 1
L2	Layer 2
LSA	Licensed Shared Access
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution Advanced
M2M	Machine-to-Machine
MAC	Media Access Control
MBMS	Multimedia Broadcast Multicast Service
MIMO	Multiple-Input Multiple-Output
MIoT	Massive Internet of Things
MME	Mobility Management Entity
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
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Networks Alliance
NVS	Network Virtualization Substrate
Ofcom	Office of Communications
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
OSI	Open Systems Interconnection
PCC	Primary Component Carrier
PHY	Physical Layer
PLMN	Public Land Mobile Network
PMR	Private (or Professional) Mobile Radio
PNF	Physical Network Function
PON	Passive Optical Network
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
RF	Radio Frequency
RN	Relay Node
RRH	Remote Radio Head
RRM	Radio Resource Management
RRU	Radio Remote Unit
SA	Service and System Aspects

SAS	Spectrum Access System
SC	Small Cells
SCC	Secondary Component Carrier
SCS	Service Capability Server
SDK	Software Development Kit
SDN	Software Defined Network
SD-RAN	Software-Defined Radio Access Network
SLA	Service Level Agreement
SME	Small and Medium Enterprises
SON	Self Organizing Network
SONAC	Service-Oriented Virtual Network Auto-Creation
TDD	Time-Division Duplex
TEDS	TETRA Enhanced Data Service
TETRA	Terrestrial Trunked Radio
TR	Technical Report
UMTS	Universal Mobile Telecommunications System
TDM	Time Division Multiplexing
TR	Technical Report
UE	User Equipment
UM	Unacknowledged Mode
V2X	Vehicle-to-X Communications
vBBU	virtual Baseband Unit
vBSC	virtual Base Station Controller
VNF	Virtual Network Function
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 Radiocommunication Conference

List of figures

List of tables

1. Introduction

The COHERENT project aims to design, develop and showcase a novel control framework for 5G heterogeneous mobile networks (HMN), which leverages the proper abstraction of physical and MAC layer in the network and a novel programmable control framework, to offer operators a powerful means to dynamically and efficiently control spectrum and radio network resources in their increasing complex HMN.

COHERENT proposes the proper abstraction of physical and MAC layer states, behaviours and functions to enable a centralized network view of the underlying radio networks with significantly reduced signalling overhead. The centralized network view with sufficient but abstracted information on spectrum, radio links, interference, network topology, load information, and physical layer reality is essential to enable optimal resource allocation in the network.

The innovative impact of the COHERENT project is the development of an additional programmable control framework, on the top of current control planes of operators' mobile networks, being aware of underlying network topology, radio environment, traffic conditions and energy consumption, and being able to efficiently coordinate wireless network resources cross the border of cells. The tools that will be developed within the project are intended to be deployable by the carriers in an incremental fashion and actually lay down the foundations for 5G and beyond mobile networks.

This document is Deliverable D2.1 "Use Cases & Architecture" for Work Package 2 "System architecture and SDK Development".

The overall objectives of WP2 are:

- To describe in detail the use cases of the project producing a set of requirements, as well as driving the definitions for the scenarios to be validated.
- To define the COHERENT reference architecture in terms of components and their interactions, and to design and implement a Software-Defined RAN controller.
- To design and develop a Software Development Kit (SDK) that will be leveraged to implement novel algorithms and protocols.

This deliverable mainly focuses on the first of the aforementioned three issues and makes initial contributions to the second one. The report defines the use cases that COHERENT solutions will address. In particular, it provides a detailed view of COHERENT's reference scenarios and analyses the requirements that need to be fulfilled by the platform in order to properly support the use cases.

The target of the deliverable is to identify meaningful and realistic use cases and scenarios which will be considered for the evaluation of the COHERENT prototype. Additionally, this deliverable considers the various types of users to be served and the types of services and applications that are to be delivered.

Moreover, the deliverable identifies initial requirements for the COHERENT architecture, which will be under continuous re-consideration. In particular, application-driven network requirements will be derived considering the defined scenarios and use cases.

Finally, using as input the requirements analysis, we investigate design aspects of COHERENT architecture. The proposed design leverages a centralized control layer where the global view of the network is gathered and exposed to the applications developers through a set of high-level programming APIs implementing the proposed abstractions.

1.1 Structure of the document

The document is organized as follows:

- **Chapter 2** presents the state of the art on 5G requirements, technology trends, and usage scenarios.
- **Chapter 3** defines a set of COHERENT scenarios and use cases approached from multiple perspectives.
- **Chapter 4** presents COHERENT requirements, and analyses them based on each defined scenario. Additionally, we attempt to map the proposed COHERENT requirements and use cases with other standards.
- **Chapter 5** investigates high-level design aspects of the COHERENT architecture.
- **Chapter 6** provides conclusions of important aspects of the document.
- An Annex with a brief presentation of use cases classification in the Third Generation Partnership Program (3GPP) TR is provided in **Annex A**.

2. State of the art on 5G requirements, technology trends, and scenarios

This chapter identifies relevant international 5G initiatives, investigates 5G business models and requirements, provides an overview of the enabling technologies for 5G, and presents research efforts in the area of management techniques for various infrastructures.

2.1 Regulation, standardization and pre-standardization organizations

5G systems will be defined by various regulation, standardization, and pre-standarization organizations. Especially important for us are the ITU-R which regulates the radio spectrum and defines the requirements for 5G systems. The 3GPP organization is a major standardization organization that will finally define one of the 5G standards to fulfil the 5G requirements. We will briefly introduce also some other related organizations since they are providing important information to the regulation and standardization work.

International Telecommunication Union Radiocommunication sector (ITU-R) is the binding treaty governing the use of spectrum and satellite orbits. It arranges World Radio Conferences (WRC's) to review and revise Radio Regulations. WRC's are organized every three or four years. The last WRC was organized in Geneva, Switzerland, in November 2015. This conference mainly concentrated on the allocation of frequencies below 6 GHz. The next WRC will be organized in 2019 where frequencies above 6 GHz will be allocated. The suggested official name for 5G systems is IMT-2020. ITU-R's publications include reports and recommendations. For 5G, ITU-R has started the Working Party WP 5D whose roadmap and future reports for the development of 5G systems are defined in [73]. The first report was published in November 2014. The published reports include [70, 71, 72]. During 2016-2017, WP 5D will define the performance requirements, evaluation criteria and methodology for the assessment of the new IMT terrestrial radio interfaces. Proposals are expected until mid-2019. In 2018-2020 the proposals will be evaluated by independent external evaluation groups and the new radio interfaces to be included in IMT-2020 will be defined in 2020.

In Europe the major organizations include CEPT, EU, ETSI, and 3GPP. CEPT is doing harmonization, European preparations for WRC's, and defining the European table of frequency allocations (http://www.cept.org). EU makes binding decisions on frequency use for member countries. ETSI is harmonizing European standards to support EU legislation for free movement of goods within the single market (http://www.etsi.org). National regulatory authorities have the sole responsibility for managing frequency use in their own country. Examples include FICORA (Finland), Ofcom (UK), BNetzA (Germany), and PTS (Sweden). 3GPP is doing the actual standardization according to the ITU-R requirements (http://www.3gpp.org). 3GPP unites seven telecommunications standard development organizations, known as "Organizational Partners" and provides their members with a stable environment to produce the Reports and Specifications that define 3GPP technologies. The project covers cellular telecommunications network technologies and thus provides complete system specifications. The specifications also consider connections for non-radio access to the core network, and for interworking with WLAN networks.

Pre-standardization or similar organizations include 5G Infrastructure Public Private Partnership (5G PPP), the Next Generation Mobile Networks (NGMN) Alliance, NetWorld2020, 4G Americas, and many others. The 5G PPP was founded in 2013 (http://5g-ppp.eu). It was initiated by the EU Commission and industry manufacturers, telecommunications operators, service providers, SMEs and research organizations. It will deliver solutions, architectures, technologies and standards for the ubiquitous next generation communication infrastructures of the coming decade. It has published 5G Vision [7].

NGMN for network operators was founded in 2005 (http://www.ngmn.org). The NGMN Alliance is developing, consolidating and communicating operator requirements to ensure that customer needs and expectations on mobile broadband are fulfilled. It has published 5G White Paper [105] and many other reports.

NetWorld2020 is the European Technology Platform (ETP) for communications networks and services (http://networld2020.eu). It was founded in 2013 by the former Net!Works and Integral Satcom Initiative (ISI) ETPs. Communications networks enable interaction between users of various types of equipment, either mobile or fixed. They are the foundation of the Internet. The NetWorld2020 ETP gathers players of the communications networks sector: industry leaders, innovative SMEs, and leading academic institutions. NetWorld2020 has published two white papers [103, 104].

4G Americas is an industry trade organization composed of leading telecommunications service providers and manufacturers (http://www.4gamericas.org). 4G Americas was founded in 2002 as a wireless industry trade association representing the 3GPP family of technologies with the name 3G Americas. The name was changed to 4G Americas in 2010. The organization's mission is to advocate for and foster the advancement and full capabilities of the LTE mobile broadband technology and its evolution beyond to 5G, throughout the ecosystem's networks, services, applications and wirelessly connected devices in the Americas. The association works within the western hemisphere. It has published 5G technology evolution recommendations [5].

2.2 5G business models and requirements

Business models will be significantly transformed in the future [7]. They will be based on shared resources and the business models will involve more and more partners delivering a part of the value. Since 5G will enable new business models in a programmable manner, application programming interfaces (APIs) should be available at different levels (resources, connectivity and service enablers) to support a variety of network and service application developers.

The 5G systems must ensure a sustainable value chain from the hardware-based infrastructure to the service-based experiences at the user level [103]. Network resources include spectrum and access infrastructure. All capacities of resource owners are shared between all operators towards providing a holistic service experience to customers. Examples include seamless vertical handover, multi-technology data load balancing, and multioperator roaming. The resource owners' business model will be based on providing the resources and underlying connections to operators. Devices will be connected to the best network at any time, either fixed or mobile.

Examples of 5G business models for operators are summarized in [105]. They include asset provider, connectivity provider, and partner service provider. If the role of the operator is an asset provider, different network infrastructure capabilities (infrastructure, platform, network) are provided as a service. Network services can be shared between two or more operators based on static or dynamic policies. As a connectivity provider best effort or enhanced Internet Protocol (IP) connectivity are provided. The IP connectivity can include a differentiated feature set, for example QoS, zero rating, latency, etc. As a partner service provider the operator capabilities can be enriched by partner capabilities or the partner offer can be enriched by the operator network or other value creation capabilities (connectivity, context, identity, etc.).

Certain key performance indicators (KPIs) called key capabilities are defined by ITU-R as requirements for 5G systems [71]. A *requirement* is a general statement about somebody's (a stakeholder, end user, etc.) needs. We will need some conceptual analysis to clarify the meaning especially of the nonfunctional requirements such as throughout, latency, energy efficiency, etc. In general, *performance* measured by the "*manner* in which *or* the *efficiency* in using the available *resources* with which something reacts or fulfils its intended purpose" [119]. Therefore it is important to know what those resources are.

The basic resources in all technical systems include materials, energy, and information [57]. Information includes data and control and it is carried by using energy. Information is an immaterial resource that is used to create some form of order and more specifically automation. During operation communication systems in general need only two of the three basic resources, namely energy and information. In digital electronics energy is sometimes used as an overall measure of the complexity of the system [95, p. 365]. In wireless communications the transfer of information is crucial and the basic resources include also time (delay, latency), frequency (bandwidth), and space (area, volume): information is delayed, it needs a certain bandwidth and for example antennas need some space. Space may also refer to the area on a chip.

Energy is perhaps the most important general basic resource. Energy is expensive for operators and users and its production has environmental effects. Most of the energy is used by the infrastructure. The consumed energy does not disappear because of the energy conservation law but is changed to radiation and heat which may be harmful to our health. Battery capacity is increasing only 50%/decade or 4%/year [136] and therefore the charging intervals cannot be significantly reduced. Mobile traffic is increasing exponentially (up to 100x/decade). Link throughput requirements are increasing up to 20 Gbit/s and for a given link distance the energy consumption is roughly proportional to the number of bits transmitted. Moore's law is slowing down will undergo a thermal noise death. Thus the energy efficiency of digital electronics will no more improve at an exponential rate. Cooling efficiency will not improve significantly, and active cooling is using energy.

System *efficiency* is usually measured with the ratio efficiency = benefits/expenditure [57, 53], for example number of bits/energy unit where bits are interpreted as benefits and energy is interpreted as expenditure. To avoid confusion, it is important that the benefits are always in the nominator and the expenditure or cost in the denominator. In this way the efficiency is improved when its numerical value is increased. A similar ratio is used even in the bit rate, defined as number of bits/time unit where the time in seconds is the expenditure in this case. An example on the opposite use is the energy efficiency metric energy per bit, which is the ratio total energy/total number of bits. It is used for example in the ratio E_b/N_0 where E_b is the energy per bit and N_0 is the noise spectral density. The problem has been noticed and the ratio number of bits/energy is now commonly used, see for example [39].

Wireless communication systems can be classified according to the most critical performance measure into the following four groups [49, 50]:

(1) *Throughput-limited systems*, for example cellular systems where the spectrum is the critical resource and the efficiency is measured using for example spectral efficiency in terms of data rate (bit rate) per frequency unit (bit/s/Hz). The benefits are now bits (information) and the expenses are time and frequency.
(2) *Energy-limited systems*, for example sensor networks where the energy is the most critical resource and the efficiency is measured using energy efficiency in terms of number of bits/energy unit (bit/J).
(3) *Delay-limited systems*, for example control systems where the most critical resource is time (delay) and it is measured in terms of seconds (s). In this case the efficiency metric is not a ratio but the expenditure itself.
(4) *Reliability-limited systems*, for example critical infrastructures where the reliability is the most critical parameter, which measures the efficiency in information transfer and is defined as the number of successfully delivered packets (benefit) divided by the total number of packets (expenditure). Thus reliability = 1 – packet error rate (unitless) and it is measured when the system is available [105].

Accordingly, ITU-R has defined three usage scenarios for 5G that include enhanced mobile broadband, corresponding to throughput-limited systems, massive machine type communications, corresponding to energy-limited systems, and ultra-reliable and low latency communications, combining reliability-limited and delay-limited systems [71]. ITU-R has also defined eight key capabilities, as shown in Figure 2-1. In most cases absolute numerical requirements are given, but for spectrum efficiency and network energy efficiency the requirements are relative to the 4G systems. The spectrum efficiency must be three times larger and the network energy efficiency must be 100 times larger than in 4G systems. Scalability will be a large problem since the 5G system may include a massive number of devices, defined as connection density of 10^6 devices/km^2 or on the average 1 device per m^2.

5G PPP has earlier defined its eight KPIs (Fig.2 in [7]). The user experienced data rate and the spectrum efficiency were not defined. They were replaced with service employment time (90 minutes, which was in 4G systems 90 days) and reliability (99.999%, which was in 4G systems 99.99 %). Some numerical requirements are also different from those of ITU-R. 5G PPP defined the energy efficiency to be only 10 times larger than in 4G systems whereas the ITU-R requirement is 100 times larger. The peak data rate is 10 Gbit/s for 5G PPP and 20 Gbit/s for ITU-R. The latency is 5 ms for 5G PPP and 1 ms for ITU-R. Otherwise the requirements are identical. More details are given in the Tables 1 and 2 presented in NGMN [105]. The requirements in those tables are divided into user experience requirements and system performance requirements. More definitions of KPIs are available in [96]. The numerical requirements are defined separately for all test cases, listed in Section 4.3 of [96].

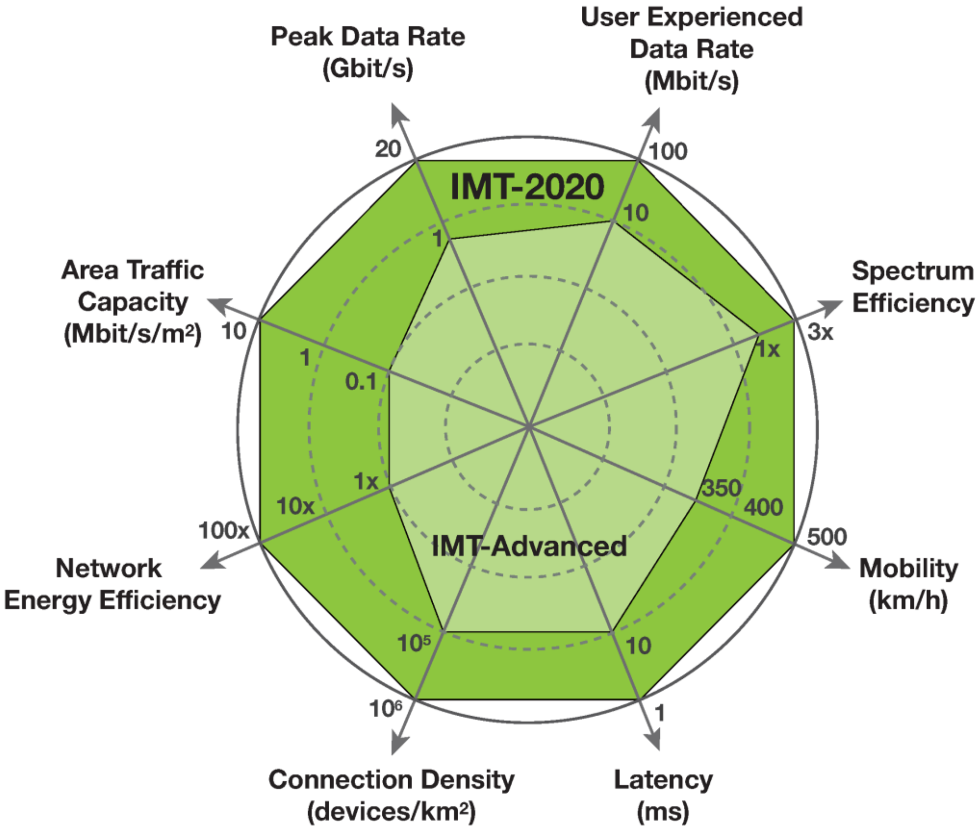

Figure 2-1 Eight key capabilities defined by ITU-R [71], including a comparison between 4G (IMT-Advanced) and 5G (IMT-2020) systems

We summarize here the definitions of key capabilities using [71] in the order they appear in Fig. 2-1. Most of them are examples of efficiency measures (efficiency = benefits/expenditure) as defined in [57, 53]. Only the latency is not presented as such a ratio. Peak data rate is the maximum achievable data rate under ideal conditions per user or device (bit/s). User experienced data rate is the achievable data rate that is available ubiquitously across the coverage area to a mobile user or device (bit/s). The spectrum efficiency is the average data throughput per unit of spectrum resource and per cell (bit/s/Hz). Mobility is the maximum speed at which a defined quality of service (QoS) and seamless transfer between radio nodes which may belong to different layers or radio access technologies can be achieved (km/h). Latency is the contribution by the radio network to the time from when the source sends a packet to when the destination receives it (in seconds). Connection density is the total number of connected or accessible devices per unit area (devices per km^2). Network energy efficiency is the quantity of information bits transmitted to or received from users, per unit of energy consumption of the radio access network (RAN) (in bit/J). Area traffic capacity is the total traffic throughput served per geographic area (in bit/s/m^2).

Reliability needs a more thorough discussion since it is not among the eight key capabilities in [71], the definition in [71] is not concrete enough, and still the definition is needed is some usage scenarios. According to some references [19, 105], there is a clear difference between reliability and availability.

Dependability is a general term that includes availability, reliability, safety, integrity, and maintainability [19]. We concentrate here on availability and reliability.

Availability is the readiness for correct service [58]. In practice it is measured as the ratio of the total time a functional unit is capable of being used during a given interval to the length of the interval [58]. It is customary to express availability in percentage, usually as 99.xxx, where xxx are numbers that complete the percentage. In NGMN [105], availability is defined as follows: “The network availability is characterized by its availability rate X, defined as follows: the network is available for the targeted communication in X% of the locations where the network is deployed and X% of the time.”

Reliability is continuity of correct service [19]. Reliability describes the ability of a system or component to function under stated conditions for a specified period of time [58]. Reliability is theoretically often defined as the complement of the probability of failure. In some cases it may be more reasonable to define it as the frequency of failures or its inverse, i.e., average time between failures [126] since this may be a more meaningful measure of user experience. Failure is an error, for example packet error. As mentioned earlier, reliability is measured only when the system is available [105]. In [105] reliability is defined to be a complement to packet error rate which follows the general definition in [58]. The reliability rate is defined as follows: the amount of sent packets successfully delivered to the destination within the time constraint required by the targeted service, divided by the total number of sent packets.” For example, 99.999 % reliability corresponds to a packet error rate of 10^{-5}. ITU-R has defined reliability as follows: “Reliability relates to the capability to provide a given service with a very high level of availability.” [71]. Thus availability and reliability are combined in this definition although in [19, 105] they are clearly separated. When these terms are used, they must be always defined to avoid confusion.

Energy efficiency also deserves a more thorough discussion. Energy efficiency metrics are summarized in [31]. Energy efficiency is the ratio of total number of correctly transmitted bits and the total energy used, including computation and communication [53, 85]. In a heterogeneous network normalization must be also done with the area. The area energy efficiency is sometimes defined as the number of bits/energy/area unit [135]. It is important to notice that dependence of the energy on the area is nonlinear, which implies that the efficiency measure is not linear and such a metric must be used with care. Between the computation and communication energies there is a classical computation-communication trade-off, which is most crucial in small cells whose radius is less than about 100 m, depending on the environment. Communication (transmission power) dominates at longer distances and computing (signal processing power) dominates at shorter distances. A requirement to energy efficiency usually comes from the power limitations of terminal devices, e.g. mobile phone and sensors. For example, if we limit the power in the physical layer of the uplink and downlink, respectively to 1 W [82] and the bit rate is 20 Gbit/s [71], the required energy efficiency is 20 Gbit/J, or after inversion 0.05 nJ/bit. As a reference value, the maximum total power for a hand-held device (without cooling) is 3 W and for a chip with appropriate cooling 200 W [68]. Since the physical layer is using altogether 2 W in hand-held terminals [82], the remaining 1 W is available for upper layers, including mainly the application processor and display [133]. On the other hand, base stations may be connected to the electricity network [18]. The total power may be hundreds of watts especially in macrocells, and because of cooling problems the functionalities must be partitioned onto several chips if the power level is over 200 W. This will affect the physical size of the implementation.

The described KPIs are requirements for various applications and it should be understood that not all of them can be fulfilled at the same time as also discussed in [71] in Fig. 4. Furthermore, some of the KPIs are peak performance requirements that can be satisfied only in certain deployment scenario.

2.3 Trends in enabling technologies

2.3.1 Overview of technology trends

Link bit rates [49, 50] and world-wide mobile data traffic [32, 41] have been increasing by a factor of one hundred in ten years. On the other hand, Moore’s law for transistors on a chip [99, 100] is approaching a thermal noise death in the beginning of the 2020s, which will cause problems in the energy efficiency of digital signal processing [51, 52, 68, 77, 81, 94, 95, 142]. The thermal noise death means that the switching energy of transistors and logic gates is approaching the thermal noise power spectral density and cannot surpass this limit called the Landauer limit, which is numerically identical to the better known Shannon limit. In addition, the Heisenberg limit regarding quantum phenomena will also become important. Significant growth is expected also for the energy consumption [42, 66]. The power consumption of mobile networks, including future wireless access points, is expected to become three times larger during 2007-2020, implying an annual growth of almost 10%. Most of the data will be video.

When the basic resources are limited and the KPIs are conflicting, some trade-offs must be made. Mathematically this is done using multi-objective optimization [91, 22]. However, there are many problems in this approach. The optimization problem is nonlinear, the optimum is not unique, there may be local optima towards which a recursive algorithm converges, and there can be stability problems in recursive algorithms because of the feedback. If the objectives or criteria are combined with some weighting, the optimality is in general lost unless linear weighted combining is used. Usually some heuristic methods are used, which will find a good enough approximation. The general problems mentioned above cannot be avoided. Therefore the theory of multi-objective optimization is still relevant.

The basic long term problems in wireless communications include high attenuation, interference between users, and high energy consumption. New technologies are always needed. On the network layer the enabling technologies include centralized cloud radio access networks (C-RAN), software-defined networks (SDN), network function virtualization (NFV), and content delivery networks (CDN) [122, 17, 38, 113]. The networks are heterogeneous consisting of cells with various radiuses: macrocells > 1 km (up to 35 km), microcells < 1 km, picocells < 100 m, and femtocells < 10 m, the trend being towards smaller cells. Some networks may exploit energy harvesting where the power level may be below 0.1 mW [118].

The C-RAN with a centralized baseband is one trend [122]. Instead of base stations it is using a distributed antenna system where the antenna sites replace the base stations. The network may be an SDN with separate data and control planes [17]. A multilayer SDN where the control plane covers the whole macrocell, but the data plane covers only smaller cells is called phantom cell, hyper-cellular network, or macro-assisted small cell [79, 87, 107]. NFV is a complementary technology to SDN and can provide the infrastructure on which SDN can run [38, 47]. We use the definition from [47]: "NFV aims to transform the way that network operators architect networks by evolving standard information technology virtualisation technology to consolidate many network equipment types onto industry standard high volume servers, switches and storage, which could be located in datacentres, network nodes and in the end user premises. It involves the implementation of network functions in software that can run on a range of industry standard server hardware, and that can be moved to, or instantiated in, various locations in the network as required, without the need for installation of new equipment." CDNs have been proposed to maximize bandwidth, improve accessibility, and maintain correctness through content replication [113]. With CDNs, content is distributed to cache servers located close to users, resulting in fast, reliable applications and Web services for the users.

Enabling technologies on the physical layer are nicely summarized in [38, 14, 26, 55, 70]. The technologies include small cells, massive multiple-input multiple-output (MIMO) systems, millimeter waves, smart devices, device-centric connectivity, and machine-to-machine (M2M) communications. Ultra-densification, mmWave, and massive multiple-input multiple-output (MIMO) are identified as the "big three" 5G technologies in [14].

Small cells have been by far the most important approach to increase system throughput because of high attenuation of radio waves [30, 36]. However, small cells are highly interference limited [140]. In order to cope with this interference a large panel of solutions are proposed today ranging from interference rejection at PHY layer to interference coordination at RRM level, like enhanced Inter-Cell

Interference Coordination (eICIC) schemes. Cooperative communication in the form of coordinated multipoint (CoMP) [69] and relaying [111] is an important trend.

Massive MIMO systems are used for increased area spectral efficiency. They are based on base stations in macro and microcells and having a large number of antenna elements. The number of elements is much larger than the number of devices in the area [93]. Antenna arrays become more practical at high frequencies, especially at millimeter waves (30-300 GHz) where the wavelength is in the order of 1-10 mm and the available bandwidths are large.

The networks and devices will be smart, which means that they have some form of automation, as included in adaptive, learning, and self-organizing systems [80]. Smart systems may have problems with robustness: we must use the right level of automation and we must guarantee that we have high enough energy efficiency since smart systems are not always energy efficient. A self-organizing network (SON) is a learning system that can change its structure [115]. SONs may be based on centralized or distributed control. An example of self-organizing networks is an ad hoc network. Self-organization is seen as an important smart technology for 5G systems.

The system must support heterogeneous networks [56], for example infrastructure-based networks such as cellular systems and WLAN systems and infrastructureless networks such as device-to-device (D2D) systems. As the coverage area of each of the heterogeneous networks is typically overlapped, traffic steering can be used to modify user distributions across the networks in order to improve network performance [101]. Furthermore, different entities may be communicating.

Smart devices can support device-centric connectivity including distributed services, D2D connectivity, local caching, and advanced interference rejection [26]. D2D communications refers to direct communication between devices allowing local exchange of user data plane traffic without going through the network infrastructure [55].

M2M systems can include critical (e.g. vehicles) and massive (e.g. sensors) deployments. M2M that is also called machine-type communications (MTC) will form the basis of the Internet of Things (IoT) with a wide range of applications, including automotive industry, public safety, emergency services, metering services, industrial automation and medical field [55]. The M2M services have a large variety of requirements, depending on the application. In some applications the data rate may be minimal, but the link reliability must be very high virtually all the time. The latency must sometimes be very low and operation must be real time. The battery life requirement may be up to 10-15 years [105].

The effective usage of unlicensed spectrum is considered as an enabler for higher date rates. While operators firstly try to improve spectral efficiency on the licensed spectrum, eventually the licenced spectrum becomes heavily utilized or congested after deployment of small cells. There is a high attraction for cellular operators to provide services by using complementary unlicensed spectrum that has the large number of unlicensed bands where small cells will be able to find interference-low channel with a high probability. The challenges raise how to offload traffic to unlicensed spectrum, how to select a suitable channel in unlicensed spectrum, how to define policies between operators using the same unlicensed band to mitigate interference and how different technologies in the unlicensed band can coexist with fair spectrum sharing.

In the next subsections, we group and provide a more detailed view regarding trends in enabling technologies.

2.3.2 Mobile broadband access and networking aspects

It is expected to see a coexistence of legacy radio access technologies (RATs, e.g. GSM/UMTS/LTE, WiFi), and new access technologies (e.g. mmWave), and also very dense multi-layer networks consisting of cells of very different sizes in the future 5G networks. The usage of ultra dense small cells in conjunction with direct D2D and V2X communication may result in great changes on the interference behavior. Therefore, new challenges are raised in interference management. Furthermore, novel approaches for mobility management are desired when highly densified networks and D2D communications are applied. METIS [109] proposed to tackle these challenges and has presented initial investigation results in [97].

2.3.2.1 Small cells densification

Small cells are similar in concept to the macrocells, but they provide shorter coverage. Known as microcells, picocells, and femtocells, they are deployed today and are promising in the future for offering improved mobile data coverage and higher capacity. Small cells present backhauling challenges that differ from those of traditional macrocells which were used in high scales in 2G and 3G networks. Several studies in small cells have been performed requiring different approaches with regards to the decomposition of the small cell functionality which is proposed to rely on IP transport network and not on expensive fiber optic links as the C-RAN remote radio heads. In addition this decomposition can provide better means for the following: i) virtual function to communicate with multiple physical remote nodes, ii) improve the mobility between the remote nodes, improve the incorporation of techniques such as statistical multiplexing, ICIC, CoMP, etc. between the nodes., iii) reduce the complexity of the management between the nodes.

2.3.2.2 Massive MIMO

Massive MIMO systems have been proposed for micro and macrocells as a complementary technology for small cells. An energy efficiency comparison is presented in [23, 88] with the assumption that a given area is covered either by a massive MIMO system or a certain number of small cells. The comparison results depend on the circuit power and on whether sleep modes are available. Small cells (femtocells and picocells) are useful for hotpots where the user density is high and massive MIMO for larger cells (microcells and macrocells), especially in urban areas. Massive MIMO is in general not efficient in macrocells in suburban and rural areas where the number of users in a cell is small. For 3D beamforming, which implies the use of massive MIMO, 3GPP has defined test cases with inter-site distances of 200 m and 500 m in urban areas [4]. The total power consumption can be greatly improved by combining massive MIMO and small cells [23, 64]. Small cells are used in hotspots and massive MIMO elsewhere in a microcell or macrocell. To deploy small cells, a high-capacity and easily accessible backhaul network is required [89]. The use of massive MIMO base stations for wireless backhaul has been also proposed, but the first results are so far included in an unpublished report, mentioned in [89]. Additional work is necessary to make such an approach more practical. Recent results are included in [53].

2.3.2.3 Moving networks

Moving networks concept is an integral part of 5G wireless cellular network and partially comprises of mobile relay and small cell concepts [61]. A moving network node or a group of such nodes can form a “moving network” that communicates with its environment, i.e. other nodes, fixed or mobile, that are inside or even outside the moving entity. Dense small cells are deployed inside the moving network (e.g. high speed train) to communicate with the users inside the moving network, while the massive MIMO unit consisting of large antenna arrays is placed outside the moving network for the sake of connection to the outside base station. It has been shown in [134] that the users in such moving networks have high data rate with considerable reduced signalling overhead.

2.3.2.4 Localized offloading through network controlled device-to-device (D2D) communications

The design premise in previous generations of mobile networks is that the infrastructure side owns the full control for the whole systems and devices obtain service by establishing a downlink and uplink connection with base stations. This kind of cell-centric architecture (infrastructure–based) of cellular systems may change in future 5G systems. Instead of controlling everything in the base stations, the architecture evolves into a *device-centric architecture* [26], where a given device (human or machine) should be able to communicate by exchange multiple information flows through possible sets of heterogeneous nodes without going through any infrastructure. It is also referred to as D2D communications. The goals are to extend the coverage of infrastructure-based networks, offload backhaul, to provide fallback connectivity or handle episodic connection disruptions, and to increase spectrum utilization and capacity.

2.3.3 Network flexibility and centralised processing

2.3.3.1 RANaaS and C-RAN

The idea of C-RAN is to provide a greater deal of control by centralized baseband processing while offloading processing between base station entities [117, 40, 8], and is aimed at achieving e.g. energy efficiency, resource efficiency and increased capacity. More specifically, some of the main advantages of C-RAN include: adaptability to non-uniform traffic and scalability; energy and cost savings; increased level of throughput and shorter delays; and, facilitated maintenance and upgrading [8]. Additionally, the C-RAN concept is an enabler for resource sharing between multiple operators and dynamic resource management under varying load and network conditions [8]. The efficiency of C-RAN depends to a large degree on the backhaul, as baseband units are separated from the radio access units and moved to the cloud for centralized processing [34]. The development towards highly flexible, configurable and programmable RANs involves virtualization and SDN control suitable for heterogeneous mobile infrastructures [117, 40, 34, 112, 102]. Typical challenges in this respect involves the trade-off between centralized and decentralized processing for resilience and efficiency, together with development of effective, data-driven management components [112] capable of acting autonomously on dynamic network and user conditions relative to expected QoS and observed QoE [117, 112, 102].

Cheng *et al.* [34] offer an example of a re-configurable backhaul prototyped as an OFDMA-based C-RAN system using radio-over the fiber technology, demonstrated by using off-the-shelf WiMAX clients where the performance benefits are shown under varying traffic conditions. The prototype implements a central processing node which manages a pool of baseband units (BBU), where each BBU can serve one or several radio access units (RAU) as well as share a user's data for flexible handover. A resource manager schedules the BBUs to send selected frames to a RAU. A reconfigurable switch allows for unicast and multicast supporting both DAS and CoMP. The output towards the RAUs is based on optical conversion and distribution, using various optical transport technologies (e.g. TDM-PON, OFDM-PON). Several cases are evaluated including coverage and interference tests, heterogeneous users and traffic profiles, how various configurations affect the overall performance, etc.

Further ideas on the development of C-RAN are provided by Rost [117] and Sabella [40], who introduce the concept of RAN-as-a-Service (RANaaS). The RANaaS is aimed at addressing the flexibility limitations that potentially exist in a typical C-RAN setting relying on a fast fronthaul [117, 40]. In RANaaS the radio functionality of the mobile infrastructure is partially split and moved into the cloud. In comparison with C-RAN, the RANaaS concept is aimed towards a significantly greater degree of flexibility, implementing a trade-off between decentralized and centralized processing depending on service requirements and network characteristics. The RANaaS concept is a step towards a generic control of ultra-dense networks consisting of low-cost small cells that are needed to ensure coverage. The type of flexibility offered by RANaaS also puts high requirements on managing the backhaul in an efficient and effective manner, such that time-critical data transactions can be made towards a central entity (via e.g. flexible routing). The architecture proposed is centered around a logically centralized SDN control for the purpose of managing the physical and virtual instances of LTE equipment interconnected via a backhaul transport network, taking the hardware capabilities and traffic demand into account.

Further, Agyapong *et al.* [112] propose a two-layer architecture consisting of a radio network and network cloud, integrating various enablers such as small cells, massive MIMO, control/user split, NFV and SDN. The proposed architecture addresses several limitations of LTE/LTE-A (e.g. capacity, data rates, delays, massive device connectivity), and is split into two logical layers - radio network with minimum L1/L2 functionalities, and network cloud layer providing higher-layer functionalities. Control plane and user plane entities implement certain control and network functions, that preferably should be located close to the base stations and the remote radio units for time-critical missions. The functions are implemented as VNFs in a software-defined infrastructure, where programmability is supported by abstracting the protocol stack, suitable for a high degree of virtualization. The authors point out that the balance between decentralization and centralization is crucial to avoid bottlenecks, single-point of failures, and to achieve reduced costs and improved performance. Moreover, the authors emphasize the need for data-driven network intelligence, i.e. intelligent algorithms that help increase network observability for the purpose of resource management, mobility management, energy efficiency, performance management, etc.

Finally, Yang *et al.* [102] present OpenRAN as an architecture for software-defined and programmable C-RANs. The authors highlight two main challenges that are addressed by OpenRAN: pluralistic standards of various heterogeneous wireless networks create management difficulties and resource in-efficiency; and, difficulty to satisfy QoS and QoE for different services that require different network characteristics. The solution offered by OpenRAN is a framework that supports virtualization for the purpose of providing a controllable, flexible and evolvable wireless network. The architecture implements three parts: a wireless spectrum resource pool which allows for virtualized RRUs (via RF virtualization technology), that can support different protocols side-by-side (e.g. UMTS and GSM); a cloud computing resource pool that hosts virtual baseband units and base station controllers in shared physical processors which constitutes a complete C-RAN; an SDN controller in the control plane of heterogeneous RANs, that abstracts and combines control functions for various vBBU and vBSCs, where each virtual instance contains an SDN agent that communicates with the controller. Altogether, these components allow for several levels of virtualization, ranging from application, cloud level, spectrum level and cooperation level virtualization.

2.3.3.2 Software-defined wireless and mobile networks

According to [37], general challenges and requirements for a 5G network encompass optimizing performance metrics, spectral efficiency, energy efficiency, delay, reliability, user fairness, QoS and implementation efficiency. Development towards software-defined cellular networks can partially contribute to achieving flexibility, scalability, and controllability in 5G networks [86]. However, the concept of software-defined mobile networks is challenging, requiring interface specifications and abstractions, as well as timely and consistent data-plane updates under highly dynamic user conditions [37] and increasing degree of cell densification [29, 86, 9]. In order to deal with high latency requirements in the control plane, hierarchical controllers may be considered as an alternative to a fully centralized solution [98].

Li *et al.* [86] highlight that one important application for SDN in cellular networks is automatic translation and processes realizing flexible policies on subscriber attributes to local packet processing rules. Scalability through local switch agent updates is given as a challenge, due to the potential load on data-plane state updates. Flexible switch functions such as deep packet inspection, header compression, and message-based protocols would be desirable for dynamic management of services. Further, they see that a software-defined infrastructure and virtualization of cellular resources can provide increased control and programmability of e.g. base station resources. The authors propose a number of extensions to the SDN concept suited for cellular networks.

A hierarchical control plane is outlined by Moradi *et al.* [98], who address the inflexibility of LTE systems due to their current topologies (core and radio) at different regions. From this design choice, problems arise regarding lack of egress points per region, increased traffic demands that put pressure on the packet gateways, and an increasing degree of signalling traffic that will not scale. The authors suggest that many of the scalability challenges can be addressed by using a hierarchical control plane; recursive label swapping, to achieve scalable end-to-end path implementation; recursive discovery protocols for scalable topology discovery; and, algorithms using abstract network state information and topologies of hierarchical controllers for the purpose of network-wide optimization. Given the aforementioned problems, a hierarchical control plane is suggested, with a controller acting on its own logical topology, and each switch operating (on a path) using a label swapping mechanism.

Further, Gudipati *et al.* [9] orient their work on managing network resources in a software-defined RAN, and address specifically the challenge of reducing the delay between a centralized controller and the individual radio elements. This is a challenge, since the radio element has a better “view” of the local state and hence can manage its resources in a more efficient manner. The densification due to small cell sizes means that control plane decisions have to be made across base stations with low latency. The proposal is to put the “elements” together in a big base station (softRAN). The physical base stations are seen as radio elements with minimal logic. The controller is implemented as a weighted graph, with each node as a radio element, the edge represents the channel strength between two nodes. Flow records and operator preferences can be encoded. The programmer specifies software for the data and control plane components and interactions using high-level graph languages. A hypervisor abstracts the RAN hardware infrastructure into a resource graph. It generates an instance of the resource graph for each network service. Programmers describe network services as functions on the abstract resource graph model, treating RAN resources (e.g., macro eNBs, small-cells) as logical entities. Possible scenarios where this is relevant encompass load balancing between base stations, and utility optimization of QoE.

Riggio *et al*. [121] take a step in the direction of defining of reusable high level programming abstractions for managing wireless networks. Said abstractions tackle state management, resource provisioning, network monitoring, and network reconfiguration. The authors introduce also a proof–of–concept implementation of a SD–RAN Controller realizing the proposed primitives and an associated Software Development Kit (SDK) allowing programmers to create and deploy new applications and services as Network Apps. The authors identify the common aspects concerning resource management in wireless access network by first acknowledging the difference between control and management and then by analyzing the requirements imposed by wireless networks. In their paper the authors draw a clear line between network control and network management. The former (control) deals with fast timescale operations executed by the elements at the edges of the network, such as scheduling in LTE network or transmission rate adaptation in WiFi networks. The latter (management) is in charge of checking whether the operating conditions for a certain policy are still met, and, if this is not the case, of reconfiguring or replacing the policy.

2.3.3.3 Virtualization in wireless and mobile networks

In general, wireless network virtualization and network sharing can significantly reduce the overall expenses of wireless network deployment and operation while enabling a high degree of flexibility [33, 138, 65]. Moreover, virtualization also enables easier migration to newer products or technologies by isolating parts of the network [33]. However, significant challenges remain to be addressed, including isolation, separation of network functions and virtualized network functions, control signalling, programmability, dynamic resource discovery and allocation, mobility management, network management and operation, and security as well as non-technical issues such as governance regulations [33, 138, 63].

For example, Costa-Perez *et al.* [138] surveys the challenges of virtualizing mobile carrier networks, including RAN sharing as well as a higher level of base station programmability and customization. The requirements of network sharing from 3GPP standard and RAN sharing enhancements are considered for mobile carrier network virtualization. A wireless resource virtualization solution which is named as the network virtualization substrate (NVS) is utilized for wireless and infrastructure resources slicing, in order to satisfy the key virtualization features such as resource isolation, customization and maximization of the base station utilization. The use case considered encompasses a virtualized base station based on NVS, which includes a slice scheduler ensuring the resource isolation and a flow scheduling framework which enables flow scheduling to different slices, tested on a picochip-based WiMAX testbed.

Liang *et al.* [33] surveys the challenges of wireless network virtualization, and outline the necessary components for wireless network virtualization: radio spectrum resource and spectrum sharing; wireless network infrastructure, i.e. the physical substrate network including sites, base stations, core network elements, transmission networks, for the purpose of infrastructure or network sharing; wireless virtual resources, created by slicing wireless network infrastructure together with the spectrum into multiple virtual slices; and, the wireless virtualization controller split into a substrate and a virtual controller for the purpose of increased manageability and programmability of virtual resources in heterogeneous networks.

Hawilo *et al.* [63] highlight the challenges of the next generation networks with focus on vEPC. In general, the demand for broadband network increases dramatically, and in effect, the network operators seek to virtualization technologies to separate network functions from the underlining proprietary hardware, in order to reduce the cost on infrastructure investments. The authors outline a general architecture composed by a) physical server; b) hypervisor (virtual machine monitor); c) the guest virtual machine; and cloud computing services and technologies such as IaaS (infrastructure as a service), Naas (Network as a service) and VNF orchestration. The architecture is exemplified in terms of virtualizing the Evolved Packet Core (EPC) together with grouping virtual EPC functions - the proposed NFV solution indicates via analysis a reduction of the network control traffic by 70%.

Zhang *et al.* [65] propose a 5G wireless network architecture, MyNet, aimed for flexible, rapid and automatic deployment of services. MyNet includes both virtualized infrastructures (virtualized base stations and cloud) and a set of logical functions related to both control plane and data planes. MyNet utilizes the programmability provided by SDN and NFV. An important key component of MyNet is service-oriented virtual network auto-creation (SONAC), which automatically selects and deploys network functions from the set of functions in MyNet to provide customized network services. Several use cases are studied, under the scenarios of user mobility management, M2M service, connectivity management and infrastructure management.

Finally, Riggio *et al*. [120] propose the concept of Programmable Network Fabric which builds upon a single platform consisting of general purpose hardware (e.g., x86) and operating systems in order to deliver three types of virtualized network resources, namely: forwarding nodes (i.e., OpenFlow–enabled switches), packet processing nodes, and radio processing nodes. The authors then propose a high-level declarative language for NFV management and orchestration. The language allows network service developers to implement custom packet processing on a specific subset of the traffic. The low-latency, high-throughtput data-plane enabled by these concepts allows supporting offloading of certain MAC features (in WiFi networks) to micro datacenters deployed at the edges of the network. Similarly, the platform is envisioned to support dynamic functional split for LTE-based small cells.

2.3.4 Mobile Edge Computing

In mobile cloud computing, the front/end mobile application offloads its functionality and computation to the back-end server, so that the resource-constraint mobile devices can utilize the computational resource of varied services. However, the back-end servers are typically hosted at the cloud datacenter, which leads to a high latencies and low bandwidth in the end-to-end communication. Some emerging mobile applications e.g. augmented reality and cloud game, are both resource-intensive and interaction-intensive. End-to-end network bandwidth and latency have great impacts on these applications when cloud resources are applied. This motivated an architecturally change on mobile cloud computing, namely mobile edge computing [44], in order to support the low latency resource intensive computations. Similar systems are Fog computing [27], Cloudlet [124], gateways [139].

Mobile edge computing extends the cloud-based Internet by introducing an intermediate layer between mobile devices and cloud. The objective is to provide smooth and low/latency service delivery from the end device to the cloud. Therefore, a new architectural element is introduced for mobile edge computing, which leads to a three-tier hierarchy: mobile device – mobile edge cloud – cloud. The servers of the mobile edge cloud are geographically distributed at the edge of the networks. Mobile edge computing bridges the mobile devices and the cloud and brings the cloud closer to the end devices. While the mobile edge computing provides localisation, enabling low latency and context awareness, the cloud provides global centralisation.

In order to satisfy the demand of large data rate and relieve the backhaul load, one promising solution is to allocate caches and computing resources at the edge of the network so that the content and the services are placed close to the end devices. Emerging technologies such as CDN virtualization are developed for enabling content caching closer to the devices. Accordingly, low latency and high QoE are achieved for delay-critical services, e.g. video editing and augmented reality. The gains of local distributed caching in small cell networks are investigated in [12, 20]. It has been shown in the perspectives of performance and algorithm in [12, 20] that it is important to realise caching in future 5G wireless networks. The caching problem has been addressed in different aspects in the literature. The proactive caching in small cell networks was studied by predicting the users behaviours in [21] and by supporting the seamless mobility in [125]. A coded caching scheme is developed in [90, 60] and its performance is further evaluated in [59, 75, 110]. [25] proposed a learning based online caching scheme.

2.3.5 Native support for machine type communications

Massive machine communication provides up- and down-scalable connectivity solutions for tens of billions of network-enabled devices, which is vital to the future mobile and wireless communication systems.

The technologies that can support some form of M2M communications are currently not capable to support the ubiquitous access of MTDs to the communication systems. Therefore, the most natural and appealing solution is to connect MTD to the cellular networks. The ubiquitous presence of cellular networks could achieve cost reduction on the network deployment and improve the coverage and mobility support for MTC. Moreover, the communication links in cellular systems are more reliable due to their regulated spectrum management. However, the traffic of most MTC applications, e.g. metering infrastructure is usually small and infrequent data generated from massive MTDs, which is different from the current human-initiated traffic. Furthermore, the MTDs are often with limited computational capability and power storage. Therefore, current cellular networks technologies are likely not able to accommodate the expected growth of MTC service.

The standards have worked on integrating MTC to the current cellular systems. For example, 3GPP has defined two kinds of communications for MTC applications: (1) communications between an MTC device and a server, and (2) communications between two MTC devices. In LTE, two entities are defined specifically for MTC, namely service capability servers (SCSs) and MTC-interworking function (MTC-IWF). The SCS entity is proposed to offer services for MTC applications hosted in the external networks. The MTC-IWF hides the internal public land mobile network (PLMN) topology and relays or translates signaling protocols to invoke specific functionalities in PLMN. MTC-IWF is also responsible for relaying trigger requests from the SCS after checking authorization and reporting the acceptance or denial of these requests [130]. Furthermore, an important enhancement for MTC is the support of direct communications between nearby devices, which is known as D2D communications [109]. Supporting D2D for MTC is already specified in LTE Release 12. Finally, applying NFV in MTC could provide virtual instances of the required hardware according to the demand [129] and reduce the equipment and operation cost of eNB by offloading functionalities to the cloud. Such an approach has been partially considered for network enhancement in future LTE releases.

2.3.5.1 Co-existence and integration with human initiated traffic

In addition to increased bit rate and energy efficiency of the terminals and of the whole system, 5G will hence be required to provide minimal latency to critical services and seamless integration of Internet of Things (IoT) nodes, and to support massive M2M communication services, all without degrading the quality of services in traditional mobile broadband services, e.g. voice, audio/video streaming, and web browsing. However, connection establishment and radio resource allocation become problematic when massive IoT devices are processed simultaneously. In LTE, the devices need to perform a random access procedure over a shared physical random access channel (PRACH) in order to set up the connection with eNB. Serious network congestion and performance degradation may happen when a large number of MTDs try to perform the random access procedure at the same time. This results in delays and inefficient utilization of radio resources for both M2M and H2H traffic.

[83] showed that great effort has been devoted to the study and test of solutions for integrating massive MTDs in the current and next cellular system architectures without causing performance degradation on conventional human-to-human (H2H) services. For example in [141], using priority based channel access, the QoS of H2H traffic could be guaranteed by assigning higher priorities to H2H traffic and lower priorities to M2M traffic. [131] proposed to assign a single bearer to a group of MTDs so that the

group bearer remains established as long as the actual data is being transferred. [129] proposed to use NFV for M2M traffic so that the M2M traffic is separated from H2H traffic.

2.3.5.2 Accommodating small bursts of data

Many MTDs frequently exchange short bursts of data with their network-side application. However, due to the fact that the amount of the signalling overhead for transmitting these small bursts of data is higher compared to the amount of the data itself, battery life, spectrum and network capacity are then wasted. Connectionless access procedures can be used to efficiently support MTC applications that only require intermittent connectivity to transmit small packets from massive number of MTDs [6]. Data aggregation can improve the efficiency of transmitting small bursts of data. Data or signalling messages could be aggregated at various locations (e.g. MTD, MTC gateways or base stations) in the networks Nevertheless, such an approach may be not applicable for delay-sensitive MTC applications due to the additional delay caused by the aggregation process [131, 143].

2.4 Scenarios and use cases

In this section we present a summary of scenarios and use cases available for 5G so far. By a *use case* we mean a specific application paradigm, while a scenario is assumed as a wide application area, where the proposed technology can be valuable. The scenarios help describing functional and nonfunctional requirements. Different authors are using different terminology for scenarios and use cases (e.g., [13] and [74]).

As discussed briefly in Section 2.1, ITU-R has defined three usage scenarios for 5G as shown in Figure 2-2 [71]. They include enhanced mobile broadband, ultra-reliable and low latency communications, and massive machine type communications. Figure 2-2 includes also some examples of use cases. The usage scenarios roughly correspond to the throughput-limited, reliability and delay limited, and energy limited systems, which makes the classification very convenient. Of course each usage scenario has its specific requirements depending on the specific application. Since the requirements may be contradictory, for different applications different requirements are applied. Thus for example it is not reasonable to require a low delay for all applications, especially if the system is geographically widely distributed as in satellite communications where the 1 ms delay requirement cannot be usually fulfilled. The following descriptions of usage scenarios are taken from [71].

- **Enhanced Mobile Broadband:** Mobile Broadband addresses the human-centric use cases for access to multi-media content, services, and data. The enhanced Mobile Broadband usage scenario will come with new application areas and requirements for improved performance and an increasingly seamless user experience. This usage scenario covers a range of cases, including wide-area coverage and hotspot, which have different requirements. For the hotspot case, i.e., for an area with high user density, very high traffic capacity is needed, while the requirement for mobility is low and user data rate is higher than that of wide area coverage. This use case is throughput-limited. For the wide area coverage case, seamless coverage and medium to high mobility are desired, with much improved user data rate compared to existing data rates. However the data rate requirement may be relaxed compared to hotspot.
- **Ultra-reliable and low latency communications:** This use case has stringent requirements for capabilities such as throughput, latency and availability. Some examples include wireless control of industrial manufacturing or production processes, remote medical surgery, distribution automation in a smart grid, transportation safety, etc.
- **Massive machine type communications:** This use case is characterized by a very large number of connected devices typically transmitting a relatively low volume of non-delay-sensitive data. Devices are required to be low cost, and have a very long battery life. This use case can be assumed to be energy-limited.

Reliability is now included among the key capabilities although it is not included in Figure 2-1 taken from the same report [71]. We emphasize that in ITU-R definitions availability and reliability are connected, which may be confusing since elsewhere reliability is measured when the system is available, see Section 2.1. Additional use cases are expected to emerge, which are currently not foreseen. Flexibility will be necessary to adapt to new use cases that come with a wide range of requirements. 5G systems should be designed in a highly modular manner so that not all features have to be implemented in all networks.

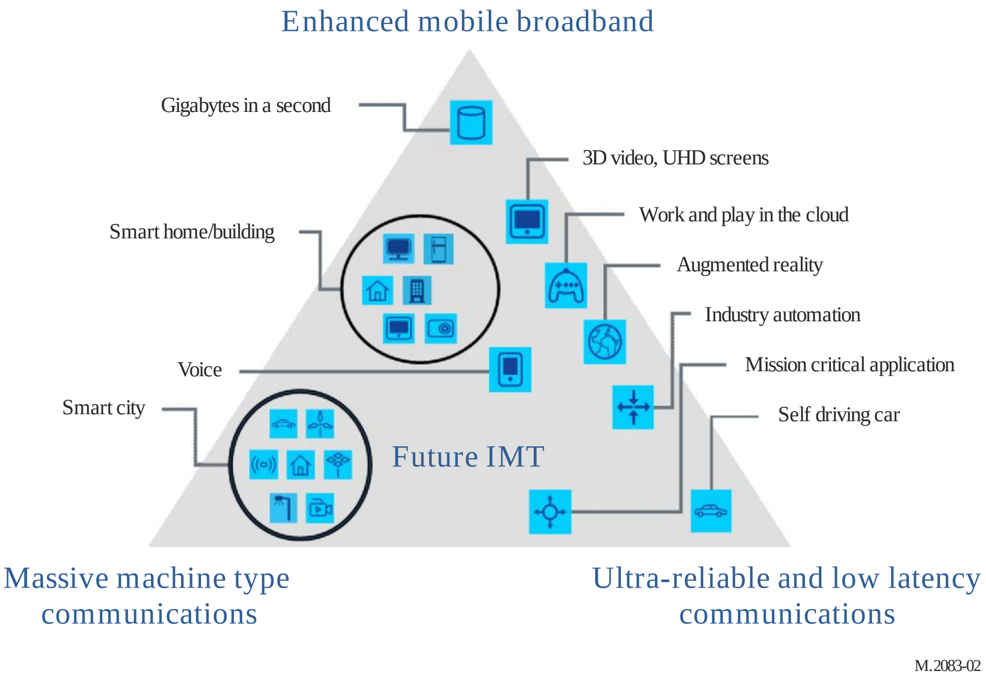

Figure 2-2 Three usage scenarios defined by ITU-R [71]

Similarly 5G PPP has defined three 5G new service capabilities that include user experience continuity, mission critical services, and Internet of Things (Fig.1 in [7]), which roughly correspond to the usage scenarios in Fig. 2-2. Each service capability is emphasizing different KPI's (Fig. 2 in [7]).

Operators have defined eight use case families [105], see Figure 2-3. Each of them can be mapped onto the ITU-R use case families. Enhanced mobile broadband includes broadband access in dense areas, broadband access everywhere, high user mobility, and broadcast-like services. Massive machine-type communications includes massive Internet of Things. Ultra-reliable and low latency communications includes extreme real-time communications, lifeline communications, and ultra-reliable communications. We may conclude that the ITU-R usage scenarios (Figure 2-2) form a good starting point for classification of different use cases and if a more detailed classification is needed, the operators' view in [105] is a good reference.

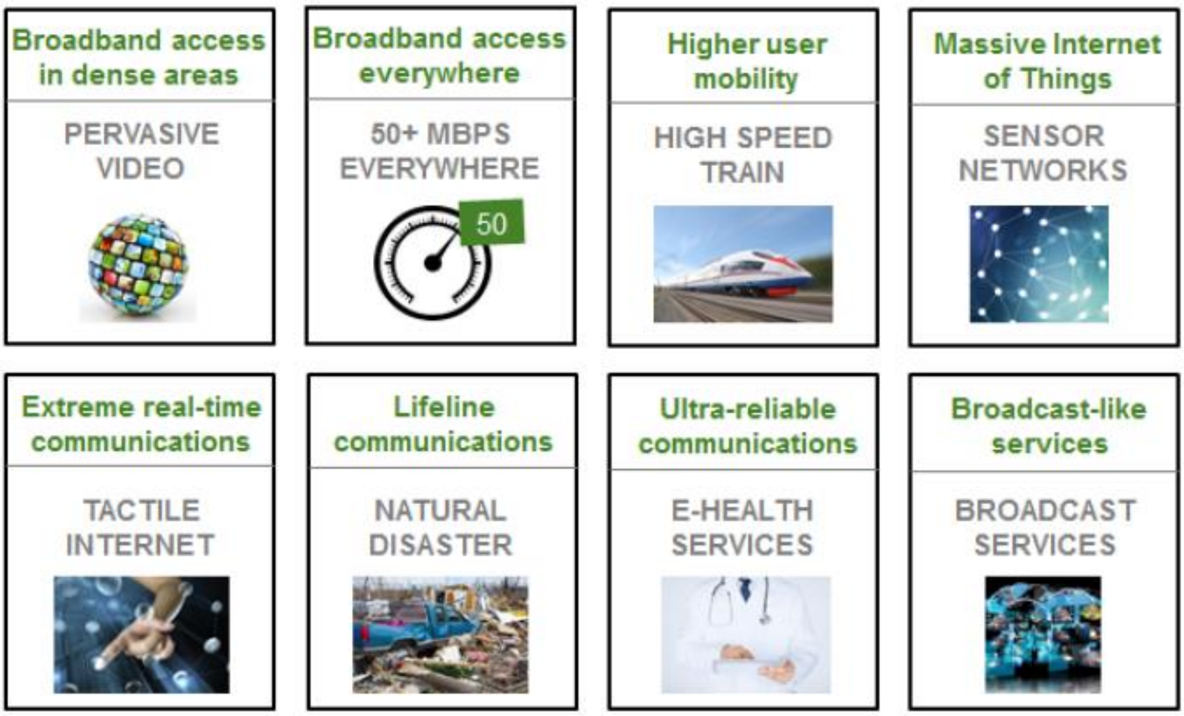

Figure 2-3 Eight use case families defined by NGMN, including examples [105]

The METIS project has defined five scenarios which include amazingly fast, great service in a crowd, ubiquitous things communicating, best experience follows you, and super real-time and reliable connections [96, 109]. In addition, METIS has defined 12 test cases, including virtual reality office, dense urban information society, shopping mall, stadium, teleprotection in smart grid network, traffic jam, blind spots, real-time remote computing for mobile terminals, open air festival, emergency communications, massive deployment of sensors and actuators, and traffic efficiency and safety. Newest results are included in the METIS-II white paper [91]. Again, the classification is rather similar to [71].

3GPP has a draft document TR 22.891 [1], which contains already more than 70 use cases which are classified into five use case categories, including enhanced Mobile Broadband (eMBB), Critical Communications (CriC), Massive Internet of Things (MIoT), Network Operation (NEO), and enhanced Vehicle-to-X Communications (eV2X). Figure 2-4 shows four of these categories. The document is still a draft and therefore the terminology has not yet been unified and the eV2X is missing from the figure. The details of the requirements are given in the appendix. Here we only present a brief summary. In many cases the report TR 22.891 refers to the earlier 5G PPP [7] and NGMN [105] requirements. There is no reference to ITU-R requirements [71], which were published quite recently.

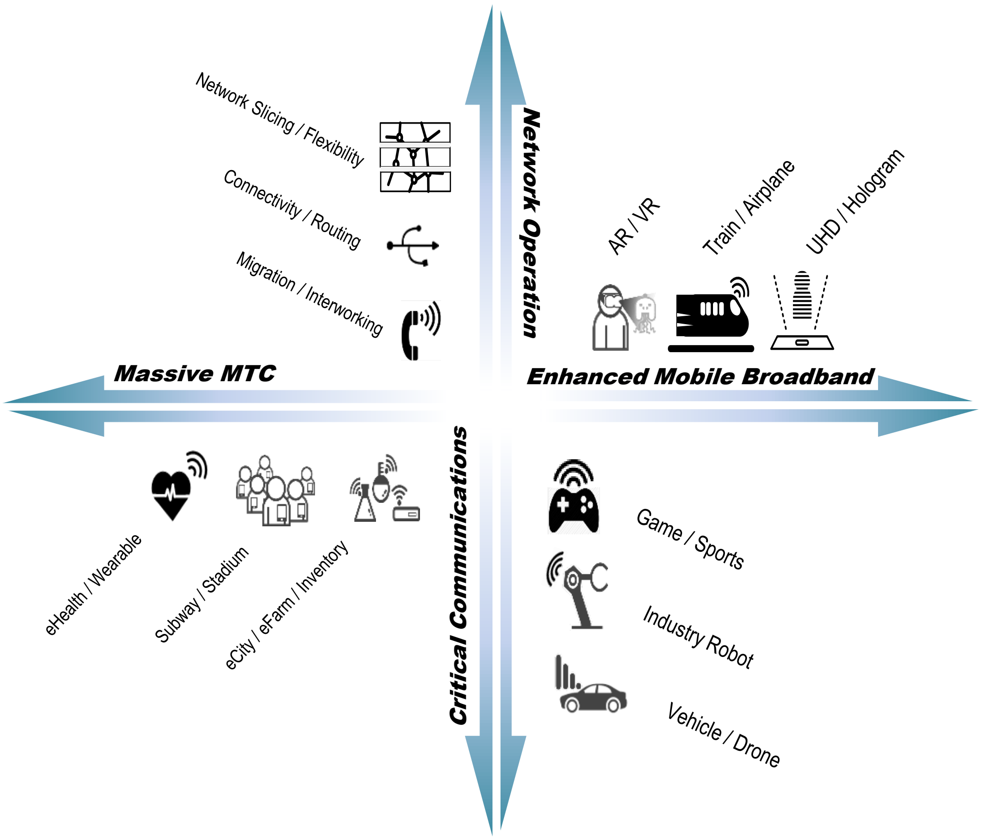

Figure 2-4 Four use case categories defined in [1]

The eMBB use case category includes three subcategories, namely higher capacity, enhanced connectivity, and higher user mobility. The eMBB category corresponds to the ITU-R usage scenario with the same name in Figure 2-2. In the higher capacity subcategory the user experienced data rate is in some cases up to Gbps level, user peak data rate is tens of Gbps, the whole traffic volume is at least at the level of Tbps/km^2, and the system must support very low latency for user experienced data exchange. Specific requirements are set to different scenarios, including mobile broadband for indoor, mobile broadband for outdoor, on-demand networking, mobile broadband services with seamless wide-area coverage, improvement of network capabilities for vehicular case, vehicular Internet and infotainment, and broadcasting support. Enhanced connectivity requires connectivity everywhere, including aerial and nautical objects. Higher user mobility shall support services in fast moving vehicles up to 500 km/h and fast moving airplanes up to 1000 km/h with enhanced user experience.

The CriC use case category includes four subcategories, namely higher reliability and lower latency; higher reliability, higher availability, and lower latency; very low latency; and higher accuracy positioning. The Cric category corresponds to the ITU-R usage scenario ultra-reliable and low latency communications (Figure 2-2). The lowest delays are in the order of 1 ms, especially in Tactile Internet which makes the cellular network an extension of our human sensory and neural system. In many other applications the delays can be relaxed and may be 10-100 ms or even 1 s. The reliability is usually defined with the packet loss rate < 10^{-5}. Reliability is in some applications measured as the fraction of transactions that cannot meet the latency or jitter constraints, and this should remain below 10^{-9}. No numerical requirement is given for availability, which should be close to 100%. For higher accuracy positioning, the positioning uncertainty should be better than 3 m at 80% of occasions. Some applications may require an uncertainty of 10 cm to avoid damage to property or life in densely populated areas.

The MIoT use case category (massive MTC in Figure 2-4) corresponds to the ITU-R usage scenario massive machine type communications (Figure 2-2). This category includes three subcategories, namely Internet of Things, smart wearables, and sensor networks. The density of connections may be up to 10^6 connections/km^2. An example of smart wearables is bio-connectivity, which is the continuous and automatic medical telemetry (e.g., temperature, blood pressure, heart-rate, blood glucose) collection via wearable sensors. The 3GPP system shall allow a battery powered sensor lifetime of multiple years while enabling a transaction rate of one every few seconds. In sensor networks the battery life time must be more than 10 years. Also range becomes a critical factor due to the low transmitter power levels of the sensors.

The NEO use case category includes seven subcategories, namely system flexibility, scalability, mobility support, efficient content delivery, self-backhauling, access, and migration and interworking. This category overlaps with the ITU-R enhanced mobile broadband usage scenario (Figure 2-2), but in some aspects also with the massive machine type communications and ultra-reliable and low latency communications. Flexibility refers to network slices, which are independent sets of network functions (potentially from different vendors) and parameter configurations. Flexibility is needed also for natural disasters, traffic routing, and resource sharing on demand and dynamically. Scalability refers to the traffic variations depending on the time of the day, on the day of the week, and on the location. Mobility support includes for example cell changes minimizing packet loss while maintaining the same IP address or if the IP address is changed, the interruption time is minimized. In high mobility scenarios the network congestion must be avoided and interference must be minimized. Efficient content delivery includes in-network caching for efficient delivery of content from an appropriate caching entity, e.g. a cache located close to the user. In self-backhauling the resources are flexibly partitioned between access and backhaul functions when supported in a common band. In the access subcategory the best connection is selected according to the traffic type. The exploitation of the existing network infrastructure must be improved. A special use case is for green radio, capable of achieving 1000 times energy efficiency compared to the legacy system. Migration and internetworking includes migration of services from earlier generations and coexistence with legacy systems.

The eV2X use case category (missing from Figure 2-4) belongs also to the ITU-R ultra-reliable and low latency communications, but it has some overlap also with the enhanced mobile broadband usage scenario (Figure 2-2). This category includes for example connected vehicles that are able to autonomous driving (self-driving cars). Therefore the system requires very low latency and very high reliability. There are various additional requirements, for example high positioning accuracy, e.g. 10 cm. The number of vehicles can exceed 10000. The absolute speed can be 200 km/h and the relative speed can be 400 km/h.

COHERENT scenarios and use cases are presented in Chapter 3. The state of the art review presented in Section 2.4 has helped us to classify the scenarios and use cases. Especially ITU-R classification will work as a starting point since ITU-R will finally define the 5G requirements. 3GPP will finally define one of the 5G standards that fulfils the requirements, which means that 3GPP’s role has to be also emphasized. There can be additional scenarios and use cases when we look at them from the software control point of view in the COHERENT project. This is the main purpose of Chapter 3.

3. COHERENT scenarios and use cases

In this chapter, we present a wide range of scenarios and use cases that COHERENT solutions will address. In particular, we provide a detailed view of COHERENT's reference scenarios. This analysis will help us then to define a set of requirements that need to be fulfilled by the platform to properly support the use cases.

The key terms related to our analysis are explained below:

- **Use case**: a specific application paradigm.
  A use case describes how the proposed technology can be used to satisfy specific needs. The use case is a single path through a diagram and focuses on a piece of functionality in a system.
- **Scenario**: *a wide application area, where the proposed technology can be valuable*.
  A scenario describes the environment in which a set of use cases can be defined. It describes the complete functionality of the system and may include multiple use cases.
- **Stakeholder**: *a party, which is involved and affected by a specific scenario or use case*.
  A stakeholder can take multiple roles.
- **Role**: *a set of specific activities by a specific scenario or use case*.
  A role could be played by different stakeholders.

The following figure depicts the scenaria described in Coherent and their underlying use cases.

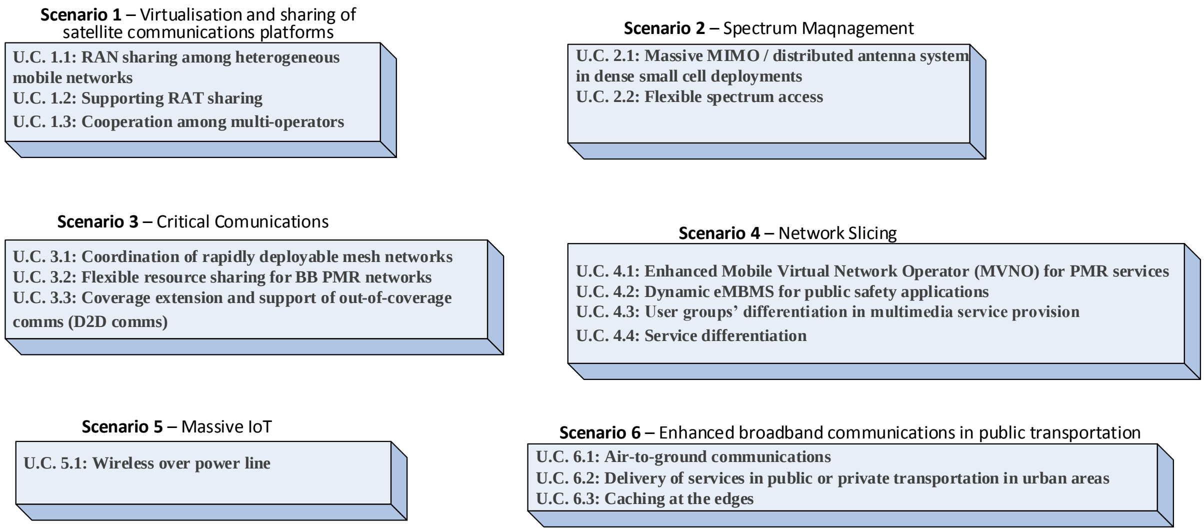

Figure 3-1 A diagram showing the complete set of scenaria with their composed use cases

3.1 Scenario 1: Network co-operation and inter-operability

Commercial aspects of operation in heterogeneous environment drive the operators need to look for solutions which allow them to share their resources. This concept consists of spectrum sharing, RAN sharing, infrastructure sharing and can be applied to various elements of network deployment and various technologies. The controller or coordinator, which can be located in the cloud is responsible to coordinate all available sharing schemes and provides control signalling for sharing in the heterogeneous mobile network.

3.1.1 Use Case 1.RS: RAN sharing among heterogeneous mobile networks

Description

The RAN sharing concept aims to allow the mobile virtual network operators to dynamically use available mobile radio resources/infrastructures/spectrums in the network which belong to other mobile network operators. This sharing scheme can reduces the capital expenditure (CAPEX) and/or the operational expenditures (OPEX) of the mobile network operators. These expenditure reductions result from the sharing of available infrastructures/radio resources/spectrum among different mobile network operators. For the infrastructure sharing case, there are further two main types of sharing: distributed RAN (D-RAN) infrastructure sharing or Cloud-RAN (C-RAN) sharing. For D-RAN infrastructure sharing case, the legacy RAN infrastructures (eNB, MME, GW) are shared among different operators, e.g., Multi-operator Core Network (MOCN) and Gateway Core Network (GWCN) architectures in 3GPP definition. As for C-RAN case, the shared eNB can be further reduced into simple Remote Radio Head (RRH) which only contains RF elements and the baseband functions, which are moved to general purpose processors (GPP) on the cloud, shared among different operators. These shared GPP resources are also called the base band unit (BBU) pool. Then the statistical multiplexing and traffic load variation of all eNBs of different operators can be fully exploited, which further reduces the CAPEX.

In Figure 3-2, different infrastructure sharing schemes are depicted. First of all, in D-RAN infrastructure sharing, the macro BS, MME, GW are shared between two operators. Moreover, if the micro BSs are further replaced with simple RRHs, the BBU pool on the cloud can be shared among different operators so the C-RAN sharing is applied. Furthermore, we can apply SDN concept to separate the control plane and data plane. In this case, operator B applies the classical implementation of RAN, while operator A allocates the data plane on micro BS/RRH and control plane signaling on the macro BS. The separation of control and data planes enables an easier control of the resources sharing. Efficient RAN sharing of both radio and infrastructure is able to guarantee performances and QoS.

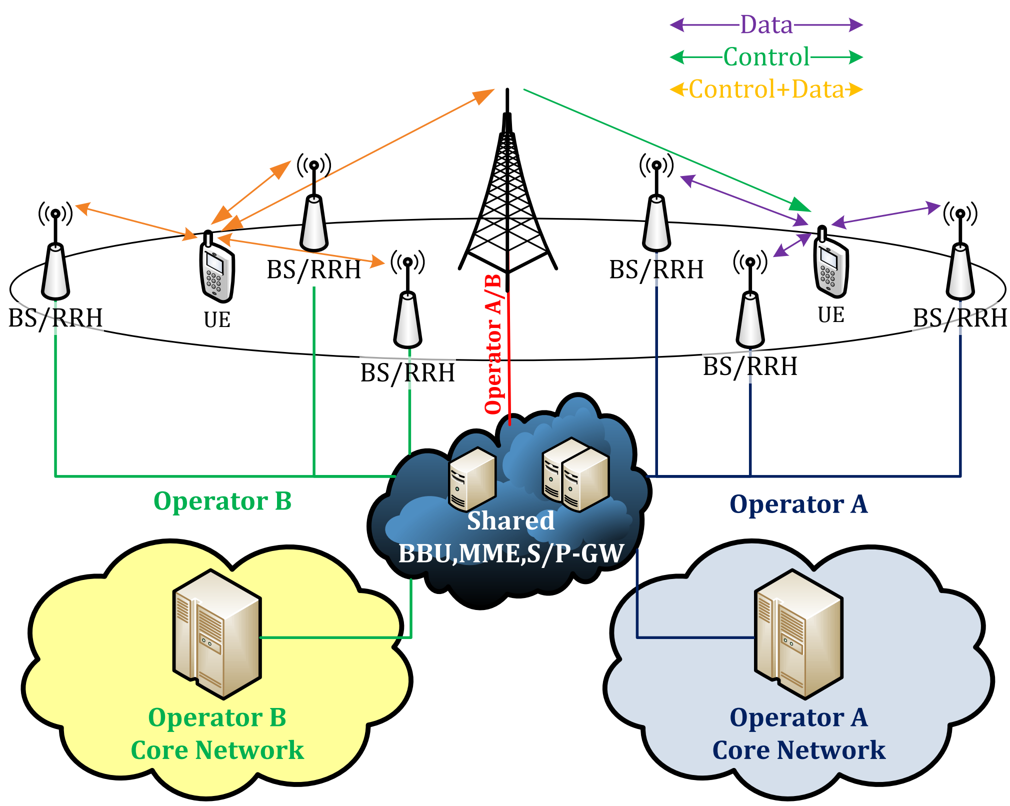

Figure 3-2 RAN sharing applies SDN

Alternatively, in spectrum sharing different component carriers (CC) of each BS can be shared via different primary component carrier (PCC) and secondary component carriers (SCCs) with different mobile network operators. If we follow the SDN logic, we could further separate the control plane which is transported via PCC and the data plane which is shared with other mobile network operators on other SCCs.

Challenges and innovation

- How to reach optimal, dynamic and parallel sharing of radio and infrastructure resources through the most appropriate abstraction?

- How to provide services over shared resources in order to be done in automated fashion and in limited amount of time?
- How to allow for resource sharing only for specific period of time?
- How to achieve the previous objective respecting delay requirements imposed by different technologies?
- How to optimally adapt/define these abstractions in HMN?
- How to reach effective RAN sharing when the UE is interacting with multiple BSs/RRHs?
- How to ensure performance isolation in the various contexts?
- How to support scaling in a fully automated way following operator needs?
- How to select the shared infrastructures for D-RAN sharing?

Benefits / key capabilities

- Better utilisation of operators' infrastructure
- Increased range and capacity
- Capacity on demand
- Better response to customers demand
- Less backhaul is required for D-RAN infrastructure sharing
- Enable dynamic C-RAN sharing for heterogeneous cell planning or legacy macro cell planning

3.1.2 Use Case 1.SR: Supporting RAT sharing

Description

Infrastructure sharing may be implemented in many forms ranging from passive sharing of base station sites and masts to sharing of core equipment. In general we consider supporting RAT sharing as complementary concept to RAN sharing since it is focused on sharing of supporting technologies which traditionaly may even not belong to MNO network. In this subsection, the use case aims to share the core and access network of different RATs, e.g., Wi-Fi. This supporting RAT scharing enables the operation similar to MVNO but over dfferent radio access technologies which are not traditionally utilised by MNO. For example, the cable operators can share the common core network entities with the legacy mobile network operators using different RAT (e.g., Cable modems equipped with WiFi transcivers). This sharing use case can reduce the CAPEX and OPEX on the core network entities for both mobile network operator and cable modem operator. Besides, it can further extend the original MNO coverage via introducing extra WiFi hotsopts.

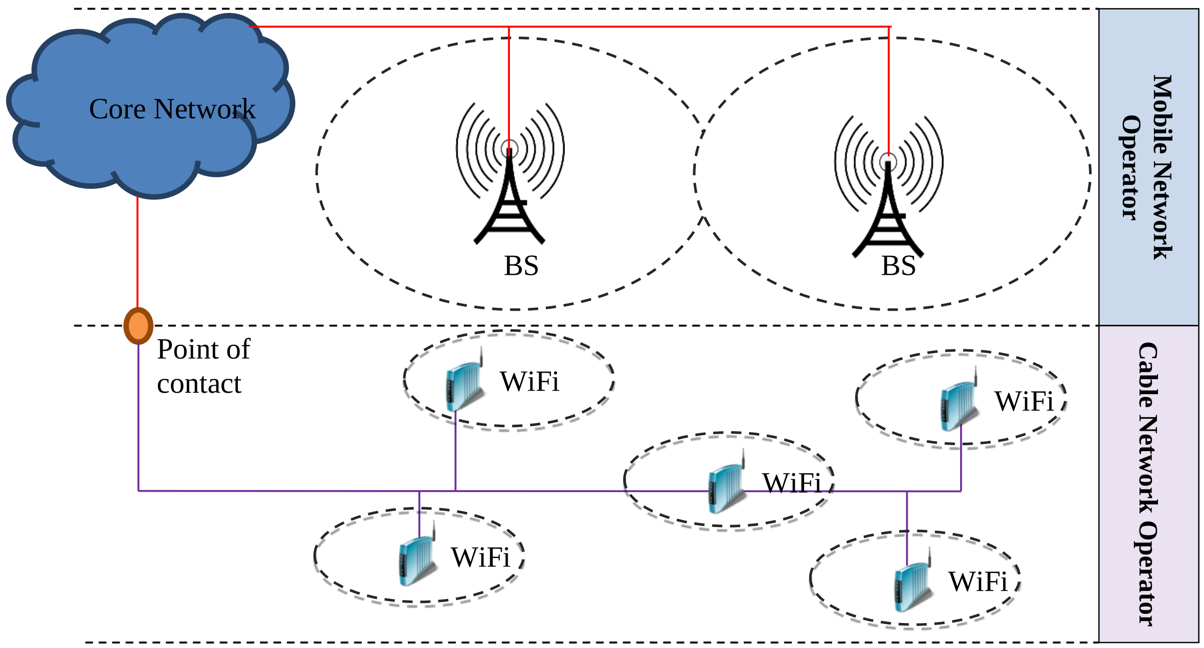

Figure 3-3 Supporting RAT sharing

Challenges and innovation

- How to provision services over shared resources so such that it is done in automated fashion and in limited amount of time?
- How to allow for resource sharing only for specific period of time?
- How to achieve QoS objectives across different technologies?
- How to ensure performance isolation in the various contexts?
- How to support scaling in a fully automated way following operator needs?

Benefits / key capabilities

- Better utilisation of operators infrastructure
- Increased range and capacity
- Capacity on demand
- Better response to customers demand
- Additional radio access opportunities/technologies

3.1.3 Use Case 1.CO: Cooperation among multi-operators

Description

Sharing of RAN or supporting RAT may be realised in the form of static agreement or may be implemented in a dynamic way. The dynamic operation is commonly provided in the form of “controller” which is a dedicated entity responsible for coordination of specific functions, for example, Licensed Shared Access (LSA) controller developed by FairSpectrum for LSA or Spectrum Access System (SAS) developed by Google for Citizens Broadband Radio Service (CBRS). In general, each operator could have a central controller which in turn could implement various services responsible for: spectrum sharing, infrastructure sharing, RAN sharing, supporting RAT sharing, etc. We think that each controller should also provide an interface which would allow it to communicate with other operators controllers. This interface should be constructed in a configurable way which could allow for spectrum/infrastructure/RAN sharing. We envision that controllers representing various operators would form a cloud of controllers. This cloud would define roles, rules and enable dynamic operation of various parties in a fast changing environment of HMN. Such a cloud would create a form of market for shared access models of operation.

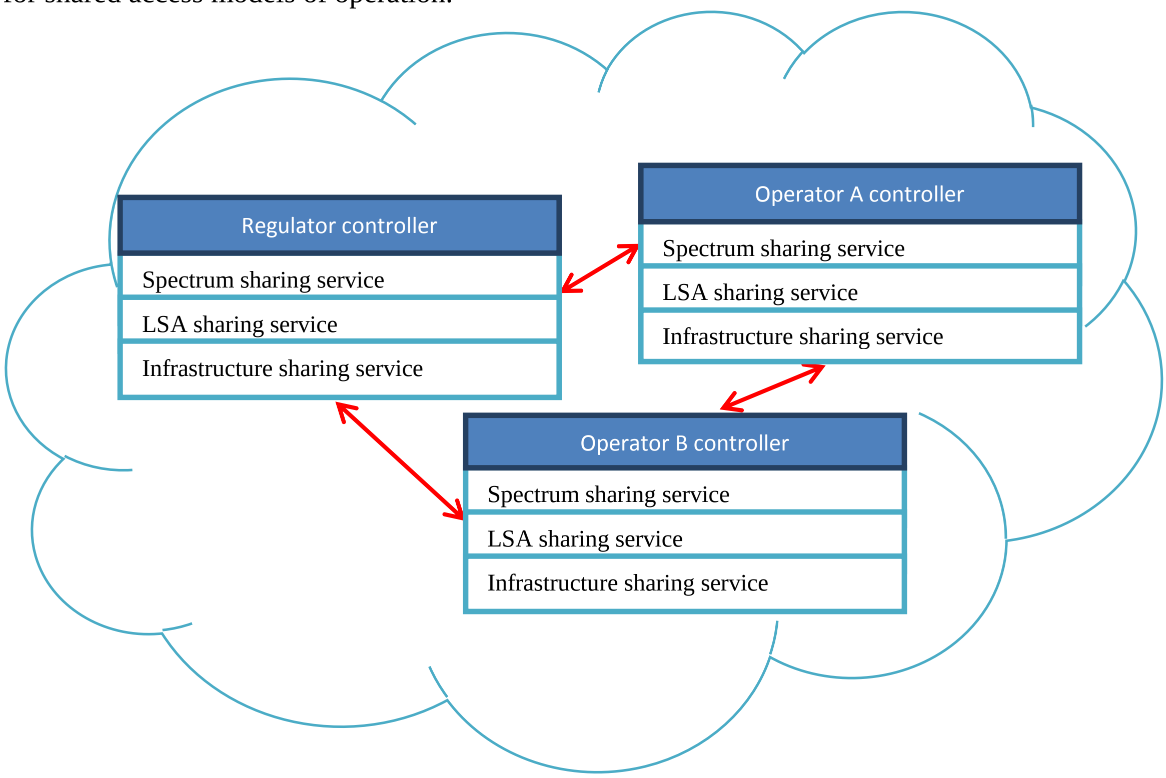

Figure 3-4 Cloud of controllers – group of controllers responsible for coordination of operation over shared HMN environment

Challenges and innovation

The key mission of each controller is to manage resources, such as spectrum, which are within its administrative domain. However, sharing of the resources requires additional interactions. These interactions and their level of complexity may vary depending on the amount of resources which are shared and the level of sharing details. Moreover, controllers may perform their interactions periodically or the interactions may be triggered by predefined events and/or event patterns (for example saturation of available bandwidth). In general, controllers should be capable of performing autonomous operation but their functions and levels of autonomicity should be well defined. Interfaces allowing for communication between controllers should be customisable while at the same time they need to provide a minimal set of predefined functions which would form a standard communication protocol between controllers.

Benefits / key capabilities

- Autonomic sharing
- On-demand resource allocation

3.2 Scenario 2: Spectrum management

Spectrum management in general is a process which aims to achieve efficient use of available frequencies. Traditionally this process has been regulated countrywide by a designated regulatory body, which allows for periodic resource allocation in a form of a license based on technical compatibility and formal aspects. However, such a periodic allocation is often provided for several years and often does not assure best solution for general public interest. In this section we focus on use cases that require unlocking additional spectrum resources by more dynamic and coordinated allocation of frequencies, distributed massive antenna systems for densely populated areas and flexible spectrum access schemes for more opportunistic allocation. These approaches aim for more efficient operation over the scarce spectrum and undermine exclusive spectrum licensing, thus allowing for more efficient collaborative approach.

3.2.1 Use Case 2.MM: Massive MIMO / distributed antenna system in dense small cell deployments

Description

The operators have started widely deployment of small cells at high-density traffic hotspots to expand network capacity and align with exponentially growing traffic demand. It has resulted in high network densification in dense urban areas (e.g. shopping malls, airports, stadiums of urban networks), where inter-cell distances could be less than 100 meters.

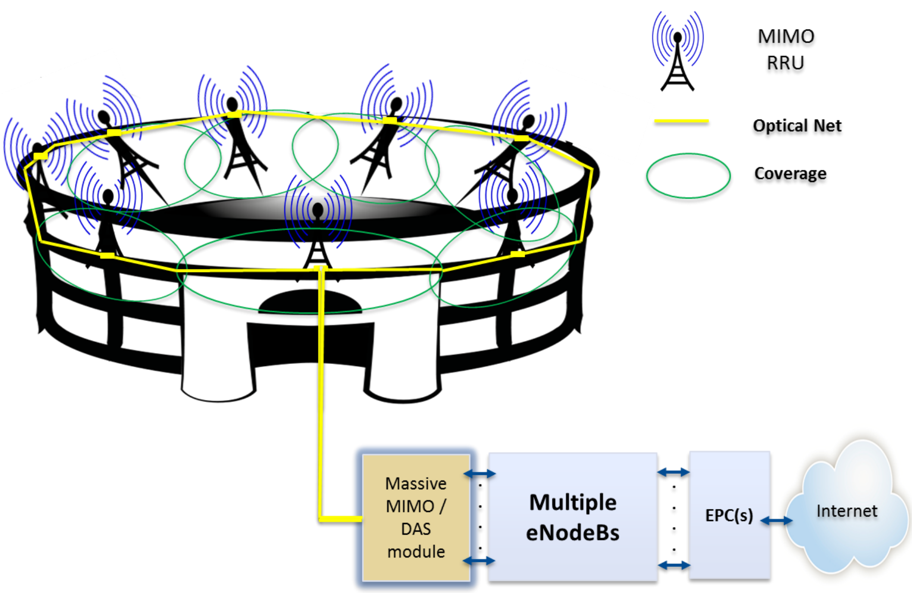

Figure 3-5 MIMO - Distributed Antenna System with optically networked remote radio units

Challenges and innovation

With decreasing cell sizes at high-density traffic hotspots, the number of cell-edge areas increases dramatically. As the users move throughout hot-spot areas covered by multiple base stations, they will be always located at cell-edge at some moments and experience performance degradation due to high interference. Mobile operators are looking for efficient densification techniques while the users are looking for consistent service in hot spot areas.

The problem is addressed in the project by using Massive or very-large MIMO eventually combined with Distributed antenna system or base station coordination (e.g. CoMP and interference alignment techniques). Massive MIMO is considered as a key technology enabling this use case. It contributes to interference elimination and thus will provide a significant improvement to cell-edge users' throughput, contributing to their consistent mobile experience. For mobile operators, it will allow a more intense reuse of the spectrum in dense areas and so increasing cell spectral efficiency and improving cell coverage. Furthermore, the combination of small cells with Massive MIMO presents new opportunities to reduce energy consumption and associated energy costs.

The theoretical investigation of massive MIMO will be based on modelling using abstractions of the physical and MAC layer (developed in WP3). The selected techniques will be adapted to a prototype implementation in WP6 demonstrating the feasibility of the approach on the commercial small base station platform with a limited number of available antennas.

There are still a lot of challenges considering massive MIMO, for example, computational complexity, distributed/coordinated processing algorithms, and synchronization of the RF units.

Benefits / key capabilities

- Network capacity expansion.
- Reuse of the spectrum in dense areas.
- Increasing cell spectral efficiency.
- Consistent service provision in hot spot areas.
- Reduction of energy consumption and cost.

3.2.2 Use Case 2.FSA: Flexible spectrum access

Description

In dense urban areas, mobile traffic volume in a region may undergo dynamically changes during the day because of rush hours or some special events happened in that region. When the traffic demand in a region is extremely high so that more spectrum is required to provide extra capacity, the mobile broadband networks shall have the flexibility to temporally provision more spectrum in the region, in order to guarantee users' quality of experience. The extra spectrum can be obtained from utilization of higher frequency bands (such as mm-waves), from intra-operator spectrum reallocation including flexible duplexing, inter-operator spectrum sharing, application of such solutions as Licensed Shared Access (LSA) from incumbent spectrum owners or Spectrum Access Systems (SAS), or other methods.

Efficient application of such scheme will be possible only when mobile network operator will possess enough spectrum opportunities to guarantee reliable backhaul and fronthaul connection, depending on the operator's network infrastructure. The ability to flexible, yet precise spectrum control and management also in the context of x-haul is important to ensure realization of the Service Level Agreements (SLA) to the operator's clients.

Authorised spectrum access (ASA) can provide a competitive advantage for an operator in today's rash business environment. The integrated device, mobile or standalone, will play the role of a hot spot for users nearby to the main user for using any unused spectrum. The control of the spectrum will be managed by the operator so that the main user will always experience the expected quality of service.

Furthermore, although much effort is and will be put on harmonization of spectrum usage around the world in the upcoming WRC15/19 conferences, it is the national regulator who defines the detailed rules of spectrum usage in given region. Moreover, special spectrum usage templates can be defined due to the realization of, e.g., specific mass event (such as visit of the VIP or organization of the mass sport contests), where many mobile users will request for high-quality up-link broadband services, services which may be provided while using the downlink FDD frequency channel.

In other cases, like industrial control, is requested a low amount of high quality licensed bi-directional spectrum which can be provided in the un-used time-frequency resources of the FDD uplink channel.

Finally, the operator can consider application of advanced traffic offloading schemes to other 3GPP and non-3GPP networks in order to realize the committed SLAs. Clearly, such offloading can be done for either capacity or coverage, allowing user-centric service delivery.

Some examples of flexible duplexing, such as TDD or DL only in the uplink FDD frequency channel were already addressed in the literature. Their applicability is conditioned by the national regulations and by the ability to manage the interference within an operator deployment and in some cases also between operators.

In consequence, the delivery of flexible spectrum access schemes can be treated as a technical enabler for 5G networks.

The presence of sophisticated spectrum management and control framework can be good candidate for 5G networks supporting flexible and effective mobile broadband access schemes. The new spectrum access schemes may be relevant to a number of vertical use cases, as identified in this document, as this use case proposes the ability to setup user-centric, service delivery supporting flexible spectrum access.

This use case has been contributed to the 3GPP TSG-SA WG1 Meeting #71 that took place in Vancouver, Canada, 19-21 October 2015.

The main involved actors in different licensing approaches are:

- Operators, who need to enforce the spectrum sharing policy and the associated billing;
- Producers, who need to develop the necessary management and control systems.

The main involved actors in flexible duplex are:

- Operators, who should procure and deploy equipment with suitable radio and protocol support
- UE producers, who should support the flexible downlink only and TDD operation in the up-link frequency channel
- Base station producers, who should support the flexible TDD operation in the up-link frequency channel.

Challenges and innovation

- For flexible spectrum usage it is necessary to change the existing European regulations for allowing the downlink transmission in the uplink FDD band.

Benefits / Key Capabilities

- The flexible spectrum usage will increase the spectral efficiency up to 40% as compared with the existing situation.

3.3 Scenario 3: Critical communications

Private Mobile Radio (PMR) systems cover a very large variety of applications for industrial companies, governmental institutions, transportation companies, security forces. For instance, security forces, like policemen, communicate usually through a proprietary PMR system. A company with a large fleet of delivery trucks may want to use a proprietary PMR system to track the vehicles or to provide communications to its employees.

As a part of PMR systems, world Public Protection and Disaster Relief (PPDR) applications have an important place for their implications on the society. PPDR services are defined as services that provides immediate and rapid assistance in situations where there is a direct risk to life or limb, individual or public health or safety, to private or public property, or the environment but not necessarily limited to these situations (Source: Commission Recommendation C(2003)2657) [35]). PPDR applications are usually dealt with by public safety actors (police, firemen, etc.).

Public safety network relies on a wired network that supports fixed wireless base-stations (BSs) providing planned coverage and bringing services to mobile entities, e.g., hand-held user equipment (UEs) or vehicle integrated devices. The following figures illustrate different topologies corresponding to possible use cases that public-safety users may encounter depending on the operational situation. These topologies differentiate based upon criteria: i) availability of the backhaul link and access to the core network, ii) BS inter-connections and iii) BS availability.

In the nominal case as shown in Figure 3-6, network can provide nominal access to public safety UEs and this case refers to the majority of operations (e.g., law enforcement, emergency services, fire intervention) occurring in covered cities and (sub)-urban environments where the network deployment has been previously designed and planned, and services are provided within a large coverage expansion.

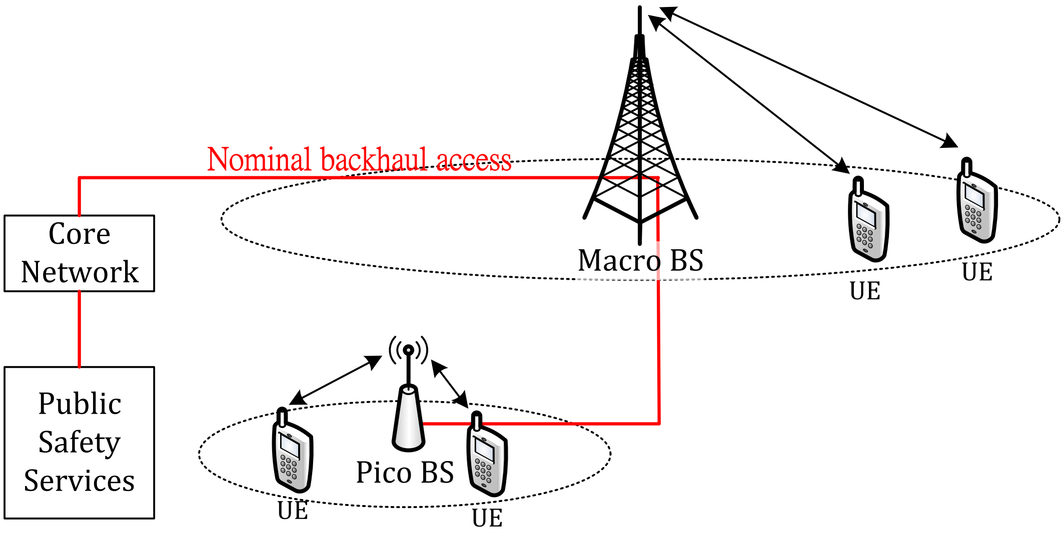

Figure 3-6 Full backhaul access scenario of public safety network

In the case of backhaul link failure due to faulty equipment (i.e., power outage or physical damages on the backhaul or RF antennas), the core network may not be fully accessible to the BSs. However, depending either on the type and the position of failure, or on the availability of backup solution, the BSs may still keep adequate inter-connectivity through wireless link access to each other, as depicted in Figure 3-7. Besides, D2D link is also supported in the case of a mobile UE that would get out of the coverage service area provided by the BSs.

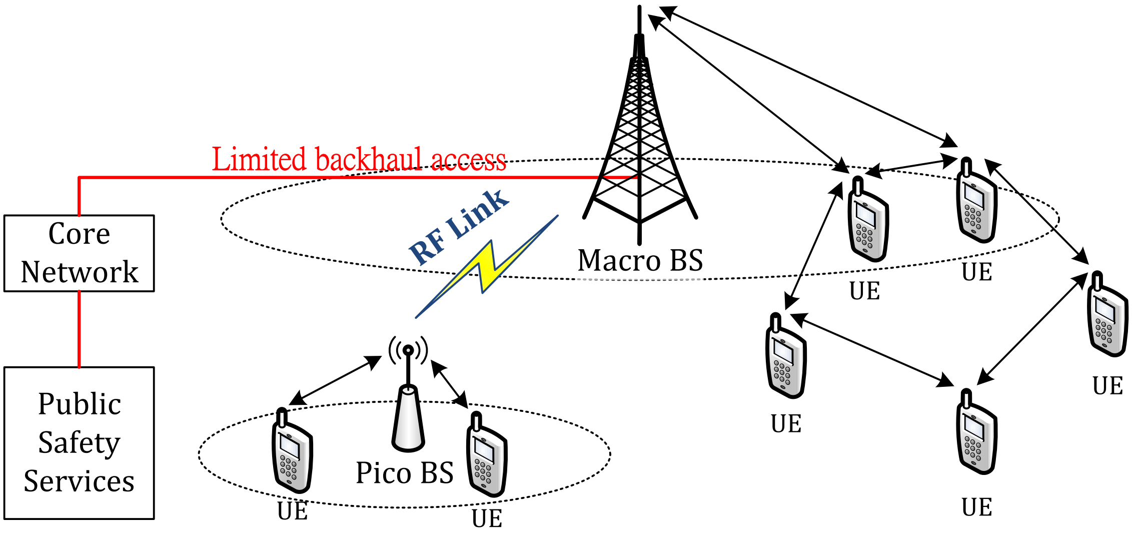

Figure 3-7 Limited backhaul access scenario of public safety network

Finally, due to UE intense mobility or no base station availability caused by the link failure or absence of coverage (e.g. underground environment, indoor environment, etc.), it is likely that only D2D communication is available between UEs as depicted in Figure 3-8. In that case, the proximity service is provided only for the D2D links among UEs. This operation mode in legacy TETRA standard is called Direct Mode Operation.

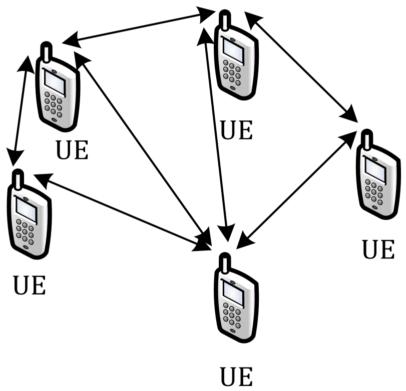

Figure 3-8 No base station coverage scenario of public safety network

PMR systems use dedicated communications systems (TETRA, TEDS, TETRAPOL, P25 or APCO 25, ACROPOLE etc.) [45, 46, 48]. These legacy systems operate on narrow bandwidth channels (e.g. in between 25 and 150 kHz for TETRA/TEDS) and were designed to satisfy voice communications and low rate data transmissions (for TETRA) with high reliability. During the past few years, public safety organizations expressed new needs in terms of improved services requiring higher data rates, for instance, video conferencing, sending detailed photos, queries on remote data bases, improved transfer of medical information, etc. These new needs fostered the work on what is called broadband PMR systems, new PMR systems able to support improved data rates while satisfying traditional requirements in terms of latency and robustness when needed. Instead of upgrading the old TETRA/TEDS ETSI standards, many industrials providing PMR systems started working on proprietary add-ons based on civilian wideband technology, namely (Worldwide Interoperability for Microwave Access) WiMAX, in order to reduce development costs. Recently the technological choice for broadband PMR system shifted towards 3GPP Long Term Evolution (LTE), as a result of the general trend to reduce network deployment cost and implementation. Today, the main PMR industrials are trying to standardize at least in part specific features and requirements of the PMR systems inside the 3GPP. As the result of this standardization work, a PMR operator will be able in the future to use a classical network operator. For instance, the efforts of the PMR community inside the Technical Specification Group Service and System Aspects (SA) brought to the creation of the SA Working Group 6 (briefly SA6) which is responsible for the definition, evolution and maintenance of technical specification(s) for application layer functional elements and interfaces supporting critical communications (e.g. Mission Critical Push To Talk). Moreover, 3GPP LTE Release (Rel) 13 contains features like D2D one-to-one and one-to-many (known as ProSe or Proximity Services at 3GPP level) and group communications provision (known as GCSE_LTE or Group Communication System Enablers for LTE at 3GPP level) which are of interest to the PMR community, since they are key elements for developing a fully-compliant PMR-over-LTE system which benefits of group communications and direct mode communications. Yet, a lot of work is still to be done in order to let LTE support for legacy PMR services.

As stated before, PMR environment is quite fragmented. PPDR services are provided to a variety of actors: police (including port/airport security, road/railway transport police), border and coastal guard, the national army (when supporting first responders after natural disasters for instance) fire services, volunteer organizations and civil protection, hospital and medical staff [76]. Even if future broadband PMR communication systems will follow a clear standardization path (which is not completely sure), the multiplicity of agencies and customers as well as the diversity of geographical coverage and of final goals of different organizations will leave unchanged the big challenge of interoperability. This issue was addressed, for instance, by the Celtic-Plus MACICO project (see [76] and references therein) which developed a TETRA network intersystem interface for addressing the need of interconnecting different countries' TETRA networks together. An important application is, for example, the coordination for the border surveillance where cross-border communications capabilities can have a big impact on the effectiveness of the missions.

A large part of PPDR services are currently identified as emerging services for 5G, see for instance the 3GPP document describing 5G SA possible use cases for the final quarter of 2015 [1]. PPDR services fall inside the group of Critical Communications, because there is no support of those cases in current 4G systems. 5G has the ambition of providing flexible systems in which a large variety of services with diverging requirements can be supported in its unique framework. This result is not obtained by a monolithic networking and radio access system performing everything, but by using an efficient management and coordination of the most adapted technical solutions to fulfill the requested performance. We notice that in France, for instance, integration of PMR services in the framework of the 5G is considered a national priority [16]. The integration of PMR systems in the 5G system should bring important advantages such as, for example, better interoperability and coordination with civilian networks and with the emerging services related to connected objects, but also reduced CAPEX. A use case update on ultra reliable comunivations was presented by a Coherent member at 3GPP TSG-SA WG1 Meeting #71bis ad-hoc, S1-153028, Vancouver, Canada, 19-21 October 2015.

3.3.1 Use case 3.MN: Coordination of rapidly deployable mesh networks

Description

In case of natural disaster, civilian networks but also PMR networks or Essential Service Operator (ESO) networks for PPDR applications may be completely destroyed or severely damaged. 5G systems should be able to provide communications even in those cases (see Sect 5.3 in 3GPP TR22.891 [1], see also NGMN 5G White Paper, section 3.2.1). Consideration on critical services, minimal energy consumption, fast system/network recovery, fast solution deployment, and location of survivals are primordial. Another important point is the transition of the wireless communications services from the moment just after the disaster, up to the moment in which normal operation of the wireless communication systems is re-established.

In the first period just after the natural disaster, providing communications services to the first aid responders is fundamental. Beside satellite and ad-hoc communications established among user terminals, another way to reestablish a network on a damaged area is to deploy wireless PMR mesh networks. Since the deployment is not done according to a well-studied deployment plan and that the network must support mobility of its nodes, the communication technology should work also on possibly unstable network topologies [114].

Rapidly deployable wireless mesh networks will operate on the same spectrum of the pre-existing PMR systems which may be completely down, or which may have kept minimal capabilities in certain areas. For example, certain base stations with independent energy supply may see that the system is out of order or that there is a malfunction of any kind and may decide (according to predefined policies) to work as isolated cells and to, therefore, adapt network topology accordingly.

The rapidly deployable wireless mesh network shall coexist or even coordinate with the rest of the system, if it is still partly operational. Moreover, the global management should allow smooth coordination during all the phases in which the previous communication system is reestablished to its full capacity.

Another use case similar to the natural disaster is when backhaul link is in failure due to a faulty equipment or power outage or physical damages on the backhaul, etc. In that case the core network may not be fully accessible to the BSs and hence they can try to establish a wireless backhaul. In this case, in order to keep on the service running, network resiliency is fundamental.

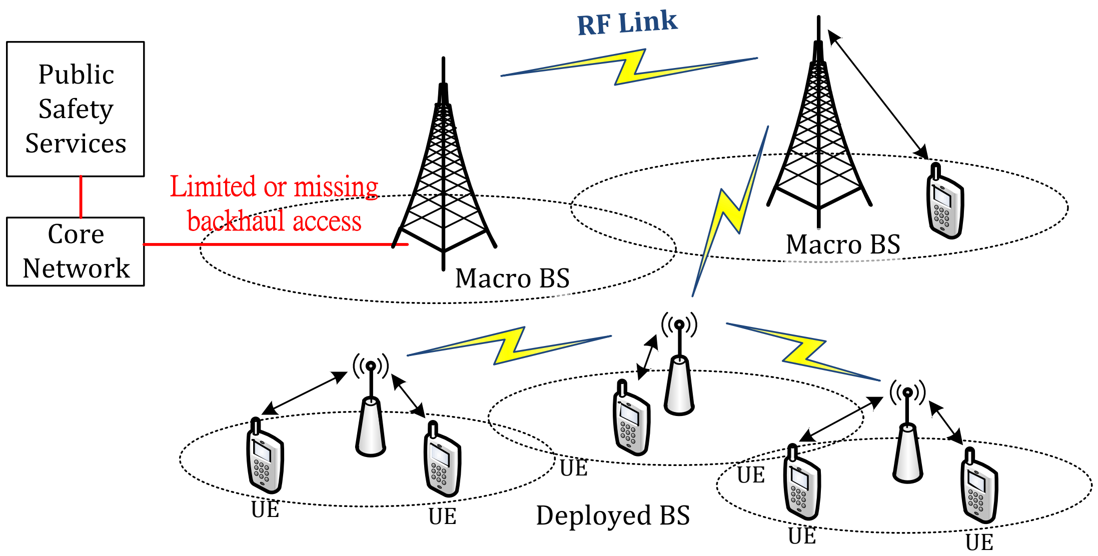

Figure 3-9 Coordination with rapidly deployable mesh networks

Challenges and innovation

The main challenges are to provide communications just after a natural disaster. We notice that up to now in 3GPP LTE, it is not possible to set up a completely independent wireless mesh network. Interoperability between the rapidly deployable mesh network and the pre-existing network is also important. Other important points are the possibility to monitor and control interference, and to manage resources dynamically in order to provide a smooth migration to the pre-existing network when it is repaired.

Notice that for the situation of disaster relief missions PMR networks can require the use of heterogeneous technologies and communications systems, e.g. satellite links to connect to remote control centres, base stations, relays, etc. Another challenge is how to build a system which allows simplified management of all these technologies and which is able to deploy them in short time with respect to the emergency time-scale which is determined, for instance, by the need of rescuing lives.

Benefits / key capabilities

- Capability to communicate without any existing infrastructure.

- Capability to provide minimal services like voice, localization, transmission of pictures with high reliability.
- Capability to provide broadband access with mastered QoS.
- Capability to manage irregular, non-optimized and possibly instable network topologies, provide routing and communication solutions to all involved actors.
- Capability to coordinate and manage different technologies (e.g. satellite communications, WiFi hot spots, etc.).
- Interoperability between the rapidly deployable mesh and the part of the pre-existing PMR network which is still operational, if it exists.
- Mobility of nodes and of relays should be supported.
- Manage interference and coordinate networks present on the area.

3.3.2 Use case 3.FS: Flexible resource sharing for broadband PMR networks

Description

National police and firemen are institutions depending on the Ministry of the Interior (MoI) and they have their own PMR networks (e.g. legacy PMR networks and/or broadband PMR networks based on LTE). These PMR networks will probably work at least on dedicated spectrum which is allocated by the MoI to the different actors all over the country, as it is the case today.

Besides public safety forces, Essential Service Operators (ESO) play a fundamental role in a country. They consist in transportation companies (railways, airports, metros, etc.), energy production plants, energy distribution companies, water production and distribution companies, etc. Therefore, providing secure communications to ESO is important as well as for classic PMR network systems.

A possible use case could be a crisis scenario in which there is a terrorist attack or other major threat in a metro station or near an energy plant. A large number of policemen, medical staff, and firemen converge to the location. Communication needs of policemen and firemen in the area increase quickly, as a result of the high number of users and of the services required, e.g., detailed real-time video transmission. In order to cover those needs, additional communication resources are required by the PMR police network because, for instance, it has no coverage in the requested area (e.g. underground or metro station), or because there is a coverage but with a limited number of frequency channels (e.g. rural area for energy plant). In those cases the 5G PMR network can redistribute resources to police forces in a minimum required time, while preserving the requested QoS and priority policies of the legacy communications.

The result is that policemen transparently use the resources of the ESO PMR network and a higher number of services can be supported with the right level of security and QoS.

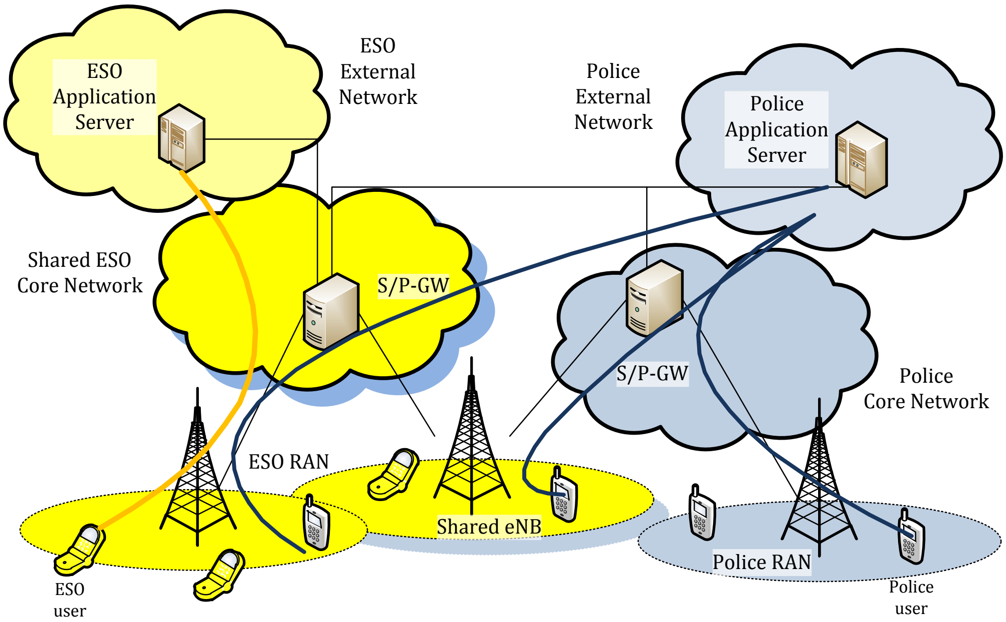

Figure 3-10 Flexible resource sharing in PMR networks

Challenges and innovation

Challenges change as a function of the relationship between the actors of the scenario. It is clear that the required service can be provided through different technical solutions.

For instance, if ESO has built its own network and operates it over frequencies granted by MoI, then ESO can support the required service for users belonging to the police by allocating them a network slice over its own resources. This strategy requires a quick coordination method between the ESO and the national police, as well as terminals able to work over the whole allocated frequency bands - which is feasible.

Another approach could be the one of multi-connectivity strategy: users belonging to the police have a multi-connectivity capability (see for instance Sect 5.64 in 3GPP TR 22.891 1[1]) so that they can simultaneously connect to the ESO network too. The ESO network will administrate the QoS of the police users according to previous agreements with the police. Since there are many PMR networks in a country, the national police should have a partnership with all of them, which may be problematic and therefore is less feasible.

Yet another approach is, for instance, when there is a unique M(V)NO which manages and administrates the networks of the national police and of the ESO. ESO and police networks will be isolated at a service layer thanks to network slicing, and reallocation of resources at network and RAN level can be done by the M(V)NO on the basis of existing agreements between the national police and the ESO. In this scenario, the MoI spectrum for PPDR services, or part of it, would be managed similarly to Licensed Shared Access (LSA).

Benefits / key capabilities

- Increased flexibility of the PMR network in order to quickly allocate resources to the network with the highest priority (for instance police over the ESO in case of crisis).
- Improved performance (in terms of data rate and of supported number of communications) of the PMR network with higher priority (e.g. police network).
- More efficient use of the precious spectrum for PMR system and PPDR services. In case of crisis, it is possible that most of ESO user will leave the place and so the spectrum of the ESO network will substantially be unused.
- Guaranteed isolation among the involved PMR networks, for security reasons but also for reasons related to management and supervision of network traffic.

- Guaranteed required QoS especially for critical services on both police and ESO networks.

3.3.3 Use case 3.CE: Coverage extension and support of out-of-coverage communications (D2D communications)

Description

Legacy PMR standards like TETRA allow communications out of coverage of the fixed cell-based infrastructure. This feature is currently not completely supported by current 4G systems, even if Proximity Services (ProSe) has been introduced in TS 36.300 in 3GPP LTE Rel-13, standardization efforts have already been done for Rel-12.

There are mainly two ways in which legacy PMR systems achieve communications out of coverage of fixed infrastructure. The first one is to support direct communications between mobile users, which is called Direct Mode Operation (DMO) in TETRA. In LTE, D2D communication, called also sidelink communications in the standard, out-of-coverage can support this mode, at least for data transmission. The second way to extend coverage is to use a UE as a relay. Both options are at least in part standardized in 3GPP but are far to be deployed.

For civilian use, out-of-coverage D2D communications or UE relay mode face some barriers, related to the difficulty for the operator to control the data flows in order to charge the spectrum use and to control QoS. Of course, privacy and security issues also exist.

For PMR use, the barriers to the adoption of out-of-coverage D2D communications are completely different. The LTE standard is introducing the good features, but the support from manufacturer is not guaranteed in short time since the PMR market is smaller than the one for consumer services. Hence it is expected that voice services will still be supported by legacy PMR systems, while data communications will progressively switch to LTE based broadband PMR systems. The management of the two systems will be a challenge.

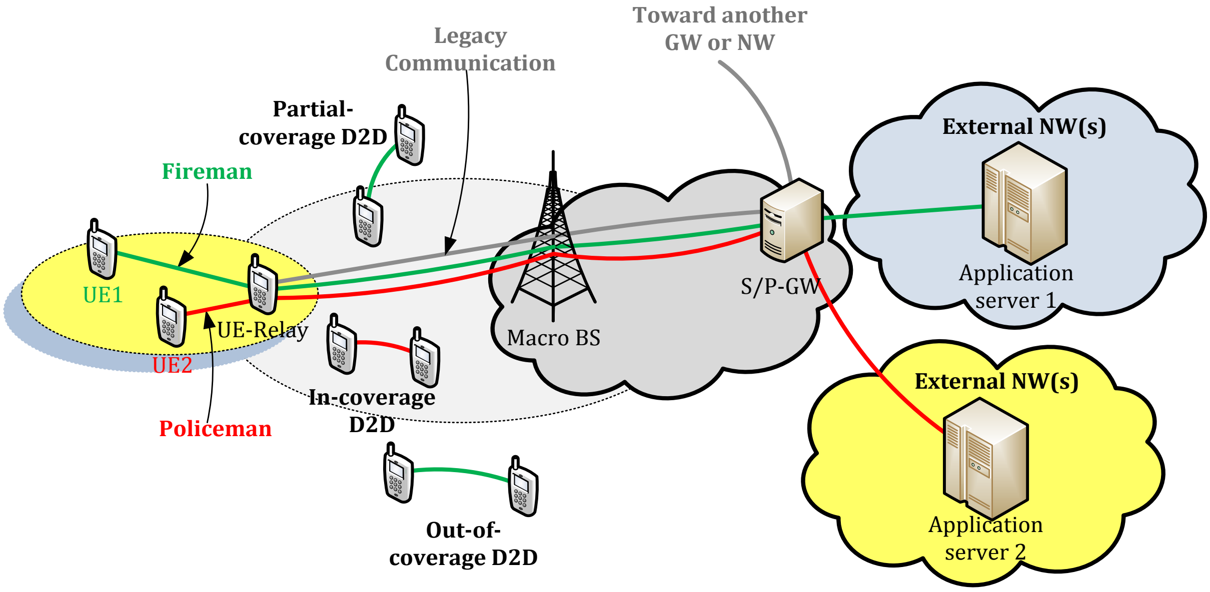

Figure 3-11 Coverage extension and support of communications without coverage

Challenges and innovation

- New architectural solutions proving transparent communication for the relayed user.
- Interference control for multiple D2D one-to-one and/or one-to-many communications among users located in the same or adjacent cells.
- Reactive solution providing a reliable communication in probably high-density scenarios and/or with high mobility.
- Fast deployment and system adaptability, providing coexistence with legacy solutions.

Benefits / key capabilities

- Relaying feature can be seen as a feature on top of D2D capability which allows relaying traffic towards/to the network from a UE and/or vice versa from the network to the UE by means of a UE-Relay. In this context the D2D have to be properly configured in order to allow relaying operation.
- D2D and relaying can help in providing potential coverage extension to eMBMS services and to general services
- D2D and relaying provide a default mode of communication when the infrastructure fails.
- D2D can help in optimizing the spectrum use.
- D2D can help in increasing the user data rate at the cell edge.
- Capability of providing at least part of the services for users out of the coverage, thanks to D2D.

3.4 Scenario 4: Network slicing

A network slice is composed of a collection of logical network functions that supports the communication service requirements of particular network functionalities. It is composed of a set of isolated network functions without sharing with other network slices. It enables parallel networks to be set up with updated software versions regarding network functionalities, which lowers the risk for the operator and ensuring minimal disruption to subscribers. It can drastically reduce the deployment time of a slice. Moreover, operators can enable more efficient business models. In this section, we present four use cases related to PMR services, public safety applications, as well as differentiation in multimedia content provision. In each use case, the provided services have different requirements (in terms of delay, resilience, performance, bandwidth, etc.), so network slicing could be applied for management of the different network functionalities needed.

3.4.1 Use case 4.MV: Enhanced Mobile Virtual Network Operator (MVNO) for PMR services

Description

A PMR network (e.g. the one of an ESO or other actors like police, fire brigade, etc.) is operated by a Mobile Virtual Network Operator (MVNO) over the network of a traditional civilian Mobile Network Operator (MNO).

The PMR network should be deployed over one or many network slices (see for instance Sect 5.2 in [1]). The PMR network shall be isolated from the rest of the MNO network not only in terms of network services but also from the security perspective (resistance to cyber-attacks included) and network performance. PMR specific services, such as Push-To-Talk (PTT) group call applications, dispatcher operation etc. shall be provided to the user or there must be the possibility to jointly manage already existing legacy PMR systems (for instance based on TETRA) and new PMR systems. In other words, MNO should be able to provide to the PMR network operated by MVNO same services as for a legacy PMR system.

The PMR network may have different needs according to the geographical area and the operational situation. For instance, in case of crisis, resources shall be allocated in order to satisfy the requested services according to their priorities. The system shall be able to dynamically manage resource allocation policies among different services of the PMR network and also with the MNO operator.

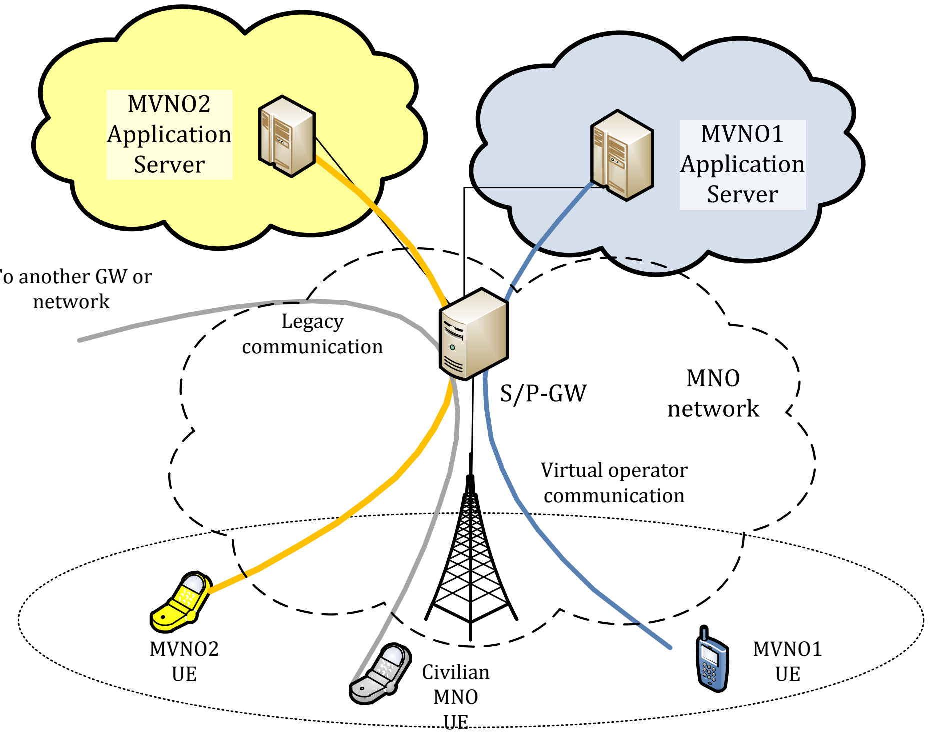

Figure 3-12 Enhanced MVNO for PMR services

Challenges and innovation

PMR services, in particular for PPDR ones, have completely different requirements with respect to the traditional services of a MNO. Even if there are efforts to introduce requirements for mission critical services in 3GPP LTE, this effort must be continued. Network slicing can be a good candidate solution for supporting a set of PMR services. However, the system must have the capability to monitor the status of the different network slices and to control the resources according to dynamic resource allocation policies. For instance, it would be good to assign dynamically more resources to the PMR network in case of crisis, but only where needed, in the crisis area.

Among network slices instantiated by the MNO, there may exist different levels of resource sharing: for instance Core Network and/or RAN can be shared. It depends on the type of technology supported by the MNO and on its sharing mode. Also Software Defined Networking (SDN) is a possible candidate technology here for supporting the flexibility in the monitoring and control of the whole network.

Notice also that, say, the national police may want to have partnerships with different M(V)NO for covering the whole country, or for providing higher data rate in dense urban areas, or for providing interoperability with connected objects (for instance, surveillance cameras, intrusion sensors, etc.). These services may be offered by different MVNOs. However, each MVNO must guarantee the level of priority and of QoS required by the PMR user.

Benefits / key capabilities

- Isolation of the PMR network service with respect to the civilian MNO services, also from the security perspective;
- End-to-End QoS management;
- Group management (such as add/remove users or form/delete new groups with new functionalities and access rights), PTT and other typical services of PMR networks;
- Dynamical resources allocation of the PMR network slice in case of crisis (dynamic policies agreed with the MNO);
- Monitor and control the global MNO network so that the coexistence of PMR network services with the other services is assured and the QoS and priority policies of the involved services and network are satisfied;

- The PMR network client must be able to supervise traffic inside its PMR network and to tune priorities, security and resource allocation polices and service QoS parameters.

3.4.2 Use case 4.eM: Dynamic eMBMS for public safety applications

Description
3GPP LTE supports multimedia broadcast/multicast services thanks to the enhanced Multimedia Broadcast/Multicast Service (eMBMS) introduced since LTE Rel-10 [84]. eMBMS was designed for the delivery of video streaming and download of multimedia content mainly. Hence it is not adapted to services with stringent latency requirements, like group calls for public safety (e.g. 300 ms for call set-up versus 4-7 s for current eMBMS bearer set-up [28]), recent works such as [28], make interesting proposals in order to optimize the resource allocation for increasing the total network efficiency. The aim is to support group calls for PMR systems, satisfying PMR requirements, with LTE eMBMS.

Challenges and innovation
An improved eMBMS controller may enhance network efficiency by controlling and configuring multicast and unicast communications for group communications in a more dynamic way. In general, it seems that further optimization can be achieved if a more flexible resource allocation at PHY/MAC layer was allowed. Another challenge is to find an architecture and protocol that are able to guarantee low group call set-up times without preconfiguring resources for the worst case. Moreover, for eMBMS services which usually operate by using a fixed MSC scheme with fixed power control and Unacknowledged Mode (UM) it is difficult to control communication of all involved users, as some users may even be in outage or unreachable. This might not be acceptable for PPDR services in which the reliability of the multicast communication is important, especially for high priority services (group call, alarms).

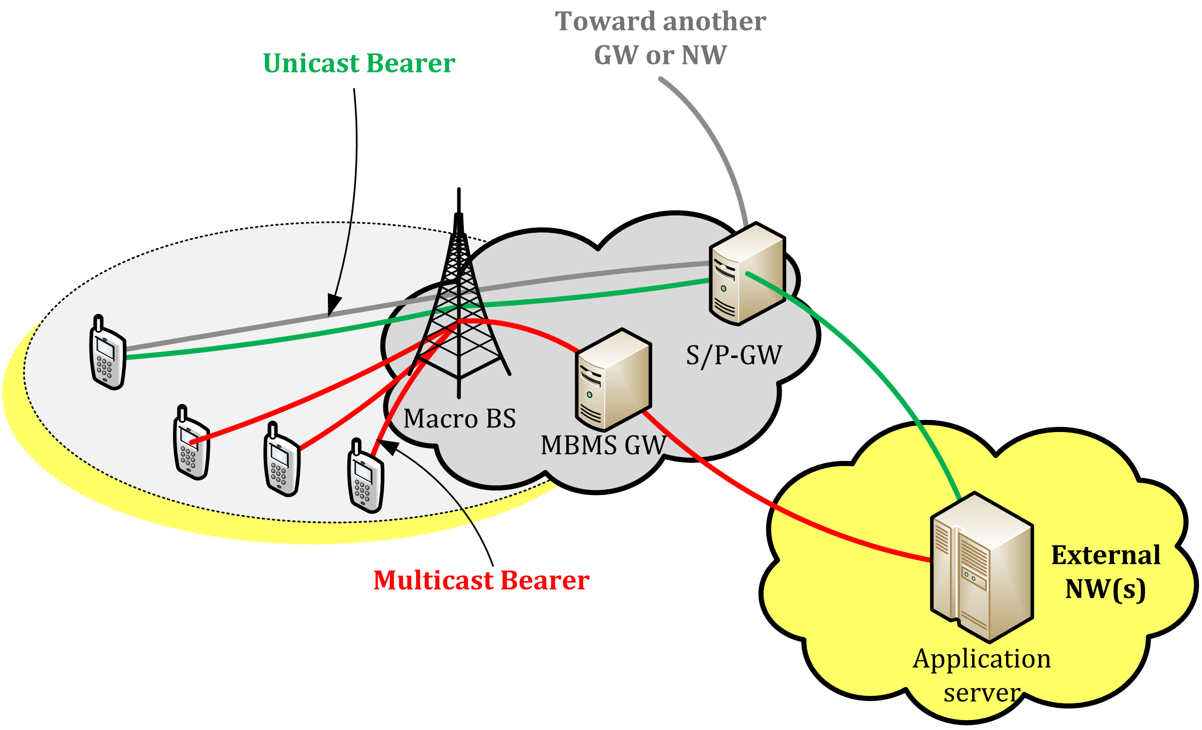

Figure 3-13 Dynamic eMBMS for public safety applications

Additional information provided by the COHERENT network view, for instance about user locations and handover combined with the link quality for multi-point links might help the eMBMS controller to improve the network spectral efficiency. Another issue may also be to provide multicast content to users located in different cells or outside eMBMS cell/unicast cell coverage. Different solutions can be envisaged.

Benefits / key capabilities

- Provide voice and multimedia multicast / broadcast services for PPDR applications (e.g. alarms to prevent a dangerous situation, natural disasters, etc.) under their specific requirements through an improved eMBMS protocol so that standardized 3GPP features can be leveraged when deploying PMR systems.
- Improved support of multiple network topologies, for instance for network coverage extensions with fixed/mobile relays or with UE used as relays.
- The system shall be able to control resources allocated to eMBMS.
- The system shall be able to control nodes in charge to support eMBMS.
- Increases network efficiency for group communications (a single eMBMS bearer is used instead of multiple unicast bearers) where a higher number of users is involved for example in PTT communications or similar activity.
- Improved downlink and uplink resource allocation for users belonging to the same group.
- Improved scheduling and more dynamic configuration/resource allocation and unicast / multicast switch based on demand.

3.4.3 Use case 4.GD: User groups' differentiation in multimedia service provision

Description

We consider a standard media service provisioning scenario, where an IPTV service is offered to a set of users with similar access profiles. In this use case, we assume that a network operator offers two types of services:

- a bouquet of TV channels in standard quality provided by a third-party, i.e., a content provider, to his end-customers, and
- a video on demand service for premium customers.

Thus, the network operator performs network management by means of assigning different network slices, dedicated for each of the aforementioned provided services correspondingly, in order to manage and provide the IPTV service with adequately high QoS/QoE. The next figure (Figure 3-14) presents an abstract view of the use case scenario:

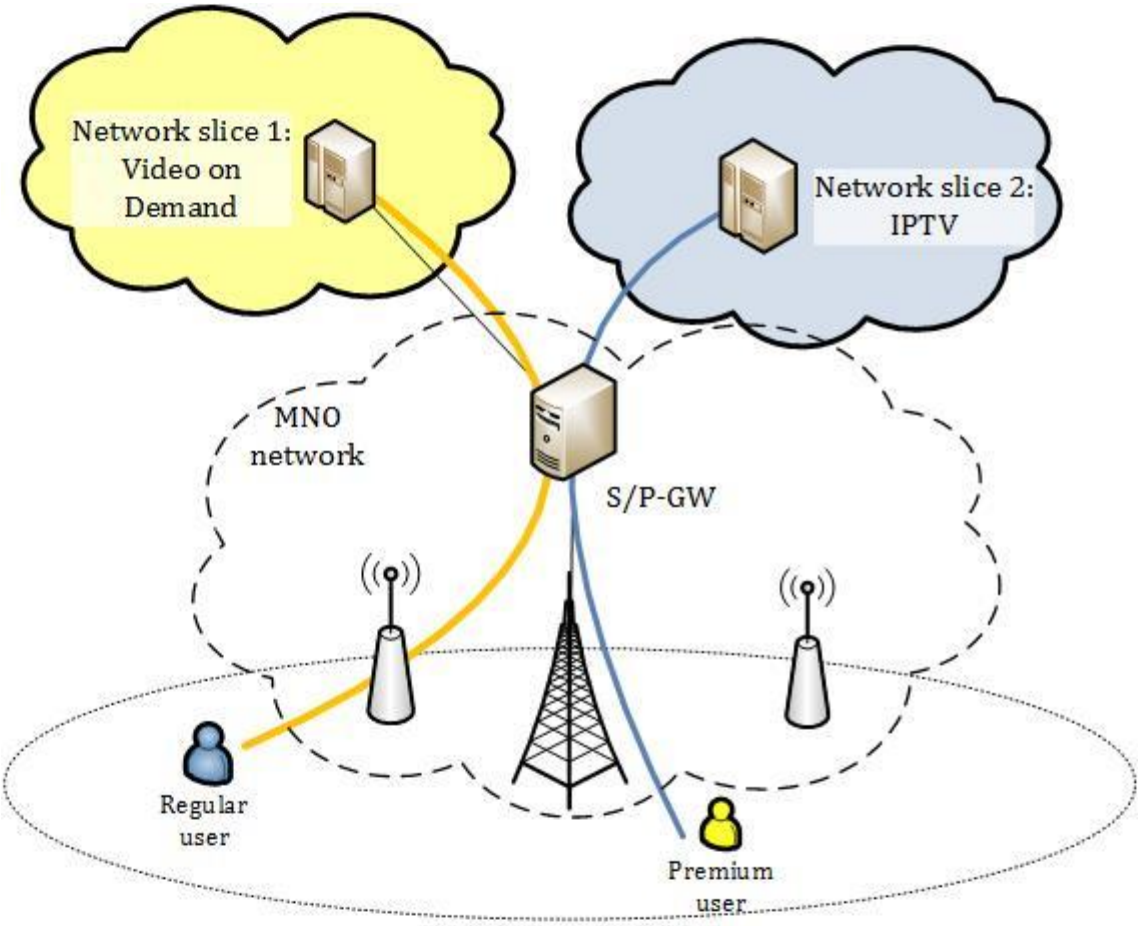

Figure 3-14 Differentiation of user groups

Benefits / key capabilities

- High data rates.
- Achieve desired QoE for each service type / user segment.
- Communication reliability.
- Use of licensed spectrum.
- Interference reduction.
- Cost reduction.

Challenges and innovation

- Satisfy different user and service needs.
- Support different network requirements
- Achieve efficient spectrum sharing and management
- Ensure interoperability among heterogeneous networks
- Propose the protocols and the structure of spectrum manager taking into account the information delivered from the abstracted and virtualized low layers.
- Support intra- and inter-operator spectrum sharing and flexible duplexing.

3.4.4 Use case 4.SD: Service differentiation

Description

Network applications have different performance requirements. Therefore, it is beneficial to separate them using logical "network slices". In general, operators RAN may operate on a set of various frequency bands. Each frequency band may feature different parameters such as capacity, latency, or jitter associated with its configuration, implementation and utilization. Moreover, each of the customer applications may have different QoS requirements. Therefore, we need application aware bandwidth/client steering capabilities, which allow bandwidth hungry users to be directed towards spectrum band which provides sufficient capacity, while latency sensitive applications should be steered towards low latency bands.

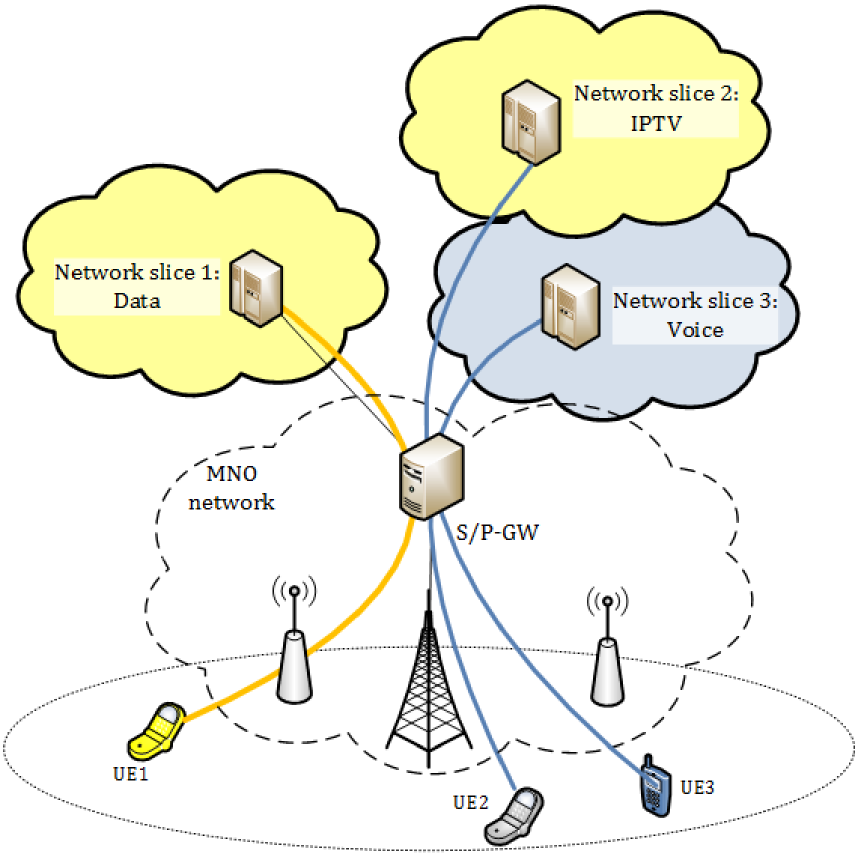

Figure 3-15 Network slicing for service differentiation

Challenges and innovation

- Satisfy different user and service needs.
- Support different network requirements
- Achieve efficient resource sharing and management
- Single client may run various applications, thus simple single selection based on application may not be feasible

Benefits / key capabilities

- Better spectrum utilisation
- Improved match between applications needs and the network.

3.5 Scenario 5: Massive IoT

The support for massive IoT is one of fundamental requirements for the next generation system. mIoT includes sensor networks, smart wearables but also the smart city authomation may be considered part of this category.

Based on 3GPP TR 22.891, and also on the Narrow Band LTE WI, the problems to resolve are related to the penetration loss (at least 20dB more link budget is required) and especially for wearables and battery powered sensors, the reduction of the energy consumption.

3.5.1 Use case 5.WoPL: Wireless over power line

Description

The combined use by an operator of 3GPP-defined RAN lower layers over the power line distribution network within buildings and over the wireless medium will increase drastically the indoor penetration for communicating with utility meters, for example for reporting electricity, gas or water consumption. An operator can use the spectrum allocated for power line communications or its licensed spectrum, preferably in under 1GHz bands.

Figure 3-16 provides an example of using power lines for transmitting a cellular RAT. The power, water, etc. meters are located in metal enclosures. In multi-dwell buildings, they are placed also within concrete walls. The best way to access them is by using power lines, as shown in this figure. A frequency adapter, which may include also a power amplifier, can be placed in a convenient location to have good propagation to the cellular or small base station or both. Given the target of wireless transmission, the RAT used over power distribution lines should be the same with the RAT used for cellular communication.

The home appliances, such as refrigerators, washing machines, TV, etc. can be connected over the power line directly to the small base station or can be connected to it over the air. However in case that there is no small base station, the same frequency adaptor can allow the direct connection to the cellular base station.

Another important aspect is the energy-efficient connection of wearables. With a frequency adapter placed in every room, the distance between the wearable device and the frequency adaptor will be of only several meters, strongly reducing the RF energy needed for transmission. In turn, this will conduct to significantly lower battery consumption.

The “Wireless over power line” enables the achievement of the following 5G-PPP KPIs:

- 10 to 100 times more connected devices.
- 10 times lower energy consumption.

It also supports the EC objectives as provided in the Liaison Letter to 3GPPP RAN meeting in Dec. 2015, in RP-151668 “LS on 5G use cases from verticals”:

“On the plus side the business potential of introducing 5G in the energy domain is exceptionally high, as it is expected to provide the necessary support not only to the critical machine type communication (MTC) applications of energy grid protection and control, *but also to the massive volume of MTC type applications of the emerging smart metering*. In summary, the anticipated performance and flexibility of 5G will enable a communication infrastructure which is able to support the emerging energy use-cases of 2020 and beyond. The ongoing evolution of the power grid into a grid supporting a much more distributed generation and storage of power as well as micro-grids would be a clear beneficiary of the high performance, but still very flexible communication architecture provided by 5G.”

In particular the following aspects of life can benefit from mobile and fixed-line easy to use services:

- *Tele-care and telemedicine*: open up new opportunities for providing medical care to the home, including ways of monitoring well-being and improving the medical information available to healthcare providers.
- *Personal health systems*: these include wearable and portable systems for monitoring, diagnosis and therapy, and supporting treatment plans for individuals with chronic disease. They are complemented by tele-monitoring and tableware. ”

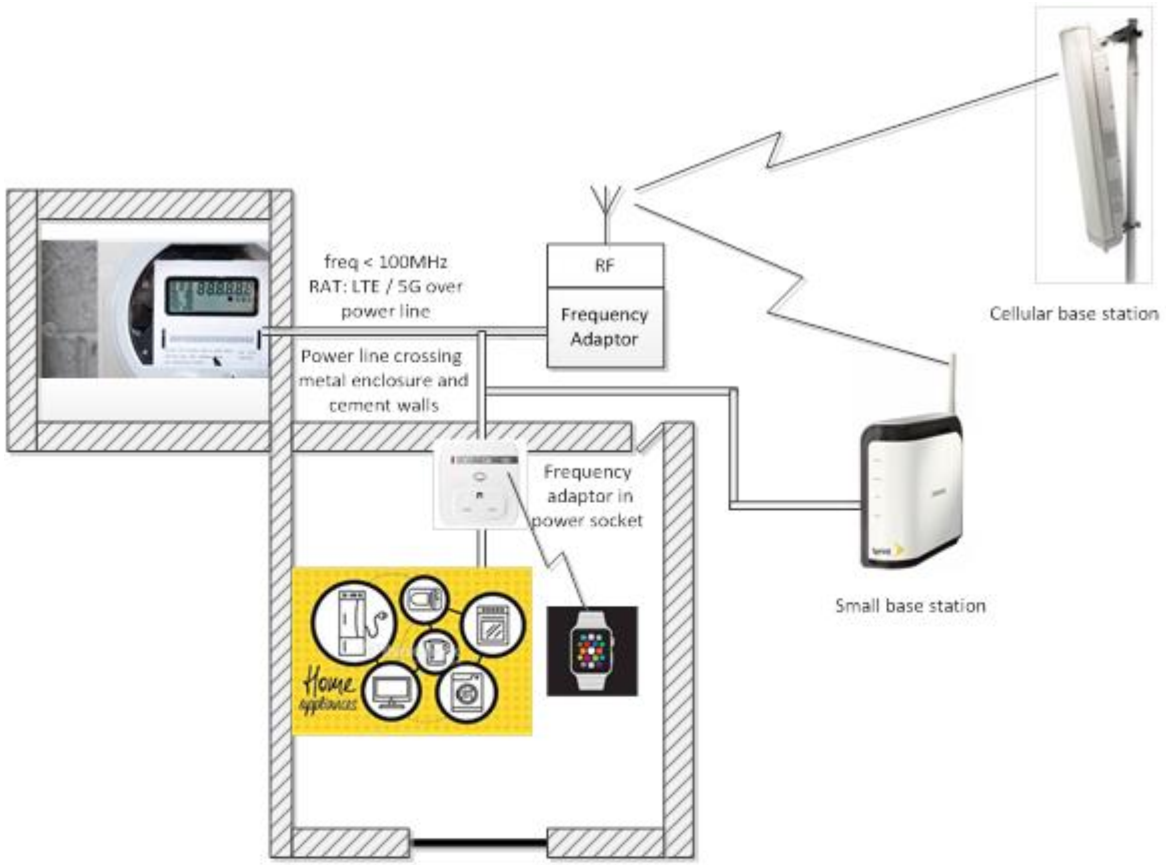

Figure 3-16 Base station connection through LTE/5G over power lines

The main involved actors are:

- Utility equipment producers, who should embed into their meters LTE/5G technology with a connection to the power lines;
- Small base station producers, who should embed in their base station a connection to the power line and an appropriate radio design
- Radio device producers, which should create chips containing radio frequency adapters, amplifiers and interfaces suitable for connection to power lines
- Operators, who should deploy this solution.

This use case was presented by a Member of Coherent at 3GPP TSG RAN ad hoc, RPa160027, RPA160001, Barcelona, Spain, January 28 - 29, 2016

Challenges and innovation

- Selection of frequency carrier to be used over power lines
- Coordination of interference over power lines and over the air, as the power line may behave as an antenna.

The innovative side of this use case is related to the using the proper communication medium for ubiquitous penetration in conjunction with the wireless technology.

Benefits / key capabilities

- Much lower deployment costs for the same level of coverage;
- Lower consumed energy for both infrastructure and mobile equipment

3.6 Scenario 6: Enhanced broadband communications in public transportation

This scenario focus on use cases with broadband communications in situations with high mobility of group of users, for instance customers accessing services in public transportation.

Providing multimedia content to users inside public transportation systems like busses and trains is a common service today. Technical solutions already exist for providing services to users in high speed trains or busses, some technical solutions using available technology are presented for instance in [3]. In the 3GPP context, the idea of a moving relay node consists in a mobile pico-cell with a wireless backbone. In order to cope with the Vehicle Penetration Loss (VPL) due to the materials of the vehicle, a moving relay node may have antennas for the backhaul on top of the vehicle and antennas for the wireless access of users inside the vehicle. Moving relay nodes are still not standardized inside 3GPP, even if there was a study item ended in Rel-12. Benefits and drawbacks of moving relays were also widely studied at least in the following research projects: ARTIST4G [127] and more recently METIS [128]. The past works have evaluated the pros en cons of considering different RATs for the backhaul and access link, architectural and RAN options, and have pointed out open problems linked to use cases like public buses in urban environment.

3.6.1 Use case 6.AG: Air-to-ground communications

Description

Customers are used to access internet, to have calls, to download videos, etc. during their everyday movements in the public transportation system (metro, busses, trains, etc.). However, such services are not widespread during flights.

This use case describes the provision of traditional services to customers in private or public airplanes during all the phases of the flight: take-off, flight at high altitude and landing. In order to offer continuous coverage of the plane different technologies may be considered, for instance, legacy 4G or 5G systems when the airplane is flying over dry land, while using satellite connection over sea. Today system use, for instance, WiFi connections inside the plane and satellite connections to join the external networks. However, spectrum for satellite communications is scarce and deploying additional satellite for increasing the system throughput has elevated costs. Providing 4G or 5G connectivity to airplanes over continents is being considered today to be a more cost-efficient solution for improving on-board services.

This use case is also described in 3GPP TR 22.891 [1] in Section 5.29 (Higher user mobility) and Section 5.66 (Broadband direct air to ground communications) and it attracts considerable attention because it is linked to a precise business case. As a matter of fact, air carrier companies constantly search for services which can differentiate them from the competitors and give them a competitive advantage or reinforce their leadership. Providing on-board broadband wireless services is definitively one of such differentiating services.

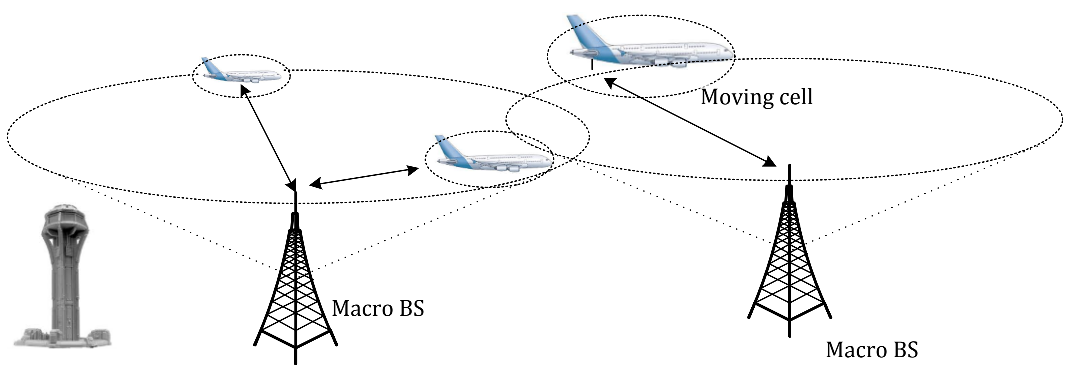

Figure 3-17 Delivery of services in public or private transportation

Challenges and innovation

One big challenge for providing 4G/5G access into planes in every phase of the flight is the validation of the transmission and reception RF equipment in the airplane environment which is subject to very strict rules for guaranteeing maximal security to passengers. The airplane has a number of other wireless communication systems which may generate non-negligible interference on cellular systems bands due to the proximity of RF equipment and to the used transmission powers. Vice versa, the cellular transmission system must control outband emissions on the receivers of the other wireless communication systems.

Another challenge is the optimization of the tradeoff between quality of service and cost for the deployment of a base station network for supporting air-to-ground communications, as well as management of spectrum [10].

Moreover, the extreme speeds and the specific features of the air-to-ground communication channel must be taken into account. First trials are quite encouraging [43] and precommercial solutions are available which state that adaptations to LTE are necessary to support the specific features of air-to-ground communications [10]. 3GPP LTE has in fact been designed for speeds up to 500 km/h which are under typical aeronautical speeds.

Management of customers handover is of course another feature which can be optimized with respect to current systems. Management of different RAT (e.g. cellular network and satellite) is also necessary.

Finally, another challenge could be the management of connections in proximity of big airport, hubs, where the number of airplanes to be served may be very high.

Benefits / key capabilities

- Provide standard unicast and multicast services to customers in airplanes in all phases of the flight with enhanced quality of experience.
- Support communications up to 1200 km/h.
- Support air-to-ground communications, i.e. aeronautical propagation channels.

3.6.2 Use case 6.MR: Delivery of services in public or private transportation in urban areas

Description

Providing services in public transportation like busses, trains, trams is essential for satisfying customers. While a lot of studies and systems are already present, the future evolution of 5G wireless systems, as well the increased demands from customers, will present new challenges to be addressed.

For instance, busses will evolve in urban environment that will see a proliferation and densification of small cells and distributed remote radio heads, going towards ultradense heterogeneous deployments. New and old techniques for increasing the data rate like massive MIMO, CoMP and distributed MIMO techniques will be used. In this environment with high interference, guaranteeing a good backhaul connection will be a challenge. A similar problem is experienced also by the access link inside the bus, depending on the band and technology (WiFi, LTE, other) used [128]. Increased coordination between cells and these moving entities will be probably needed.

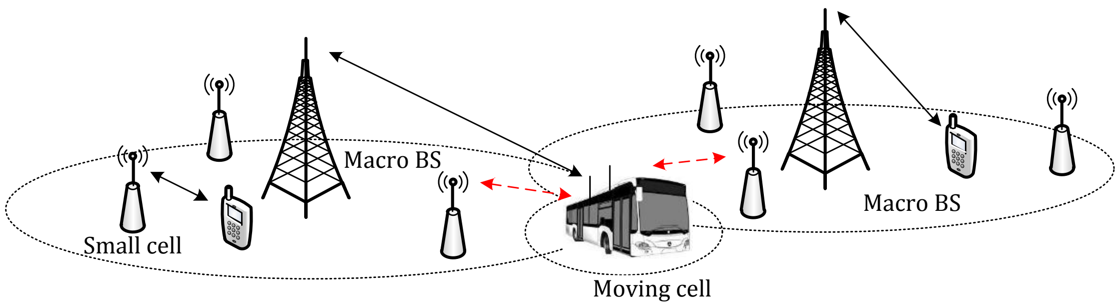

Figure 3-18 Delivery of services in public or private transportation in urban areas

Challenges and innovation

One challenge is to leverage on COHERENT framework in order to improve the support of moving relays, depending on their environment. For instance, a better management of the interference in dense HMNs in urban environment can be beneficial for better supporting moving relays installed on busses.

Another channel is the management of hand-over and mobility of both the moving relay and the users attached to it. As noticed by [128], in dense urban environment, busses can be very close and unwanted handovers between moving relays can be an issue. The same situation happens at bus stops, where users may be attached to the wrong bus.

Benefits / key capabilities

- Improve the management of mobility of moving relays and moving cells (i.e. the users attached to the moving relay).
- Improve the management of interference for moving cells, especially in dense heterogeneous network scenarios.

3.6.3 Use case 6.CA: Caching at the edges

Description

Customers are used to benefit from a variety of services of the wireless cellular networks while being in the public transportation. In order to improve the efficiency of the system, a content caching entity close to the users, typically on the public transportation vehicle, can help in improving the delivery of multi-cast services. This use case is partially overlapped with Section 5.36 (In network and device caching) in [1]. We consider caching by focusing on public transportation vehicles, since they are good candidates for caching content, e.g. multimedia flows, in order to increase the spectral efficiency of the network. In fact customers access typically to popular content, like informations about public transportation network status, news, popular videos, etc. Caching at the relay node may improve the quality of experience of users and at the same time alleviate the burden on the wireless backhaul link.

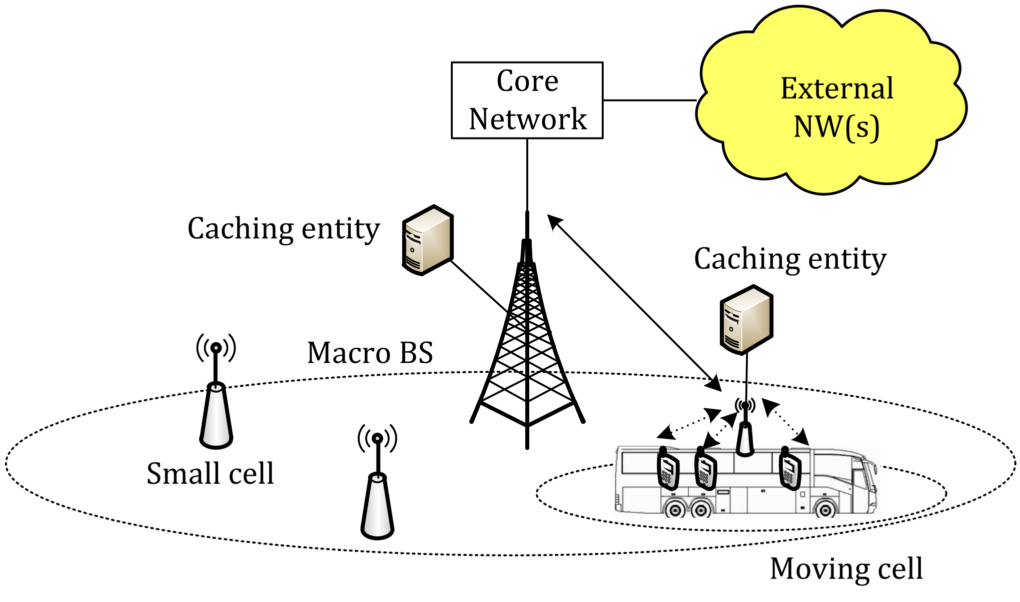

Figure 3-19 Content caching for public transportation services

Challenges and innovation

Content caching for wireless networks has received increasing attention in recent years. Studies are already presents about the good placement of caching entities in 3GPP cellular network architecture, see for instance [137]. One challenge with moving cells is to understand if there is a real gain and how much gain if the caching entity is put at the moving relay side. Gains are to be quantified in different ways, for instance throughput of nodes under coverage for the moving cells, throughput gains also on the backhaul link.

Recently content caching has been studied in the framework of software defined networking for LTE wireless networks [78]. Another challenge is to see how to insert content caching for moving cells in a SDN framework and see if further optimization of the system can be achieved by this.

Benefits / key capabilities

- Improve spectral efficiency, especially of multi-cast services for instance by exploiting local caching of the information at the relay node.
- Enable a flexible deployment of content caching entity in nodes belonging to the network, but also in nodes which may belong to an end-user (relays on busses).
- Enable the deployment of multiple content caching entities and their coordination in order to serve users.
- Improve quality of experience.

3.7 Summarizing COHERENT scenarios and use cases

In this section, we summarize the proposed scenarios and use cases that COHERENT intend to investigate. The first scenario is about *network cooperation*. In particular, we focus on RAN sharing among heterogeneous mobile network, infrastructure sharing, as well as enabling co-operation among operators via using cloud controllers. The second scenario is about *spectrum management*. In particular, we present a use case for distributed antenna system in small cell deployments, as well as a use case for flexible spectrum access. Our third scenario focuses on *critical communications*. Our use cases present coordination within mesh networks, resource sharing for broadband PMR networks, as well as coverage extension and support of D2D communications. The fourth scenario is related to *network slicing*. We present four use cases related to PMR services offered by an MVNO, public safety applications using eMBMS, as well as differentiation (in terms of user groups and services) in multimedia content provision. The next scenario is about *massive IoT*, in which a use case about wireless over power line is presented. Our last scenario describes *broadband communications in public transportation*. The included use cases present air-to-ground communications, service delivery in urban areas, as well as enabling content caching.

Our proposed scenarios and use cases are briefly presented in Table 3-1. For simplicity, we use acronyms per scenario / use case, which will be used as a reference for the next section, which is about requirements.

Table 3-1 Summary of COHERENT scenarios and use cases

Number of Scenario / Use case	Scenario / Use case Description	Acronym
Scenario 1	**Network co-operation and inter-operability**	**SC1**
Use Case 1.1	RAN sharing among heterogeneous mobile networks	UC1.RS
Use Case 1.2	Supporting RAT sharing	UC1.SR
Use Case 1.3	Cooperation among multi-operators	UC1.CO
Scenario 2	**Spectrum management**	**SC2**
Use Case 2.1	Massive MIMO / distributed antenna system in dense small cell deployments	UC2.MM
Use Case 2.2	Flexible spectrum access	UC2.FSA
Scenario 3	**Critical communications**	**SC3**
Use Case 3.1	Coordination with rapidly deployable mesh networks	UC3.MN
Use Case 3.2	Flexible resource sharing for broadband PMR networks	UC3.FS
Use Case 3.3	Coverage extension and support of out-of-coverage communications (D2D communications)	UC3.CE
Scenario 4	**Network slicing**	**SC4**
Use Case 4.1	Enhanced Mobile Virtual Network Operator (MVNO) for PMR services	UC4.MV

Use Case 4.2	Dynamic eMBMS for public safety applications	UC4.eM
Use Case 4.3	User groups' differentiation in multimedia service provision	UC4.GD
Use Case 4.4	Service differentiation	UC4.SD
Scenario 5	**Massive IoT**	**SC5**
Use Case 5.1	Wireless over power line	UC5.WoPL
Scenario 6	**Enhanced broadband communications in public transportation**	**SC6**
Use Case 6.1	Air-to-Ground Communications	UC6.AG
Use Case 6.2	Delivery of services in public or private transportation in urban areas	UC6.MR
Use Case 6.3	Caching at the edges	UC6.CA

4. COHERENT Requirements

4.1 COHERENT requirements consideration

In this chapter, we will perform an initial analysis of the requirements of the COHERENT system. In particular, we derive a set of requirements based on the type of users, services and applications extracted by the aforementioned scenarios and use cases. The requirements are presented in Table 4-1.

Table 4-1 COHERENT requirements

Number of requirement	Requirement description	Associated scenario / use case
R#1	COHERENT shall ensure the necessary physical and MAC layer abstraction	ALL
R#2	COHERENT shall enable the management of heterogeneous mobile networks	UC3.MN, UC6.MR, UC6.AG
R#3	COHERENT shall support flexible architecture (e.g. centralized and distributed) in order to cope with different use cases	ALL
R#4	COHERENT shall provide flexible and coordinated spectrum management	UC2.FSA, UC3.MN, UC3.FS, UC4.MV, UC5.WoPL, UC6.MR
R#5	COHERENT shall support efficient radio resource management	UC3.MN, UC3.FS UC4.MV, UC4.eM, UC6.MR
R#6	COHERENT shall be aware of spectrum usage	UC2.FSA, UC3.FS, UC3.MN, UC3.CE
R#7	COHERENT shall support programmable control based on the low-layer abstraction	ALL
R#8	COHERENT shall provide well-defined open interfaces and protocols to support programmable control	ALL
R#9	COHERENT shall support PPDR services	UC3,MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM
R#10	COHERENT shall be able to manage/support legacy PMR systems based on known standards (e.g. TETRA)	UC3.FS, UC3.MN, UC4.MV
R#11	COHERENT shall be able to manage future broadband PMR systems	UC3.MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM
R#12	COHERENT shall be able to manage legacy 4G systems and solutions	UC3.FS, UC3.MN, UC3.CE, UC4.MV
R#13	COHERENT shall support typical requirements of PMR networks and services	UC3.MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM
R#14	COHERENT shall be able to dynamically manage computational resources if SDN is used	UC1.RS, UC1.SR, UC3.FS, UC4.MV
R#15	COHERENT shall be able to support heterogeneous RAN over heterogeneous medium (e.g. satellite, ad-hoc networks, cellular over the air networks, cellular over power line networks)	UC1.RS, UC3.MN, UC5.WoPL, UC6.AG
R#16	COHERENT shall be able to dynamically manage storage resources if SDN is used	UC3.FS, UC6.CA
R#17	Network operations in COHERENT, e.g. network configuration and resource management, should be real-time, dynamic, on-demand and automated	ALL

R#18	COHERENT shall reduce the equipment costs, and the operating costs	ALL
R#19	COHERENT shall reduce energy consumption and the corresponding energy costs	ALL
R#20	COHERENT shall enable accountability mechanisms in order to monitor spectrum / resource leases by third parties (e.g. Mobile Virtual Network Operators)	UC1.RS, UC1.SR, UC1.CO, UC3.FS, UC4.MV
R#21	COHERENT shall be extendable to support collaboration different MNOs to jointly use the available resources	UC2.FSA, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#22	COHERENT shall be able to take as an input policies defined by the MNOs and MVNO and their customers for resource sharing	UC2.FSA, UC3.FS, UC4.MV
R#23	COHERENT shall maximize mobile operators' investment on existing sites and available licensed spectrum without buying new spectrum	UC2.MM, UC2.FSA UC5.WoPL
R#24	COHERENT shall use best propagation medium (air, power line) for achieving coverage and energy efficiency for mIoT	UC5.WoPL
R#25	COHERENT architecture shall support separation of user plane and control plane	ALL
R#26	COHERENT architecture shall support virtualization of user plane and control plane	ALL
R#27	COHERENT architecture shall support flexible virtualization of sub-layer components of the user plane and control plane	ALL
R#28	COHERENT shall allow for distributed monitoring, aggregation and analysis at various levels of the architecture, supporting self-organization as well as distributed and centralized control/coordination.	ALL
R#29	COHERENT architecture shall support application servers located in the general Internet, RAN, at the network edge or behind an enterprise router	UC4.MV, UC4.eM, UC4.GD, UC4.SD, UC6.CA
R#30	COHERENT architecture shall support multi-point radio communication of an application server with an UE	UC5.WoPL
R#31	COHERENT shall support moving cells for high speed applications, e.g. mobile broadband communications in airplanes	UC6.AG
R#32	COHERENT shall enable operators to create network slices	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#33	COHERENT shall enable operators to manage network slices	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#34	COHERENT shall enable operators to dynamically slice the network based on pre-defined KPIs	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#35	COHERENT shall enable operators to operate different network slices in parallel	UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#36	COHERENT shall be able to dynamically manage core network resources shared among multiple network slices.	UC4.MV, UC4.eM, UC4.GD, UC4.SD

R#37	COHERENT shall enable operators to identify certain end-points or groups of end-points / subscribers to be associated with particular network slices	UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#38	COHERENT shall enable operators to define required criteria for different markets or services to customise network slices	UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#39	COHERENT shall enable operators to allow authorized third parties to create and scale their own network slice via appropriate APIs	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#40	COHERENT shall enable operators to authorize third parties to manage a network slice configuration via appropriate APIs	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#41	COHERENT shall enable a user equipment to obtain service from a specific network slice during initial selection of network	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#42	COHERENT shall reconnect a user equipment in an idle state to the same network slice which it was connected before	UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#43	COHERENT shall allow a user equipment to transit from one network slice to another network slice belonging to the same operator	UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#44	COHERENT shall enable the operator to manage and supervise network traffic over one or many slices	UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#45	COHERENT shall enable the operator to allow authorized third parties to manage and supervise network traffic over one or many slices	UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#46	COHERENT shall allow management priority for PMR services and in particular PPDR services. This shall be done also over network slices managed by the operator or an authorized third party	UC3.MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#47	COHERENT shall allow management security and privacy of PMR users and in particular governmental users. This shall be done also over network slices managed by the operator or an authorized third party	UC3.MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#48	COHERENT shall enable dynamically management of QoS of PMR users inside network slices managed by the operator or an authorized third party	UC3.MN, UC3.FS, UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#49	COHERENT shall be able to dynamically manage RAN resources shared among multiple network slices	UC1.RS, UC3.MN, UC3.FS, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#50	COHERENT shall guarantee a mode of operation without fixed infrastructure (for instance, among mobiles nodes or base stations and mobile nodes without relying on a fixed infrastructure)	UC3.MN
R#51	COHERENT shall be able to provide minimal services like voice, localization, transmission of pictures with high reliability for PPDR applications (typically after a natural disaster).	UC3.MN
R#52	COHERENT shall be able to manage irregular, non-optimized and possibly instable network topologies with acceptable performance	UC3.MN, UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD, UC6.AG, UC6.MR

R#53	COHERENT shall be able to control and manage interference in heterogeneous networks and also while upgrading the status of the networks (for instance during maintenance for rebooting the network after a major natural disaster, while coexisting with a temporary network)	UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#54	COHERENT shall manage the mobility of nodes in heterogeneous networks with heterogeneous RATs (satellite systems, cellular systems, ad-hoc mobile networks, etc.)	UC3.MN, UC6.MR
R#55	COHERENT shall support out-of-coverage communications for PMR services	UC3.FS, UC3.MN, UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#56	COHERENT shall support extended eMBMS for PMR applications, satisfying PMR requirements	UC3.CE, UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#57	COHERENT shall support dynamic management of DL/UL split and duplexing modes	UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#58	COHERENT shall support management of moving cells or moving relays	UC6.MR, UC6.AG, UC6.CA
R#59	COHERENT shall support management of interference in dense urban heterogeneous networks	UC6.MR, UC3.CE
R#60	COHERENT shall manage multiple-hop relaying in a transparent way for the final user	UC3.CE, UC3.MN
R#61	COHERENT shall guarantee high reliability and acceptably low latency also for multi-hop communications and D2D communications (e.g. for PMR services)	UC3.CE, UC3.MN
R#62	COHERENT shall be able to monitor and control network resources such that spectral efficiency is increased	UC3.CE, UC6.AG, UC6.MR, UC6.CA
R#63	COHERENT shall contribute to interference elimination and provide a significant improvement to cell-edge users' throughput, contributing to their consistent mobile experience	UC2.MM
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	UC2.MM
R#65	COHERENT shall be able to provide a coordination interface in combination with advanced scheduling and MAC protocols	UC2.MM
R#66	COHERENT shall be able to support centralized base station equipment that is separated from the associated RRHs (remote radio heads) to support a distributed antenna system (DAS)	UC2.MM
R#67	COHERENT shall be able to provide control and coordination interface between neighbouring cells / clusters	UC2.MM
R#68	COHERENT shall guarantee privacy of user subscription data for network slices	UC3.FS, UC3.CE UC4.MV, UC4.eM, UC4.GD, UC4.SD
R#69	COHERENT shall have a fall-back mode of operation in case of natural disaster or when the fixed infrastructure is out-of-order	UC3.MN, UC3.CE

R#70	COHERENT shall be able to monitor and control network resources such that reliability, availability and network resilience are improved if requested	ALL
R#71	COHERENT shall support flexible spectrum usage schemes depending on the various metrics and parameters gathered by the operator	UC2.FSA, UC5.WoPL
R#72	COHERENT shall coordinate the use of different flexible spectrum modes per time-frequency resource to avoid interferences higher than perceived in normal operation	UC2.FSA, UC5.WoPL
R#73	COHERENT shall support the change of spectrum utilization strategy either on a millisecond basis (for, e.g., activation/deactivation of flexible duplexing scheme) or in much longer scale	UC2.FSA,
R#74	COHERENT shall support the change of the spectrum utilization strategy and be transparent to the end user.	UC2.FSA, UC5.WoPL
R#75	COHERENT shall support the flexible spectrum access to prioritize services for emergency situations in public safety scenarios.	UC2.FSA
R#76	COHERENT shall have the capability of supporting different spectrum access requirements of multiple tenants in the system.	UC2.FSA, UC5.WoPL
R#77	COHERENT shall have the capability of cost efficiency in connecting the mIoT devices	UC5.WoPL
R#78	COHERENT shall have the capability of energy efficiency in connecting the mIoT devices	UC2.FSA, UC5.WoPL
R#79	COHERENT shall have the capability of spectrum efficiency in connecting the mIoT devices	UC2.FSA, UC5.WoPL
R#80	COHERENT shall support autonomous and distributed self-organizing infrastructure capabilities.	ALL

4.2 Mapping COHERENT requirements with other standards

4.2.1 Mapping with NGMN

A comprehensive list of requirements are identified by NGMN [105]. As shown in Figure 4-1, it is categorised into six dimensions, namely

- User experience
- System performance
- Device
- Enhanced service
- New business model
- Network deployment, operation and management

In this section, a subset of COHERENT requirements are mapped into these six dimensions in NGMN. We do not include all our proposed requirements, since a fraction of them either could be mapped to more than one NGMN requirements group, or they are not explicitly related to them.

4.2.1.1 User Experience

Consistent user experience

- R#63 COHERENT shall contribute to interference elimination and provide a significant improvement to cell-edge users' throughput, contributing to their consistent mobile experience

Mobility

- R#31 COHERENT shall support moving cells for high speed applications, e.g. mobile broadband communications in airplanes
- R#54 COHERENT shall manage the mobility of nodes in heterogeneous networks with heterogeneous RATs (satellite systems, cellular systems, ad-hoc mobile networks, etc.)
- R#58 COHERENT shall support management of moving cells or moving relays

Figure 4-1 NGMN 5G requirements

4.2.1.2 System Performance

Spectrum Efficiency

- R#62 COHERENT shall be able to monitor and control networks resources such that spectral efficiency is increased
- 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

Connectivity Transparency

- R#2 COHERENT shall enable the management of heterogeneous mobile networks
- R#15 COHERENT shall be able to support heterogeneous RAN over heterogeneous medium (e.g. satellite, ad-hoc networks, cellular over the air networks, cellular over power line networks)
- R#12 COHERENT shall be able to manage legacy 4G systems and solutions
- R#43 COHERENT shall allow a user equipment to transit from one network slice to another network slice belonging to the same operator
- R#60 COHERENT shall manage multiple-hop relaying in a transparent way for the final user

4.2.1.3 Enhanced service

Security

- R#40 COHERENT shall enable operators to authorize third parties to manage a network slice configuration via appropriate APIs
- R#46 COHERENT shall allow management priority for PMR services and in particular PPDR services. This shall be done also over network slices managed by the operator or an authorized third party
- R#47 COHERENT shall allow management security and privacy of PMR users and in particular governmental users. This shall be done also over network slices managed by the operator or an authorized third party
- R#68 COHERENT shall guarantee privacy of user subscription data for network slices

Resilience and high availability

- R#50 COHERENT shall guarantee a mode of operation without fixed infrastructure (for instance, among mobiles nodes or base stations and mobile nodes without relying on a fixed infrastructure)
- R#51 COHERENT shall be able to provide minimal services like voice, localization, transmission of pictures with high reliability for PPDR applications (typically after a natural disaster).
- R#52 COHERENT shall be able to manage irregular, non-optimized and possibly instable network topologies with acceptable performance
- R#55 COHERENT shall support out-of-coverage communications for PMR services
- 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#69 COHERENT shall have a fall-back mode of operation in case of natural disaster or when the fixed infrastructure is out-of-order

4.2.1.4 New business model

Partner service provider and XaaS assert Provider

- R#20 COHERENT shall enable accountability mechanisms in order to monitor spectrum / resource leases by third parties (e.g. Mobile Virtual Network Operators)
- R#38 COHERENT shall enable operators to define required criteria for different markets or services to customise network slices

4.2.1.5 Network deployment, operation and management

Cost Efficiency

- R#18 COHERENT shall reduce the equipment costs, and the operating costs
- R#23 COHERENT shall maximize mobile operators' investment on existing sites and available licensed spectrum without buying new spectrum

Energy Efficiency

- R#19 COHERENT shall reduce energy consumption and the corresponding energy costs

Flexibility and scalability

- R#4 COHERENT shall provide flexible and coordinated spectrum management
- R#5 COHERENT shall support efficient radio resource management
- R#14 COHERENT shall be able to dynamically manage computational resources if SDN is used
- R#17 Network operations in COHERENT, e.g. network configuration and resource management, should be real-time, dynamic, on-demand and automated.
- R#27 COHERENT architecture shall support flexible virtualization of sub-layer components of the user plane and control plane
- R#34 COHERENT shall enable operators to dynamically slice the network based on pre-defined KPIs
- R#36 COHERENT shall be able to dynamically manage core network resources shared among multiple network slices
- R#48 COHERENT shall enable dynamically management of QoS of PMR users inside network slices managed by the operator or an authorized third party
- R#49 COHERENT shall be able to dynamically manage RAN resources shared among multiple network slices.
- R#57 COHERENT shall support dynamic management of DL/UL split and duplexing modes

Operations awareness

- R#1 COHERENT shall ensure the necessary physical and MAC layer abstraction
- R#6 COHERENT shall be fully aware of spectrum usage

- R#37 COHERENT shall enable operators to identify certain end-points or groups of end-points / subscribers to be associated with particular network slices

Operation efficiency

- R#3 COHERENT shall support flexible architecture (e.g. centralized and distributed) in order to cope with different use cases.
- R#7 COHERENT shall support programmable control based on the low-layer abstraction
- R#8 COHERENT shall provide well-defined open interfaces and protocols to support programmable control
- R#21 COHERENT shall be extendable to support collaboration different MNOs to jointly use the available resources
- R#22 COHERENT shall be able to take as an input policies defined by the MNOs and MVNO and their customers for resource sharing
- R#25 COHERENT architecture shall support separation of user plane
- R#26 COHERENT architecture shall support virtualization of user plane and control plane
- R#29 COHERENT architecture shall support application servers located in the general Internet, RAN, at the network edge or behind an enterprise router
- R#30 COHERENT architecture shall support multi-point radio communication of an application server with an UE
- R#32 COHERENT shall enable operators to create network slices
- R#33 COHERENT shall enable operators to manage network slices
- R#35 COHERENT shall enable operators to operate different network slices in parallel
- R#39 COHERENT shall enable operators to allow authorized third parties to create and scale their own network slice via appropriate APIs
- R#41 COHERENT shall enable a user equipment to obtain service from a specific network slice during initial selection of network
- R#42 COHERENT shall reconnect a user equipment in an idle state to the same network slice which it was connected before
- R#44 COHERENT shall enable the operator to manage and supervise network traffic over one or many slices
- R#45 COHERENT shall enable the operator to allow authorized third parties to manage and supervise network traffic over one or many slices
- R#53 COHERENT shall be able to control and manage interference in heterogeneous networks and also while upgrading the status of the networks (for instance during maintenance for rebooting the network after a major natural disaster, while coexisting with a temporary network)
- R#59 COHERENT shall support management of interference in dense urban heterogeneous networks
- R#65 COHERENT shall be able to provide a coordination interface in combination with advanced scheduling and MAC protocols
- R#66 COHERENT shall be able to provide centralized base station equipment that is separated from the associated RRHs (remote radio heads) to support a distributed antenna system (DAS).
- R#67 COHERENT shall be able to provide control and coordination interface between neighbouring cells / clusters

4.2.2 Mapping with 3GPP

3GPP has provided a study of use cases in 5G networks, with identifying the related high level potential requirements. In this section, the mapping of COHERENT use cases to 3GPP is provided. Furthermore, the corresponding requirements of each 3GPP use case and its mapped COHERENT use cases are listed as well.

4.2.2.1 Mapping of use cases and scenarios

Table 4-2 Mapping COHERENT to 3GPP uses cases

3GPP Use cases	**SC 1 Network Cooperation**			**SC 2 Spectrum Mgt**		**SC 3 Critical Comm.**			**SC 4 Network Slicing**				**SC 5 Massive IoT**	**SC 6 Public Transportation**		
	U1. RS	U 1. S R	U1. CO	U2. MM	U2. FSA	U3. MN	U3. FS	U3. CE	U4. MV	U4.e M	U4. GD	U4. SD	U5. Wo PL	U6. AG	U6. MR	U6. CA
Mobile broadband for hotspots scenario				X												
On-demand networking	X	X			X	X	X	X		X						
Mobile broadband services with seamless wide-area coverage														X	X	X
Improvement of network capabilities for vehicular case															X	X
Connectivity Everywhere						X	X	X						X	X	X
Higher User Mobility														X	X	X
Ultra reliable communication						X	X	X	X	X						
Local UAV Collaboration						X		X								
Network slicing							X		X	X	X	X				
Lifeline communications / natural disaster						X	X	X								
Routing path optimization when server changes														X	X	X
Network capability exposure									X							

Ad-Hoc Broadcasting										X						
Flexibility and scalability	X	X	X		X								X			
Network enhancements to support scalability and automation	X	X														
In-network caching														X	X	X
Wireless Self-Backhauling						X		X								
Multi Access network integration	X	X														
Multiple RAT connectivity and RAT selection	X	X														
Temporary Service for Users of Other Operators in Emergency Case						X	X	X								
Migration of services from earlier generations	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Coexistence with legacy systems	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X

4.2.2.2 Mapping of Requirements

Table 4-3 Mapping COHERENT to 3GPP requirements

3GPP Use cases	***Requirements***	***COHERENT use cases***
Mobile broadband for hotspots scenario	- Support of the user experienced data rate up to Gbps of level while the user is moving slowly. - Support of the peak data rate at tens of Gbps while the user is moving slowly. - Support of the whole traffic volume in the area at least the level of Tbps/ km^2.	UC2.MM, UC2. FSA

<table>
<tr><td></td><td>- Support of flexible and efficient backhaul especially outdoor.</td><td></td></tr>
<tr><td>On-demand networking</td><td>- Support of the peak data rate at tens of Gbps while the user is moving slowly.
- Provide guaranteed user experience for mobile broadband services like live video in areas with a high UE density which requires user experienced data downlink rate of 300Mbps and uplink rate of 50Mbps in 200-2500 /km^2 connection density.
- Provide consistent users experience when terminals enter the areas with a high UE density which requires user experienced data downlink rate of 50Mbps and uplink rate of 25Mbps in 2000/km^2 connection density.
- Adjust the network capacities dynamically based on the variation of demand and performance indicators.</td><td>ALL</td></tr>
<tr><td>Mobile broadband services with seamless wide-area coverage</td><td>For wide area coverage, support of user experienced data rate for mobile broadband services anytime and anywhere, e.g., 100Mbps. The above requirement assumes reuse of an existing base station site grid.
- Support fast-moving end-users, e.g., 500km/h.
- Support high connection density for high speed scenarios, e.g., 500 active UEs simultaneously.
- Support low latency for high speed scenario.</td><td>UC3.MN, UC3.FS, UC3.CE, UC6.AG, UC6.MR, UC6.CA</td></tr>
<tr><td>Improvement of network capabilities for vehicular case</td><td><table>
<tr><th>Application</th><th>Average End User Throughput</th><th>Latency (end-to-end)</th><th>Latency (over the air)</th></tr>
<tr><td>High Definition Video 8K (streaming)</td><td>< 100 Mbps (DL)</td><td>< 1 s</td><td>< 200 ms</td></tr>
<tr><td>High Definition Video (conversational)</td><td>< 10 Mbps (DL/UL)</td><td>< 150 ms</td><td>< 30 ms</td></tr>
<tr><td>Cloud Computer Games with 4K 3D graphics – Low Latency Applications</td><td>< 50 Mbps (DL/UL) (UL is needed for multiplayer game computation in user device)</td><td>< 7.5 ms (10 times less than in for real time games)</td><td>< 1.5 ms</td></tr>
</table></td><td>UC3.MN, UC3.FS, UC3.CE, UC6.AG, UC6.MR, UC6.CA</td></tr>
<tr><td>Connectivity Everywhere</td><td>- Provide aerial object and/or nautical objects with reliable mobile broadband connectivity.
- Provide reliable low-latency connectivity between aerial objects.</td><td>UC2.MM</td></tr>
<tr><td>Higher User Mobility</td><td>- Support enhanced mobile broadband services in fast moving vehicles (e.g. up to 500 km/h) with enhanced user experience.
- Support enhanced connectivity services in fast moving airplanes (e.g. up to 1000 km/h) with enhanced user experience.</td><td>UC1.RS, UC1.SR, UC1.CO, UC2.MM</td></tr>
<tr><td>Ultra reliable communication</td><td>- Services in this category require very low data error rate. Some of them also require very low latency, i.e. for industrial automation with delays of one ms</td><td>UC3.MN, UC3.FS, UC3.CE</td></tr>
<tr><td>Local UAV Collaboration</td><td>- Latency of 10 ms as collaboration requires vehicle altitude and position control loops to synchronize. Latency is required on the order of the control loop bandwidths.
- Near 100% reliability as instability and crashing of UAV could result from loss of communications. Control functions depend on this communication.
- Security to be provided at the level for current aviation Air Traffic Control (ATC) for command and control of vehicles in controlled airspace.</td><td>UC3.MN, UC3.CE</td></tr>
</table>

	- Priority, Precedence, Preemption (PPP) needed as failure to transmit communications in reliable and timely manner could result in loss of property or life. - Position accuracy within 10 cm due to multiple UAVs that may need to collaborate in close proximity to one another.	
Network slicing	- Based on operator's policy, the system shall be able to define minimal services necessary in case of disaster that are conditional on e.g. subscriber class (i.e. access class), communication class (i.e. emergency call or not), device type (i.e. Smart phone or IoT device), and application. Examples of those minimal services are communications from specific high priority users, emergency calls, and a disaster-message-board type of application that helps people reconnect with friends and loved ones in the aftermath of disasters. - Those minimal services shall be available in case of disaster. - During the recovery phase of disaster, the service continuity of those minimal services that start being provided should be ensured.	UC4.MV, UC4.eM, UC4.GD, UC4.SD
Lifeline communications / natural disaster	- Based on operator's policy, the system shall be able to define minimal services necessary in case of disaster that are conditional on e.g. subscriber class (i.e. access class), communication class (i.e. emergency call or not), device type (i.e. Smart phone or IoT device), and application. Examples of those minimal services are communications from specific high priority users, emergency calls, and a disaster-message-board type of application that helps people reconnect with friends and loved ones in the aftermath of disasters. - Those minimal services shall be available in case of disaster. - During the recovery phase of disaster, the service continuity of those minimal services that start being provided should be ensured.	UC3.MN, UC3.FS, UC3.CE
Routing path optimization when server changes	- Subject to the service agreement between the operator and the service provider, the network shall enable hosting of services (including both MNO provided services and 3rd party provided services) closer to the end user to improve user experience and save backhaul resources. - Support routing of data traffic to the entity hosting services closer to the end user for specific services of a UE. - Enable efficient user-plane paths between a UE and the entity hosting the service closer to the end user even if the UE changes its location during communication. - Enable charging, QoS, and Lawful Interception (LI) for services hosted closer to the end user.	UC1.RS, UC1.SR, UC1.CO
Network capability exposure	- To provide the network capabilities to the 3rd party ISP/ICP. - Service providers can be capable of configuring and managing the service via e.g. open API, while operators will have the option to manage and evolve the network.	UC1.RS, UC1.SR, UC1.CO, UC2.MM, UC2. FSA
Ad-Hoc Broadcasting	- Support an interface from an external Ad-Hoc Broadcast management system that manages broadcast requests. - Reserve groups of resources temporarily for scheduled Ad-Hoc Broadcasts. - Allow the UE to receive broadcasts selected by the user (from the Ad-Hoc Broadcast management system) in accordance with any appropriate authorisations.	UC3.MN, UC3.FS, UC3.CE, UC6.AG, UC6.MR, UC6.CA
Flexibility and scalability	- Using resources (compute, network and storage resources) in more than one geographic area by the system shall be supported without requiring manual re-configuration of neighbouring nodes, without	ALL

	service disruption, and while avoiding additional signalling due to unnecessary UE's re-attachments (e.g. due to loss of call state information in the network).	
Network enhancements to support scalability and automation	- Guarantee the service experience of the subscribers during the network scaling and automation operation. - Moreover, the existing mechanisms (e.g. load balancing, network function selection) which are closely related with the effect of the network scalability and automation operation need to be enhanced.	UC1.RS, UC1.SR, UC1.CO
In-network caching	- Efficiently deliver or forward content from in-network entities controlled by the operator. - Provide charging, Lawful Interception (LI) and QoS differentiation for content delivered from an in-network caching entity. - Enable a flexible deployment of content caching entity located at multiple locations within the network (e.g. at various radio sites and local aggregation points). - Support a content caching entity that is capable of being integrated within a device under the control of the operator. - Authorized UE shall be able to receive cached content broadcasted by content caching entity. - Efficient delivery of content from an appropriate caching entity, e.g. a cache located close to the user.	UC6.AG, UC6.MR, UC6.CA
Wireless Self-Backhauling	- Flexible partitioning of resources between access and backhaul functions when supported in a common band, including quasi-static provisioning of separate access and backhaul resources, and dynamic allocation of access and backhaul resources, e.g., based on current local conditions. - Autonomous neighbour discovery and link setup, self-configuration of addressing and forwarding plane, and autonomous integration into core/OAM. - Use of multiple RATs to increase service availability and network resiliency. - Multihop wireless network topologies. - Network topologies with redundant connectivity and paths to minimize service disruptions due to network dynamics. - Dynamic adaptation to topology changes (e.g., due to node additions, node failures, link fluctuations.)	UC1.RS, UC1.SR, UC1.CO, UC3.MN, UC3.CE
Multi Access network integration	- Enable the capability to connect to multiple non-3GPP and 3GPP access networks in order to allow the operator to improve the efficiency in the exploitation of the network infrastructure and to provide the best capabilities to end-user. - Support at least mobility between 3GPP and non-3GPP networks with optional session continuity, capability for the UE based on network control to select the access to connect to.	UC1.RS, UC1.SR, UC1.CO, UC2.MM, UC2. FSA
Multiple RAT connectivity and RAT selection	- Provide data transmission by using both the 5G New RATs and E-UTRA simultaneously. - Select a radio access (either a 5G New RAT or E-UTRA) to assign each data flow, taking into account e.g. service, traffic characteristics, radio characteristics, and UE's moving speed.	UC1.RS, UC1.SR, UC1.CO, UC2.MM, UC2. FSA
Temporary Service for Users of Other Operators in Emergency Case	- Support temporary service for users of other than home operators as temporary users in emergency case by serving operator policy. - Support defining the limited set of necessary communication services and acceptable terminal features for temporary users by serving operator policy. - Support an appropriate level of communications security for temporary service.	UC3.MN, UC3.FS, UC3.CE

Migration of services from earlier generations	- Support the following operational requirement defined in previous releases of EPS: RAN Sharing	UC1.RS, UC1.SR, UC1.CO, UC2.MM, UC2. FSA
Coexistence with legacy systems	- Support seamless handover and Inter System Mobility between 5G RAT(s) and E-UTRAN. - Seamless handover between the 5G RAT(s) and GERAN or UTRAN is not required.	UC1.RS, UC1.SR, UC1.CO, UC2.MM

5. Vision from SDN perspective of COHERENT architecture design

This section presents a vision from SDN perspective of COHERENT architecture design, which will enable the development of the COHERENT scenarios and use cases in the future. Note that the deliverable D2.2 “System architecture and abstractions for mobile networks”, which is part of Task 2.2 “System architecture”, will provide the detailed system architecture used within the COHERENT project.

The key innovation of COHERENT lies in the unified and programmable control framework for 5G heterogeneous radio access networks. The COHERENT architecture is driven by a key characteristic, namely flexibility. Software-Defined Networking (SDN) has emerged as a new intelligent architecture for network programmability. By leveraging SDN concept, COHERENT architecture achieves flexible networking by providing programmable control framework for 5G heterogeneous radio access networks, where the network functionalities and RAT configurations could be tailored to various use cases. Therefore, the COHERENT architecture is designed by utilising the reference SDN architecture proposed in RFC 7426 as a base. It is further devised with the aim to fundamentally improve the control and coordination among heterogeneous radio access networks, and to enable an open control framework which evolves with new radio access techniques.

This section is structured as follows. In Section 5.1, we present an overview of the RFC 7426 reference architecture and terminology. Following this, Section 5.2 introduces a conceptual view of COHERENT SDN architecture. The design of COHERENT architecture is under continuous consideration and will be presented in deliverable D2.2, which is part of Task 2.2 “System architecture”. Finally, a preliminary vision of COHERENT slicing and resource mapping is described in Section 5.3.

5.1 Reference SDN architecture in RFC 7426

In this section, we present an overview of the RFC 7426 reference architecture and terminology. RFC 7426 provides a concise reference for SDN architecture and its terms for the SDN research community. Figure 5-1 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. Note that the service and the applications may be implemented in a modular or distributed way in any of architectural layers. Therefore, they could span multiple planes as shown in Figure 5-1. Interested readers please refer to RFC 7426 for more details.

5.2 COHERENT SDN architecture design aspects

The COHERENT SDN architecture is inspired by the insights on the abstraction of low-layer states, behaviors and functions, with the aim to fundamentally improve the control and coordination among heterogeneous radio access networks, and to enable an open control framework which evolves with new radio access techniques. The COHERENT SDN architecture spans multiple planes as depicted in Figure 5-2, including data plane, control plane and application plane. Additionally, two abstraction layers are proposed for COHERENT SDN architecture, namely the ***infrastructure resource abstraction layer*** and ***network service abstraction layer***. The infrastructure resource abstraction layer abstracts the underlying physical and MAC layer to the control plane, while the network service abstraction layer provides service abstractions for the applications and services.

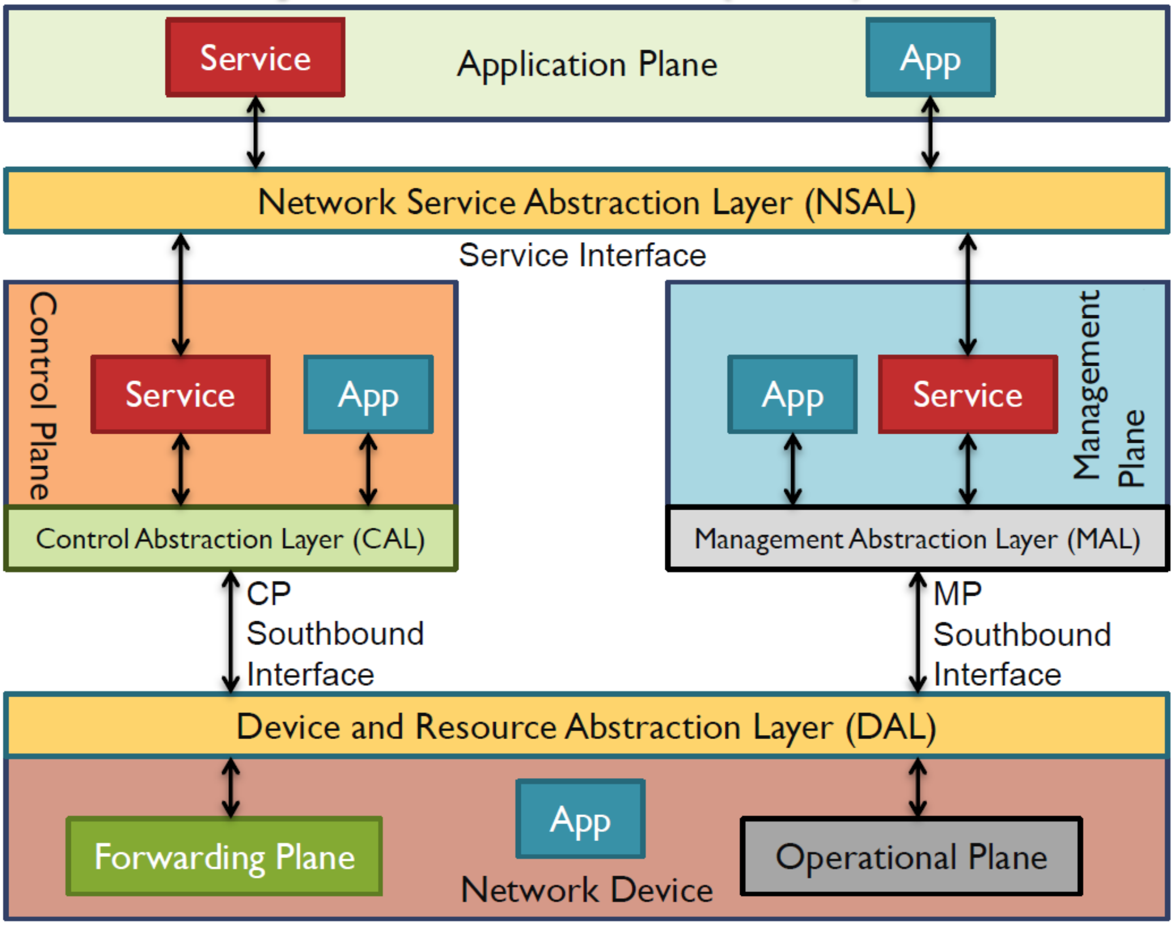

Figure 5-1 RFC 7426 SDN layer architecture

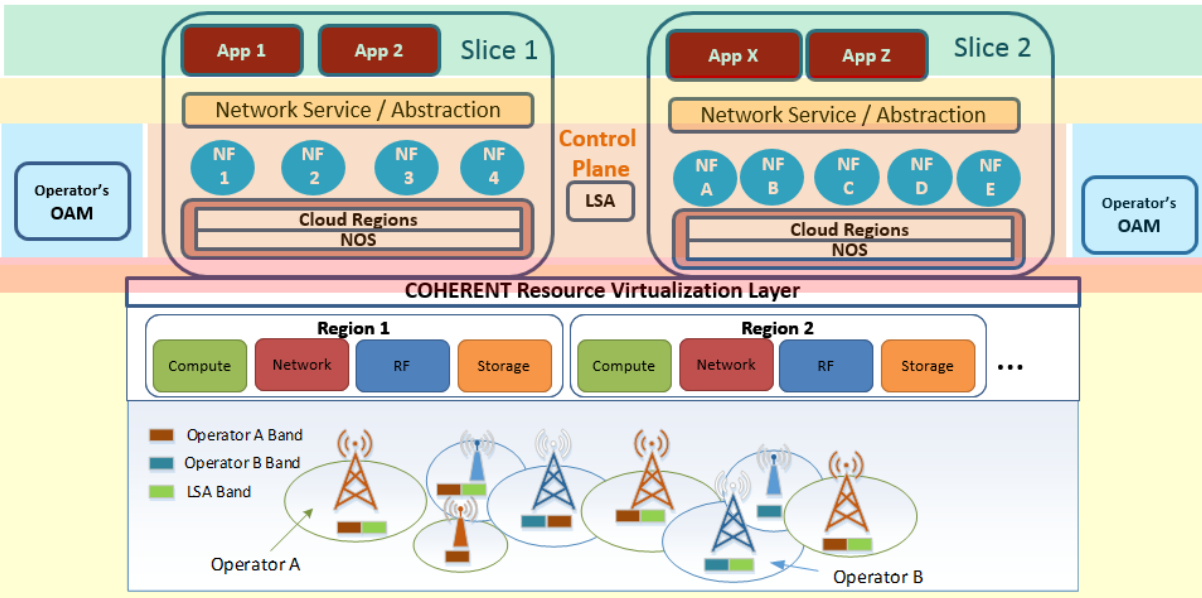

Figure 5-2 COHERENT preliminary high-level SDN architecture

5.2.1 Infrastructure resource abstraction layer

A key element in the preliminary COHERENT SDN architecture is the introduction of an ***infrastructure resource abstraction layer*** between data and control planes. The separation of data and control planes allows the applications to programmatically control the heterogeneous mobile networks with lower complexity. The ***infrastructure resource*** could be the resources of a mobile network, comprising access nodes, cloud nodes (processing or storage resources), networking nodes and associated links. 5G devices (e.g. mobile handheld devices, IoTs) are also considered in COHERENT as the infrastructure resource since they may act as a relay/ hub or a computing/storage resource, e.g. D2D communications.

Through the infrastructure resource abstraction layer, the infrastructure resources are exposed to higher layers and to the end-to-end management and orchestration entity. The abstraction may be expressed by one or more abstraction models. Examples of forwarding-plane abstraction models are Forwarding and Control Element Separation (ForCES) [62], OpenFlow [108], YANG model [24], and SNMP MIBs [116], which are designed for the fixed networks.

The COHERENT intends to deal with the insufficiency of the abstraction models for heterogeneous radio access networks. More specifically, this layer provides the abstraction of the low-layer (physical layer and MAC layer) network states and behaviours of different underlying mobile networks. For example, the abstraction of spectrum usage in physical and MAC layer could enable necessary and affordable spectrum usage information for more efficient and flexible spectrum management by operators, spectrum regulators or other players in this domain. Furthermore, proper abstraction of physical and MAC layer will significantly reduce the signalling to implement physical layer cooperative technologies and the coordination between network entities for more efficient and scalable spectrum management and interference management.

5.2.2 Control plane

The control plane in COHERENT SDN architecture configures the data plane according to the environment (e.g. channel condition) or an operator’s policy (e.g. middleboxes, application types). The COHERENT controller consists of a Network Operating System (NOS) running a collection of application modules, such as radio resource management for RAN sharing and spectrum sharing, mobility management, and traffic steering. The handling of data plane often requires multiple application modules. Therefore, the NOS should coordinate the application modules and unify a single set of forwarding decisions in each network devices. Owing to the distributed nature of control plane, the control plane can span over multiple cloud regions. The concept of the cloud region could be in different aspects, e.g. geographical region, radio access networks v.s. core networks, multi-operators, and multi-RAT.

5.2.3 Application plane and network service abstraction layer

Applications and services that use services from the control plane form the application plane in the COHERENT SDN architecture. COHERENT will provide application-centric network service abstraction in order to shield the upper applications layer and users from tedious and diverse configurations at the underlying network infrastructure among multiple networking domains. This feature implies a better understanding of global network infrastructure.

5.3 COHERENT slicing and resources

A network slice is defined as a collection of specific network functions and RAT configurations, which are aggregated together for some particular use cases or business applications. Therefore, 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 [105]. Different network slices contain different network functions and configuration settings. Figure 5-3 illustrates an example of multiple network slices in COHERENT SDN architecture concurrently operated on the same infrastructure. For example, latency may be critical for a network slice supporting eHealth/Public safety use case. For such a slice, some network functions or storage resources can be located at the edge of the networks. The COHERENT controllers should coordinate the mapping of the infrastructure resources for network slices and the management of the shared infrastructure resources and functions among multiple network slices. One example of a shared function is the scheduler of RAN sharing, which is typically shared among multiple network slices.

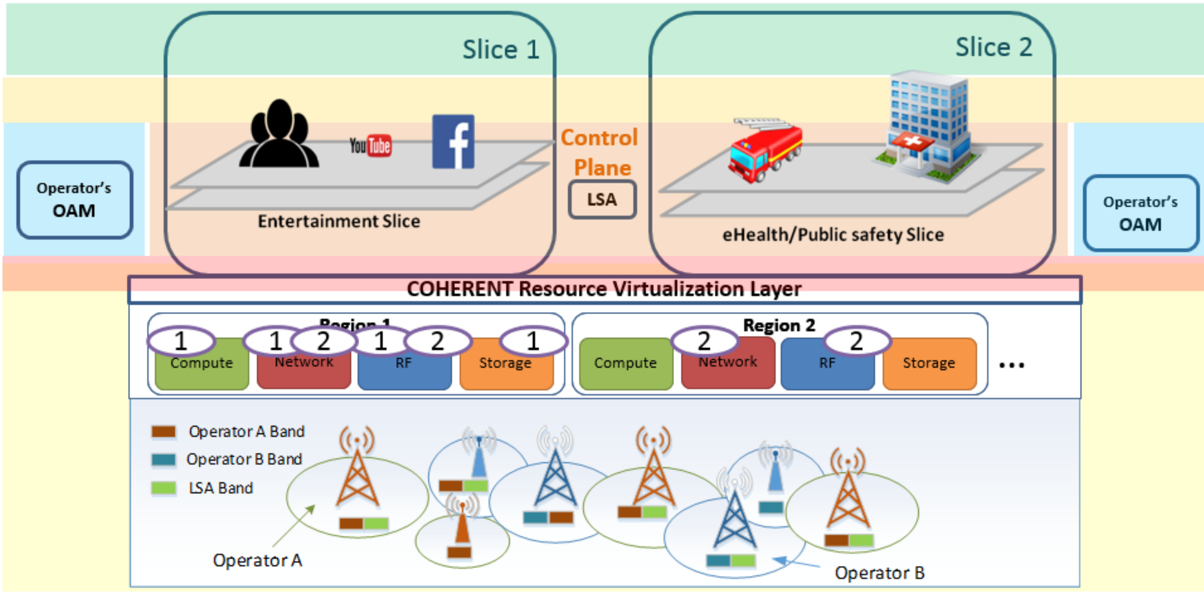

Figure 5-3 COHERENT slicing and resource mapping

6. Conclusions

This document is Deliverable D2.1 “Use Cases & Architecture” for Work Package 2 “System architecture and SDK Development”. WP2 deals with scenarios and use cases definition, requirements and system architecture and SDK development. WP2 provides useful input to WP3 (low-layer abstraction and control), WP4 (spectrum management) and WP5 (programmable control methods and algorithm implementations for inter-cell coordination and resource allocation in HMN).

Current report defines the scenarios and use cases that COHERENT solutions will address. In particular, it provides a detailed view of COHERENT’s reference scenarios and analyses the requirements that need to be fulfilled by the platform in order to properly support the use cases. Finally, using the requirements analysis as input, we investigate design aspects of COHERENT SDN architecture.

Before the presentation of the identified scenarios and use cases, we provide an overview of the enabling technologies for 5G, identify relevant international fora and present research efforts in the area of management techniques for various infrastructures. In particular, 5G requirements, as well as basic resources and KPIs are presented. Additionally, we present technology trends for 5G and summarize the available scenarios and use cases. Our main goal is to revisit the related work, and consider how COHERENT fits within the 5G research context.

After revisiting the related work, we present a wide range of scenarios and use cases that COHERENT solutions will address. The first scenario is about *network cooperation*. In particular, we focus on RAN sharing among heterogeneous mobile network, infrastructure sharing, as well as enabling co-operation among operators via using cloud controllers. The second scenario is about *spectrum management*. In particular, we present a use case for distributed antenna system in small cell deployments, as well as a use case for flexible spectrum access. Our third scenario focuses on *critical communications*. Our use cases present coordination within mesh networks, resource sharing for broadband PMR networks, as well as coverage extension and support of D2D communications. The fourth scenario is related to *network slicing*. We present four use cases related to PMR services offered by an MVNO, public safety applications using eMBMS, as well as differentiation (in terms of user groups and services) in multimedia content provision. The next scenario is about *massive IoT*, in which a use case about wireless over power line is presented. Our last scenario describes *broadband communications in public transportation*. The included use cases present air-to-ground communications, service delivery in urban areas, as well as enabling caching at the edges.

Our next step was to derive a set of requirements based on the type of users, services and applications extracted by the aforementioned scenarios and use cases. We also attempt to provide a mapping between COHERENT requirements and other standards.

Finally, we investigate design aspects of COHERENT SDN architecture in a technology-agnostic manner, which will enable the development of the COHERENT scenarios and use cases in the future. Such proposal will be further elaborated and adapted within deliverable 2.2 “System architecture and abstractions for mobile networks”, which will provide the proposed system architecture of COHERENT. This work is part of Task 2.2 “System architecture” within Work Package 2 “System architecture and SDK Development”.

In order to select and adapt those existing use cases and requirements which are relevant to COHERENT approach, in Annex A we analysed the use cases and especially the requirements available at this time in 3GPP. It should be noted that, on the other side, we have contributed in 3GPP SA1 for influencing some of the 5G requirements.

References

1. 3GPP TR 22.891 Draft, "Feasibility study on new services and markets technology enablers; stage 1 (Release 14)", v1.1.0 (2015-11), November 2015, http://www.3gpp.org/ftp/specs/archive/22_series/22.891.
2. 3GPP S1-153171, "Use case categorization", Vodafone, 3GPP TSG-SA WG1 Meeting #71bis ad-hoc, Vancouver, Canada, 19-21 October 2015
3. 3GPP TR 36.836, "Technical specification group radio access network: mobile relay for evolved universal terrestrial radio access (E-UTRA). Technical report. http://www.3gpp.org.
4. 3GPP TR 36.897 Draft, "Study on elevation beamforming / Full-Dimension (FD) Multiple Input Multiple Output (MIMO) for LTE (Release 13)", v13.0.0 (2015-06), June 2015.
5. 4G Americas, "5G Technology Evolution Recommendations," 4G Americas, http://www.4gamericas.org/files/2414/4431/9312/4G_Americas_5G_Technology_Evolution_Recommendations_-_10.5.15_2.pdf.
6. 4G Americas, 4G Americas' Recommendations on 5G Requirements and Solutions.
7. 5G PPP, "5G Vision: The 5G Infrastructure Public Private Partnership: The next generation of communication networks and services," 5G PPP, February 2015, https://5g-ppp.eu/roadmaps/.
8. A. Checko, *et al.* "Cloud ran for mobile networks - a technology overview." Communications Surveys & Tutorials, IEEE 17.1 (2014): 405-426.
9. A. Gudipati, *et al.* "SoftRAN: Software defined radio access network." Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking. ACM, 2013.
10. Alcatel-Lucent, "Using air-to-ground LTE for in-flight ultra-broadband", strategic white paper, May 2015.
11. Allied Business Intelligent, "The Cellular M2M Module Market", 2007.
12. E. Altman, K. Avrachenkov, and J. Goseling, "Coding for caches in the plane," arXiv preprint: 1309.0604, 2013.
13. D. Amyot and A. Eberlein, "An evaluation of scenario notations and construction approaches for telecommunication system development," Telecommunication Systems, vol. 24, pp. 61-94, no. 1, 2003.
14. J. G. Andrews *et al.*, "What will 5G be?" IEEE Journal on Selected Areas in Communications, vol. 32, pp. 1065–1082, June 2014.
15. A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, H. Tullberg, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project". Communications Magazine, IEEE, 52(5), 26-35, 2014.
16. Actions de Souveraineté Télécoms - Groupe de Travail : 5G en France, "Analyse, enjeux et recommandations pour la 5G en France", September 2015.
17. B. Astuto *et al.*, "A survey of software-defined networking: Past, present, and future of programmable networks," IEEE Communications Surveys & Tutorials, vol. 16, pp. 1617-1634, Third Quarter 2014.
18. G. Auer *et al.*, "How much energy is needed to run a wireless network?" IEEE Wireless Communications, vol. 18, pp. 40 - 49, no. 5, 2011.
19. A. Avizienis *et al.*, "Basic concepts and taxonomy of dependable and secure computing," IEEE Transactions on Dependable and Secure Computing, vol. 1, pp. 11-33, January-March 2004.
20. E. Bastug, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," EURASIP Journal on Wireless Communications and Networking, accepted, 2015.

21. E. Bastug, J. L. Guenego, and M. Debbah, “Proactive small cell networks,” in ICT 2013, Casablanca, Morocco, May 2013.
22. E. Bjornson *et al.*, “Multiobjective signal processing optimization: The way to balance conflicting metrics in 5G systems ,” IEEE Signal Processing Magazine, vol. 31, pp. 14 – 23, November 2014.
23. E. Bjornson, M. Kountouris, and M. Debbah, “Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination,” in Proc. Int. Conf. Telecommun. (ICT), Casablanca, Morocco, May 2013.
24. M. Bjorklund, “YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)”, RFC 6020, October 2010, <http://www.rfc-editor.org/info/rfc6020>.
25. P. Blasco and D. Gunduz, “Learning-based optimization of cache content in a small cell base station,” arXiv preprint: 1402.3247, 2014.
26. B. Boccardi, W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Communications Magazine, vol. 52, pp. 74–80, February 2014.
27. F. Bonomi, R. Milito, J. Zhu and S. Addepalli, “Fog computing and its role in the internet of things”. In Proceedings of MCC '12, workshop on Mobile cloud computing, pp. 13-16, 2012.
28. L. Carlà, R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, “LTE enhancements for Public Safety and Security communications to support Group Multimedia Communications”, ArchiveX preprint 1501.03613v1, January 2015.
29. B. Carlos, *et al.* “An architecture for software defined wireless networking.” Wireless Communications, IEEE 21.3 (2014): 52-61.
30. V. Chandrasekhar *et al.*, “Femtocell networks: A survey,” IEEE Commun. Mag., vol. 46, pp. 59-67, Sep. 2008.
31. T. Chen *et al.*, “Energy efficiency metrics for green wireless communications,” in Proc. WCSP’10.
32. Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019,” Cisco, San Jose, CA, White Paper, 3 February 2015.
33. C. Liang, and F. R. Yu. “Wireless network virtualization: A survey, some research issues and challenges.” Communications Surveys & Tutorials, IEEE 17.1 (2015): 358-380.
34. L. Cheng, *et al*. “The case for re-configurable backhaul in cloud-RAN based small cell networks.” INFOCOM, 2013 Proceedings IEEE. IEEE, 2013.
35. “Commission Recommendation of 25 July 2003 on the processing of caller location information in electronic communication networks for the purpose of location-enhanced emergency call services”, document number C(2003) 2657, July 2003.
36. D. C. Cox and H. Lee, “Physical relationships,” IEEE Microwave Magazine, vol. 9, no. 4, pp. 89 – 94, August 2008.
37. C. Wang, *et al*. “Cellular architecture and key technologies for 5G wireless communication networks.” Communications Magazine, IEEE 52.2: 122-130, 2014.
38. P. Demestichas *et al.*, “5G on the horizon: Key challenges for the radio-access network,” IEEE Vehicular Technology Magazine, vol. 8, pp. 47-53, no. 3, 2013.
39. L. Deng, *et al.*, “A unified energy efficiency and spectral efficiency tradeoff metric in wireless networks,“ IEEE Communications Letters, vol. 17, pp. 55–58, January 2013.
40. D. Sabella, *et al*. “RAN as a service: Challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network.” Future Network and Mobile Summit (Future Network Summit), 2013. IEEE, 2013.
41. Ericsson, “Ericsson Mobility Report,” Ericsson, Stockholm, Sweden, June 2015.
42. Ericsson, “Ericsson Energy and Carbon Report,” Ericsson, Stockholm, Sweden, June 2013.

43. Ericsson web site, http://www.ericsson.com/news/121101-ericsson-tests-lte-in-extreme-conditions_244159017_c
44. ETSI, “Mobile-Edge-Computing White Paper”, September 2014.
45. ETSI EN 300 392-2 V3.2.1 Terrestrial Trunked Radio (TETRA); Voice plus Data (V+D); Part 2: Air Interface (AI), 2007-09.
46. ETSI TR 102 580 V1.1.1 Terrestrial Trunked Radio (TETRA); Release 2; Designer’s Guide; TETRA High-Speed Data (HSD); TETRA Enhanced Data Service (TEDS) (2007-10).
47. ETSI, “Network functions virtualisation: An introduction, benefits, enablers, challenges & call for action,” white paper, October 2012, http://portal.etsi.org/NFV/NFV_White_Paper.pdf
48. ETSI TS 100 392-2 V3.6.1, Terrestrial Trunked Radio (TETRA); Voice plus Data (V+D); Part 2: Air Interface (AI), 2013-05.
49. G. Fettweis, “A 5G wireless communications vision,” Microwave Journal, 15-12-2012.
50. G. Fettweis and S. Alamouti, “5G: Personal mobile internet beyond what cellular did to telephony,” IEEE Communications Magazine, vol. 52, pp. 140 - 145, no. 2, 2014.
51. G. Frantz, “Digital signal processor trends,” IEEE Micro, vol. 20, pp. 52–59, no. 6, 2000.
52. G. Frantz, “Signal core: A short history of the digital signal processor,” IEEE Solid-State Circuits Magazine, vol. 4, pp. 16–20, Spring 2012.
53. Z. Gao *et al.*, “MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network,” IEEE Wireless Communications, vol. 22, pp. 1536-1284, October 2015.
54. D. Goodman and N. Mandayam, “Power control for wireless data,” IEEE Personal Communications, pp. 48-54, Apr. 2000.
55. A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” IEEE Access, vol. 3, pp. 1206-1232, 2015.
56. R. Q. Hu and Y. Qian, “An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems,” IEEE Communications Magazine, pp. 94-101, May 2014.
57. V. Hubka and W. E. Eder, Theory of Technical Systems: A Total Concept Theory for Engineering Design. Springer-Verlag, 1988.
58. IEEE Standards Dictionary, ieeexplore.ieee.org/xpls/dictionary.jsp.
59. N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, “Femtocaching: Wireless video content delivery through distributed caching helpers,” in IEEE INFOCOM, pp.1107–1115, 2012.
60. J. Hachem, N. Karamchandani, and S. Diggavi, “Coded caching for heterogeneous wireless networks with multi-level access,” arXiv preprint: 1404.6560, 2014.
61. F. Haider et al., “Spectral efficiency analysis of mobile Femtocell based cellular systems”, in Proc. IEEE ICCT, pp. 347-351, Jinan, China, Sep. 2011.
62. J. Halpern, and J. H. Salim, “Forwarding and Control Element Separation (ForCES) Forwarding Element Model”, RFC 5812, March 2010,<http://www.rfc-editor.org/info/rfc5812>.
63. H. Hawilo, *et al*. “NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC).” Network, IEEE 28.6 (2014): 18-26.
64. J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, “Making smart use of excess antennas: Massive MIMO, small cells, and TDD,” Bell Labs Technical Journal, vol. 18, pp. 5-21, Sep. 2013.
65. H. Zhang, *et al.* ”5G wireless network: MyNET and SONAC.” Network, IEEE 29.4 (2015): 14-23.
66. IEA, “More data, less energy making network standby more efficient in billions of connected devices,” International Energy Agency (IEA), Paris, 2014.

67. “iJOIN: Interworking and JOINt Design of an Open Access and Backhaul. Network Architecture for Small Cells based on Cloud Networks” [Online: http://www.ict-ijoin.eu]
68. International Technology Report for Semiconductors 2012 (ITRS 2012), www.itrs.net.
69. R. Irmer *et al.*, “Coordinated multipoint: Concepts, performance, and field trial results,” IEEE Communications Magazine, vol. 49, pp. 102-111, February 2011.
70. ITU-R, “M.2320 Future technology trends of terrestrial IMT systems”, November 2014, http://www.itu.int/pub/R-REP-M/en.
71. ITU-R,”M.2083 IMT Vision – ‘Framework and overall objectives of the future development of IMT for 2020 and beyond’,” July 2015, http://www.itu.int/rec/R-REC-M/en.
72. ITU-R, “M.2376 Technical feasibility of IMT in bands above 6 GHz,” July 2015, http://www.itu.int/rec/R-REC-M/en.
73. ITU-R, WP 5D documents, http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx.
74. A. Jacobson *et al.*, The Unified Software Development Process. Addison-Wesley, 1999.
75. M. Ji, G. Caire, and A. F. Molisch, “Wireless device-to-device caching networks: Basic principles and system performance,” arXiv preprint: 1305.5216, 2013.
76. P. Kämppi, J. Rajamäki, S. Tiainen and R. Leppänen (eds.), “MACICO: Multi-Agency Cooperation In Cross-border Operations Samples of Evidence Series: Volume 4”, Laurea Publications, Espoo, Finland, 2014.
77. R. W. Keyes, “Physical limits of silicon transistors and circuits,” Reports on Progress in Physics, vol. 68, pp. 2701–2746, December 2005.
78. M. Kimmerlin, J. Costa-Requena, J. Manner, D. Saucez, Y. Sanchez., “Caching Using Software-Defined Networking in LTE Networks,” Extension paper. 2015. <hal-01117447v2>.
79. Y. Kishiyama *et al.*, “Future steps of LTE-A: Evolution towards integration of local area and wide area systems,” IEEE Wireless Communications, vol. 20, pp. 12–18, February 2013.
80. S. Kline, “Conceptual Foundations for Multidisciplinary Thinking,” Stanford University Press, 1995.
81. R. Landauer, “Irreversibility and heat generation in the computing process,” IBM Journal of Research and Development, vol. 5, pp. 183–191, July 1961.
82. M. Lauridsen *et al*., “An empirical LTE smartphone power model with a view to energy efficiency evolution,” Intel Technology Journal, vol. 18, pp. 172-193, Issue 1, 2014.
83. A. Laya *et al.*, “Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives,” IEEE Commun. Surveys Tutorials, vol. 16, no. 1, pp. 4-16, Dec. 2014.
84. D. Lecompte, F. Gabin, “Evolved Multimedia Broadcast/Multicast Service (eMBMS) in LTE-Advanced: Overview and Rel-11 Enhancements”, IEEE Communications Magazine, pp. 68-74, November 2012.
85. P. Lettieri and M.B. Srivastava, “Advances in wireless terminals,” IEEE Personal Communications, vol. 6, pp. 6 – 19, no. 1, 1999.
86. L. Z. Li, Z. Mao, and J. Rexford. “Toward software-defined cellular networks.” IEEE European Workshop on Software Defined Networking (EWSDN), 2012.
87. J. Liu *et al*., “CONCERT: A cloud-based architecture for next-generation cellular systems,” IEEE Wireless Communications, vol. 21, pp. 14–22, December 2014.
88. W. Liu *et al.*, “Energy efficiency comparison of massive MIMO and small cell network,” in Proc. GlobalSIP’14.
89. L. Lu *et al.*, “A overview of massive MIMO: Benefits and challenges,” IEEE Journal of Selected Topics in Signal Processing, vol. 8, pp. 742-758, October 2014.

90. M. Maddah-Ali and U. Niesen, “Fundamental limits of caching,” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856–2867, May 2014.
91. P. Marsch *et al.*, “Preliminary view 5G RAN architecture and functional design,” METIS-II white paper (draft), January 2015.
92. R. T. Marler and J. S. Arora, “Survey of multi-objective optimization methods for engineering,” Struct Multidisc Optim, vol. 26, pp. 369-395, 2004.
93. T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Transactions on Wireless Communications, vol. 9, pp. 3590-3600, November 2010.
94. I. L. Markov, “Limits on fundamental limits to computation,” Nature, vol. 512, pp. 147–154, 14 August 2014.
95. C. Mead and L. Conway, Introduction to VLSI Systems. Reading, MA: Addison-Wesley, 1980.
96. ICT-31766 METIS project, “Scenarios, requirements and KPIs for 5G mobile and wireless system,” Deliverable D1.1, ICT-317669-METIS/D1.1, April 2013.
97. ICT-31766 METIS project, “Summary on preliminary trade-off investigations and first set of potential network-level solutions”, Deliverable D4.1, September 2013.
98. M. Moradi, *et al*. “SoftMoW: Recursive and reconfigurable cellular WAN architecture.” Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies. ACM, 2014.
99. G. E. Moore, “Cramming more components onto integrated circuits,” Electronics, vol. 38, pp. 114–117, April 1965.
100. G. E. Moore, “Progress in digital integrated electronics,” in Proc. International Electron Devices Meeting, pp. 11–13, 1975.
101. P. Munoz *et al.*, “Mobility-based strategies for traffic steering in heterogeneous networks,” IEEE Communications Networks, pp. 54-62, May 2013.
102. M. Yang, *et al*. “OpenRAN: a software-defined ran architecture via virtualization.” ACM SIGCOMM computer communication review. Vol. 43. No. 4. ACM, 2013.
103. Networld2020, “5G: Challenges, research priorities, and recommendations,” Joint White Paper, Networld2020, September 2014, http://networld2020.eu/sria-and-whitepapers/.
104. Networld2020, “White paper for research beyond 5G,” Draft White Paper, v. 1.0, Networld2020, October 2015, http://networld2020.eu/sria-and-whitepapers/.
105. NGMN, 5G White Paper, NGMN Alliance, 17 February 2015, https://www.ngmn.org/5g-white-paper.html.
106. NIST, “Framework and Roadmap for Smart Grid Interoperability Standards Release 1.0 (Draft)”, 2009.
107. Nokia, “5G radio access: System design aspects,” Nokia Networks, Espoo, Finland, White Paper, 2015.
108. Open Networking Foundation, “The OpenFlow Switch Specification, Version 1.4.0”, October 2013, <https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.4.0.pdf>.
109. A. Osseiran *et al.*, “Scenarios for 5G Mobile and Wireless Communications: the Vision of the METIS Project,” IEEE Commun. Mag., vol. 52, no. 5, pp. 26-35, May 2014.
110. J. Paakkonen, C. Hollanti, and O. Tirkkonen, “Device-to-device data storage for mobile cellular systems,” in IEEE Globecom Workshops (GC Wrokshops), pp. 671–676, December 2013.
111. R. Pabst *et al.*, “Relay-based deployment concepts for wireless and mobile broadband radio,” IEEE Communications Magazine, vol. 42, pp. 80 - 89, September 2004.

112. P. Agyapong, *et al*. “Design considerations for a 5G network architecture.” Communications Magazine, IEEE 52.11 (2014): 65-75.
113. G. Pallis and A. Vakali, “Insight and perspectives for content delivery networks,” Communications of the ACM, vol. 49, pp. 101-106, January 2006.
114. A. Polydoros, N. Dimitrou, G. Baldini *et al.*, “Public protection and disaster relief communication system integrity”, IEEE Vehicular Technology Magazine, pp 1-60, December 2014.
115. C. Prehofer and C. Bettstetter, “Self-organization in communication networks: Principles and design paradigms,” IEEE Communications Magazine, vol. 43, pp. 78 – 85, no. 7, 2005.
116. R. Presuhn, “Management Information Base (MIB) for theSimple Network Management Protocol (SNMP)”, STD 62, RFC 3418, December 2002, <http://www.rfc-editor.org/info/rfc3418>.
117. P. Rost, *et al.*, “Cloud technologies for flexible 5G radio access networks.” Communications Magazine, IEEE 52.5 (2014): 68-76.
118. J. M. Rabaey *et al.*, “PicoRadio supports ad hoc ultra-low power wireless networking,” Computer, vol. 33, pp. 42 – 48, no. 7, 2000.
119. Random House Webster’s Concise College Dictionary. Random House, 1999.
120. R. Riggio, I. G. B. Yahia, S. Latre, and T. Rasheed, “Scylla: A Language for Virtual Network Functions Orchestration in Enterprise WLANs.” In Proc. of IEEE NOMS, Instanbulm Turkey, 2016.
121. R. Riggio, M. K. Marina, J. Schulz-Zander, S. Kuklinski, T. Rasheed, “Programming Abstractions for Software-Defined Wireless Networks.” IEEE Transactions on Network and Service Management, vol.12, no.2, pp.146-162, June 2015.
122. P. Rost *et al.*, “Cloud technologies for flexible 5G radio access networks,” IEEE Communications Magazine, vol. 52, pp. 68-76, May 2014.
123. Safety Forum, “Recommendation of the DG eCall, for the introduction of the pan-European eCall”, 2006.
124. M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, “The Case for VM-Based Cloudlets in Mobile Computing.” IEEE Pervasive Computing, vol. 8, no. 4, pp. 14–23, 2009.
125. V. A. Siris, X. Vasilakos, and G. C. Polyzos, “Efficient proactive caching for supporting seamless mobility,” arXiv preprint: 1404.4754, 2014.
126. M. I. Skolnik, “Introduction to Radar Systems”, 3rd ed. New York: McGraw-Hill, 2001.
127. Y. Sui, J. Vihrälä, A. Papadogiannis, M. Sternad, W. Yang, and T. Svensson, “Moving cells: a promising solution to boost performance for vehicular users”. IEEE Commun. Mag. 51, pp 62–68, June 2013.
128. Y. Sui, I. Guvenc, and T. Svensson, “Interference management for moving networks in ultra-dense urban scenarios”, EURASIP Journal on Wireless Communications and Networking, April 2015.
129. T. Taleb *et al.*, “Lightweight Mobile Core Networks for Machine Type Communications,” IEEE Access, vol. 2, pp. 1128-1137, Sep. 2014.
130. T. Taleb and A. Kunz, “Machine Type Communications in 3GPP Networks: Potential, Challenges, and Solutions,” IEEE Commun. Mag., vol. 50, no. 3, pp. 178- 184, March 2012.
131. T. Taleb and A. Ksentini, “On Alleviating MTC Overload in EPS,” Ad Hoc Networks, vol. 18, pp. 24-39, July 2014.
132. “TROPIC: Distributed computing, storage and radio resource allocation over cooperative femtocells” [Online: http://www.ict-tropic.eu]
133. B. Wang *et al.*, “End-to-end power estimation for heterogeneous cellular LTE SoCs in early design phases,” in Proc. PATMOS’14.

134. C. X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks,"' IEEE Commun. Mag., vol. 52, no. 2, pp. 122_130, Feb. 2014.
135. W. Wang and G. Shen, "Energy efficiency of heterogeneous cellular network," in Proc. VTC'10 Fall.
136. M. S. Whittingham, "History, evolution, and future status of energy storage," Proceedings of the IEEE, vol. 100, pp. 1518-1534, Centennial Special Issue, 2012.
137. S. Woo, E. Jeong, S. Park, J. Lee, S. Ihm, and K. Park, "Comparison of caching strategies in moders cellular backhaul networks," Proceeding of the 11th annual international conference on Mobile systems, applications, and services (MobiSys '13), p: 319-332.
138. C. Xavier, *et al.* "Radio access network virtualization for future mobile carrier networks." Communications Magazine, IEEE 51.7 (2013): 27-35.
139. T. Zachariah, N. Klugman, B. Campbell, J. Adkins, N. Jackson, and P. Dutta. "The Internet of Things Has a Gateway Problem". In ACM HotMobile'15, pp. 27–32, 2015.
140. X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," IEEE Transactions on Communications, vol. 63, pp. 1881-1894, May 2015.
141. K. Zheng *et al.*, "Challenges of Massive Access in Highly Dense LTE-Advanced Networks with Machine-to-Machine Communications," IEEE Wireless Commun., pp. 12-18, June 2014.
142. V. V. Zhirnov and R. K. Cavin, "Future microsystems for information processing: Limits and lessons from living systems," IEEE Journal of the Electron Device Society, vol. 1, pp. 29-47, February 2013.
143. K. Zhou and N. Nikaein, "Packet Aggregation for Machine Type Communications in LTE with Random Access Channel," Proc. IEEE WCNC, pp. 262-267, April 2013.

A. Annex: Use cases classification in 3GPP TR

In order to select and adapt those existing use cases and requirements which are relevant to COHERENT approach, we have analysed the use cases and especially the requirements available at this time in 3GPP. It should be noted that, on the other side, we have contributed in 3GPP SA1 for influencing some of the 5G requirements.

3GPP SA1, as part of its 5G studies, is developing now under the SMARTER Study Item the Use Cases and the Building Blocks regarding service use cases. At this time the draft 3GPP TR 22.891 [1] is available, which already contains more than 70 use cases. Part of these use cases were adopted based on the existing documents, such as [105], while other use cases are direct contributions of 3GPP members, including those from COHERENT. The contributions to 3GPP TR 22.891 [1] continue at the time of writing this deliverable.

At the SA1 meeting in October 2015 a Work Item was generated for the consolidation of the Use Cases in separate standards named Building Blocks, having as scope the derivation of requirements. We perceive that the use case classification can be efficiently used for understanding the SA1 directions and we will use as reference the contribution S1-153171 [2], prepared by the SMARTER Rapporteur. It should be noted that there are a number of service use cases grouped based on the represented vertical markets and a horizontal group (NEO), applicable to several vertical markets. The main vertical markets overlap with vertical services defined in ITU-R, 5GPP, etc.

SA1 has agreed in October 2015 with the classification below. Some use cases appear in different groups. We include here a short summary and the associated requirements, as included in draft TR 22.891 [1]:

A.1 Enhanced Mobile Broadband (eMBB)

A.1.1 Higher capacity

Mobile broadband for indoor scenario
Potential system requirements:

- Support of user experienced data rate up to Gbps of level.
- User peak data rate at tens of Gbps.
- Whole traffic volume in the area at least the level of Tbps/ km^2.
- Support of very low latency for user experienced data exchange.

Mobile broadband for hotspots scenario
Potential system requirements:

- Support of the user experienced data rate up to Gbps of level while the user is moving slowly.
- Support of the peak data rate at tens of Gbps while the user is moving slowly.
- Support of the whole traffic volume in the area at least the level of Tbps/ km^2.
- Support of flexible and efficient backhaul especially outdoor.

On-demand networking (duplicate with NEO)
Examples: networking only during a football match; different traffic at traffic lights, while waiting, and on the move.

Potential system requirements:

- Provide guaranteed user experience for mobile broadband services like live video in areas with a high UE density which requires user experienced data downlink rate of 300Mbps and uplink rate of 50Mbps in 200-2500 /km^2 connection density.
- Provide consistent users experience when terminals enter the areas with a high UE density which requires user experienced data downlink rate of 50Mbps and uplink rate of 25Mbps in 2000/km^2 connection density.

- Adjust the network capacities dynamically based on the variation of demand and performance indicators.

Mobile broadband services with seamless wide-area coverage (duplicate with eV2X)
Examples: urban areas, rural areas, high-speed railways and fast ways between cities.

Potential system requirements:
For wide area coverage, support of user experienced data rate for mobile broadband services anytime and anywhere, e.g., 100Mbps. The above requirement assumes reuse of an existing base station site grid.

- Support fast-moving end-users, e.g., 500km/h.
- Support high connection density for high speed scenarios, e.g., 500 active Ues simultaneously.
- Support low latency for high speed scenario.

Virtual presence (duplicate in CriC)
The goal is to provide interactive services for high data rate zones (e.g. Office environments) as described in section 3.2.1 of the NGMN [105].

Example: Phil works in a multinational company which has offices in many big cities. He has regular meetings with colleagues based in other countries. He uses to have real time 360° video communications: he wears Virtual Presence glasses, allowing to be merged in a meeting room where he can see all his other colleagues sitting around a table. He can interact with them in real time as if they were just in front of him.

Potential system requirements:

- Provide high bandwidth (bidirectional) and low latency.

Improvement of network capabilities for vehicular case (duplicate with eV2X)
Example: Video/gaming services to vehicular (up to 60 km/h) users (cars or buses) in city centre, driving on a six lanes in a street canyon surrounded by at least six-story buildings. The mobile network deployment can be either macro or micro cellular, or ultra dense network (UDN). 1/3 of the active users are engaged in real-time video (4K), 1/3 in non-real-time (8K video) and 1/3 in gaming (which includes a 3D 4K video). Each car: 1 active user (range 0.5-2). Density of cars per lane: 5 cars per 100 m lane on four lanes (car lanes). For simplicity, all cars are assumed to move at constant 60 km/h. Each bus: 20 active users (range 10-30). Density of buses per lane: 3 buses per 100 m of lane on two lanes (bus lanes). For simplicity, all buses are assumed to move at constant 60 km/h.

Potential system requirements:
In the table below are examples of required user throughputs and latencies.

Table A-1 Examples of required user throughputs and latencies

Application	Average End User Throughput	Latency (end-to-end)	Latency (over the air)
High Definition Video 8K (streaming)	< 100 Mbps (DL)	< 1 s	< 200 ms
High Definition Video (conversational)	< 10 Mbps (DL/UL)	< 150 ms	< 30 ms
Cloud Computer Games with 4K 3D graphics – Low Latency Applications	< 50 Mbps (DL/UL) (UL is needed for multiplayer game	< 7.5 ms (10 times less than in for real time games)	< 1.5 ms

computation in
user device)

Vehicular internet and infotainment
Potential system requirements:

- Provide a consistent data rate that is high enough to support the chosen media:
- For internet browsing and general information at least 0.5 Mb/sec.
- For high quality music streaming at least 1 Mb/sec.
- For standard quality video streaming at least 5 Mb/sec.
- For high quality (up to UHD) video streaming at least 15 Mb/sec.
- A latency of no more than 100ms for internet browsing shall be provided.
- Deliver the required connection quality up to 200 km/hr.
- Provide the required connection quality in densely populated roads where up to 2000 vehicles in a given service area 1km^2 will be accessing data. The vehicles could be moving at speeds ranging from 0 km/h (e.g. in a traffic jam) to 200 km/h.

Broadcasting support (duplicate with NEO)
Potential system requirements:

- Support an interface from external Broadcasters' management systems.
- Reserve groups of resources for Broadcast channels.
- Support broadcast of lossless state of the art video streams such as 4k UHD.
- Elect to receive a reduced quality version of the broadcast for display on their device screen (typically less than 12") or a full quality version of the same channel for presentation to a video presentation monitor (typically much larger than their device, ie:32" to 72"screen size).
- Allow the UE to receive broadcasts selected by the user (from the Broadcaster's management system) in accordance with any appropriate authorisations by the Broadcaster (UHD: 3840 x 2160, 50FPS, AVC ~300Mbit/s, UHD: 3840 x 2160, 50FPS, HEVC ~ 150Mbit/s).

A.1.2 Enhanced Connectivity

Connectivity everywhere
Example**:** Commercial, emergency or recreational UAVs (Unmanned Aeronautical Vehicle) will be controlled by various control centers such as local/federal agency or owner of the UAVs.

Potential system requirements:

- Provide aerial object and/or nautical objects with reliable mobile broadband connectivity.
- Provide reliable low-latency connectivity between aerial objects.

A.1.3 Higher user mobility

Higher user mobility
Potential system requirements:

- Support enhanced mobile broadband services in fast moving vehicles (e.g. up to 500 km/h) with enhanced user experience.
- Support enhanced connectivity services in fast moving airplanes (e.g. up to 1000 km/h) with enhanced user experience.

Vehicular internet & infotainment
(see description within eMBB)

A.2 Critical communications (CriC)

A.2.1 Higher reliability and lower latency

Ultra reliable communication

Examples: Industrial control systems (from sensor to actuator, very low latency for some applications), mobile health care, remote monitoring, diagnosis and treatment (high rates and availability), real time control of vehicles, road traffic, accident prevention (location, vector, context, low Round Trip Time RTT), wide area monitoring and control systems for smart grids, communication of a critical information with preferential handling for public safety scenarios.

Potential system requirements:

- Services in this category require very low data error rate. Some of them also require very low latency, i.e. for industrial automation with delays of one ms.

Virtual presence
See eMBB.

Remote control
Example: In the future, UAV (unmanned aerial vehicle) will be widely used for delivery of packages (e.g. A company plans to use UAVs to deliver goods), which improves delivery efficiency. In a situation where first aid personnel cannot arrive promptly to the scene of an emergency, UAVs could be used to collect video information on site and deliver emergency equipment.

Potential system requirements:

- Round trip latency less than 150 ms, including all network components.
- High reliability for fast-moving end-users (e.g. 120km/h); reliability goal is [near 100%]
- Seamless connection for fast-moving end-users.
- Reliability to be at the same level for current aviation Air Traffic Control (ATC). Link supports command and control of vehicles in controlled airspace.
- Priority, Precedence, Preemption (PPP) mechanisms shall be used to ensure sufficient reliability metrics are reached.
- Position accuracy within 10 cm to avoid damage to property or life in densely populated areas.

Cloud Robotics
Example: Robotics considers a new paradigm where robots and automation systems exchange data and perform computation via networks. Cloud robotics would allow robots to offload compute-intensive tasks like image processing and voice recognition and even download new skills instantly.

Potential system requirements:

- Support Ues of high density distribution to upload synchronized audio, video and data in real time.
- Support end to end latency lower than 10ms.

Industrial Factory Automation
Example: In these applications, a controller interacts with large number of sensors and actuators (up to 300), typically confined to a rather small manufacturing unit (e.g. 10m x 10m x 3m). The resulting S/A density is often very high (up to $1/m^3$). Many of such manufacturing units may have to be supported within close proximity within a factory (e.g. up to 100 in assembly line production, car industry).

Potential system requirements:

- Support cycle times of 1ms to 2ms.
- Transaction jitter should be below 10usecs.
- Reliability, measured as the fractions of transactions that cannot meet the latency or jitter constraints, should remain below 10-9.
- Each transaction should support a payload of 50 to 100 Bytes
- For factory automation, required range is up to 10-20m.

Industrial Process Automation

Example: The use case requires support of a large number of sensor devices (10k) per plant as well as highly reliable transport (packet loss rate <10^{-5}). Further, power consumption is critical since most sensor devices are battery-powered with a targeted battery lifetimes of several years while providing measurement updates every few seconds. Also, range becomes a critical factor due to the low transmit power levels of the sensors, the large size of the plant and the high reliability requirements on transport. Latency requirements typically range between 100ms and 1s. Data rates can be rather low since each transaction typically comprises less than 100B.

Potential system requirements:

- Support 10k sensor nodes within an area of 10sqkm.
- Reliability for the transport of transactions, measured as the fractions of packet losses, should remain below 10^{-5}.
- Transaction latency of 50-100ms, defined as the overall cycle time between a sensor reading and action from process controller.
- Allow a battery powered sensor lifetime of multiple years while enabling a transaction rate of one every few seconds
- Transactions should be sufficiently integrity- and confidentiality-protected.

Low-delay speech coding

Example: When voice is used in a highly interactive environment, e.g., a multiplayer game or a virtual reality meeting, the requirements on the speech coding delay become tougher to meet, and current coding delays are too high. To support interactivity, the one-way delay for speech should be 10 ms (or lower).

Potential system requirements:

- Support speech with very low one-way service latency 10 ms.

Local UAV Collaboration

Example: Unmanned aerial vehicles (UAVs) local vehicle collaboration can act as a mobile sensor network to autonomously execute sensing tasks in uncertain and dynamic environments while being controlled by a single user. Accuracy in sensing tasks is increased when deploying a team of UAVs versus just one as there are multiple vantage points using multiple sensors. Examples of uses for deploying a team of UAVs include:

- Searching for an intruder or suspect
- Continual monitoring of natural disasters
- Performing autonomous mapping
- Collaborative manipulation of an object (e.g. picking up corners of a net or picking up a log).

Potential system requirements:

- Latency of 10 ms as collaboration requires vehicle altitude and position control loops to synchronize. Latency is required on the order of the control loop bandwidths.
- Near 100% reliability as instability and crashing of UAV could result from loss of communications. Control functions depend on this communication.
- Security to be provided at the level for current aviation Air Traffic Control (ATC) for command and control of vehicles in controlled airspace.
- Priority, Precedence, Preemption (PPP) needed as failure to transmit communications in reliable and timely manner could result in loss of property or life.
- Position accuracy within 10 cm due to multiple UAVs that may need to collaborate in close proximity to one another.

A.2.2 Higher reliability, higher availability and lower latency

Connectivity for drones

Example: A drone and remote control are connected to the mobile network. The drone is piloted with remote mode, data being transmitted via the network. The drone transmits video and with other data, such as infrared pictures.

Potential system requirements:

- Round trip latency less than 150 ms, including all network components.
- Due to consequences of failure being loss of property or life, reliability goal is near 100%.
- Reliability to be at the same level for current aviation Air Traffic Control (ATC). Link supports command and control of vehicles in controlled airspace.
- Priority, Precedence, Preemption (PPP) mechanisms shall be used to ensure sufficient reliability metrics are reached.
- Position accuracy within 10 cm to avoid damage to property or life in densely populated areas.
- Provide continuous wireless coverage, high speed uplink bandwidth at least 20Mbps, for a flying UE at low altitude of 10-1000 meters with the high speed as maximum as 300km/h.

Industrial control
Potential system requirements:

- Support very low latency (~1 ms).
- Support very high reliability.
- Support very high availability.
- Support high uplink data rate (tens of Mbps per device in a dense environment).

A.2.3 Very low latency

Tactile Internet
Example: Extremely low latency in combination with high availability, reliability and security will define the character of the Tactile Internet", makes the cellular network an extension of our human sensory and neural system.

Potential system requirements:

- Support very low latency (~1 ms).
- Support very high reliability.
- Support connections that are very difficult to block, modify, or hijack.

Localized real-time control
Potential system requirements:

- Support extremely high reliability and extremely low latency (1-10 ms) for data transmission.

Extreme real-time communications and the tactile Internet
Examples**:**

- Truly immersive, proximal cloud driven virtual reality.
- Remote control of vehicles and robots, real-time control of flying/driving things.
- Remote health care, monitoring, diagnosis, treatment, surgery.

Target 1ms delay implies endpoints must be physically close. Maximum distance between endpoints depends on delay budget per link.

Potential system requirements:

- Support 1ms one-way delay between mobile devices and devices in the nearby internet.

A.2.4 Higher accuracy positioning

High Accuracy Enhanced Positioning (ePositioning) (duplicate with NEO)
Potential system requirements:

- Support higher accuracy location capability less than 3 m at 80% of occasions.

A.3 Massive Internet of Things (MIoT)

A.3.1 Internet of Things

Light weight device configuration
Example: On the lower end of the device function range, devices may not need to be equipped with an IMS client, and yet it would still be desirable to activate such a device remotely. A light weight configuration mechanism may be used to provide the configuration information to the device.

Potential system requirements:

- Support devices (e.g., smart meter) with limited communication requirements and capabilities (e.g., devices without an IMS client).
- Support a lighter weight signalling for device configuration (i.e., service parameters) than is currently available in EPS.

IoT Device Initialization
Potential system requirements:

- Not relevant to lower layers.

Subscription security credentials update
Potential system requirements:

- Not relevant to lower layers.

Devices with variable data
Potential system requirements:

- Enhance efficiency and flexibility for both low throughput short data bursts and high throughput data transmissions (e.g., streaming video) from the same device.
- Support efficient signaling mechanisms (e.g., signaling is less than payload).
- Reduce signaling overhead for security needed for short data burst transmission, without reducing the security protection provided by 4G 3GPP Systems.

Domestic Home Monitoring
Potential system requirements:

- Enable mobiles that can operate as static concentrators of IoT capillary traffic towards 3GPP networks.
- Support the integration of mobile, IoT capillary concentrator systems into Home Base stations.

Massive Internet of Things M2M and device identification
Examples:

- A building climate control system. There is a climate control server that communicates with all kinds of sensors/actuators (temperature, humidity, valves, et cetera) in the building. The climate control server may also communicate with sensors/actuators used by other systems (e.g. door sensors can be used for the security system and for climate control), it may use external sensors (e.g. local weather sensors), and it will communicate with external devices for notifications and remote control (e.g. with the building manager's phone).
- Another relevant example is Wearable Devices. NGMN [105] mentions: "Fitness-related applications, such as activity and body monitoring applications that track walking, running, and biking activities, metabolic rate, cardiovascular fitness, sleep quality, etc. will constitute a significant vertical market in M2M services. Some of these applications will utilize body or personal area networks to collect biometric information and then use cellular networks to transmit it back to centralized data acquisition sites".

Within the Internet-of-Things there will be very high densities of connections. NGMN mentions an active connection density of 200,000 / km^2. [7] mentions a device density of 1 M / km^2.

Potential system requirements:
Support network servers/applications and devices to identify, address and reach other devices, in a consistent manner independently of how these devices are connected.

A.3.2 Smart wearables

Bio-connectivity
Examples: As per the 4G Americas white paper: “Bio-connectivity, which is the continuous and automatic medical telemetry (e.g., temperature, blood pressure, heart-rate, blood glucose) collection via wearable sensors.

Potential system requirements:

- Support a mechanism that provides security, authentication and authorization for Ues which only support 3GPP device-to-device communication.

A.3.3 Sensor networks

Wide area monitoring and event driven alarms
Potential system requirements:

- Support efficient transfer of infrequent uplink data for low power devices which only participate in mobile-originated communication scenarios.
- Support resource efficient mechanism to provide service parameters and activate groups of low power devices.
- Support significantly increased device power efficiency (e.g., battery life up to more than 10 years).
- Support efficient data transmission with limited resource and signalling usage.
- Support high density massive connections (e.g.1 million connections per square kilometre) in an efficient manner.
- Support significant coverage enhancement (e.g., 20dB better coverage than Rel 99 GPRS system).

Low mobility devices (duplicate with NEO)
Potential system requirements:

- Support an efficient mechanism to accept information from large numbers of stationary devices with reduced mobility management (e.g., handover support, idle mode mobility management) and a source efficient mechanism to provide information to a stationary device.

Materials and inventory management and location tracking
Potential system requirements:

- Support a resource efficient mechanism to accept information from large numbers of locally dense devices, possibly simultaneously.
- Support mechanisms to enable sufficient indoor and outdoor coverage (e.g., 20dB better coverage than legacy Rel 99 GPRS system) for a large number of locally dense low power devices.
- Support a mechanism to manage resource (e.g., radio resources) sharing by large numbers of locally dense devices efficiently.
- Support communication service for high density of devices up to (e.g., 1 million devices per km^2), with high mobility at minimum of 100 km/h and with reduced battery consumption.
- Support high positioning accuracy in both outdoor and indoor scenarios (e.g., 0.5m).

A.4 Network operation (NEO)

A.4.1 System flexibility

Network slicing
Examples:

- Self-automated car in a smart city: Bob starts his self-automated driving car that relies on V2X communication. While sitting in the car, Bob initiates a HD video streaming service through the infotainment system available in the car. In this example, the V2X communication requires a low-latency but not necessarily a high throughput, whereas, the HD video streaming requires a high throughput but is tolerate to the latency. Both services are assumed to be provided by the same operator.
- Healthcare robot: A robot that is monitored by the healthcare service provider takes care of elderly people at home. The robot sends a regular report of health status and the activities interacting between the robot and the elderly people to the healthcare operator. The robot also allows the elderly people to do any Internet like services (e.g., web-surfing, hearing streaming music, watching a video) or even making a call to their doctor directly in case of emergency. Both services are assumed to be provided by the same operator.

Potential system requirements:

- Allow operator to compose network slices, i.e. independent sets of network functions (e.g. potentially from different vendors) and parameter configurations, e.g. for hosting multiple enterprises or MVNOs etc.
- Identify certain terminals and subscribers to be associated with a particular network slice.
- Enable a UE to obtain service from a specific network slice e.g. based on subscription or terminal type.
- Operate different network slices in parallel with isolation that e.g. prevents data communication in one slice to negatively impact services in other slices.
- Conform to service-specific security assurance requirements in a single network slice, rather than the whole network.
- Support elasticity of network slice in term of capacity with no impact on the services of this slice or other slices.
- Allow the operator to authorize third parties to create, manage a network slice configuration (e.g. scale slices) via suitable APIs, within the limits set by the network operator.
- Allow changing the slices with minimal impact on the ongoing subscriber's services served by other slices, i.e. new network slice addition, removal of existing network slice, or update of network slice functions or configuration.
- Support E2E (e.g. RAN, CN) resource management for a network slice.

Lifeline communications / natural disaster

Potential system requirements:

- Based on operator's policy, the system shall be able to define minimal services necessary in case of disaster that are conditional on e.g. subscriber class (i.e. access class), communication class (i.e. emergency call or not), device type (i.e. Smart phone or IoT device), and application. Examples of those minimal services are communications from specific high priority users, emergency calls, and a disaster-message-board type of application that helps people reconnect with friends and loved ones in the aftermath of disasters.
- Those minimal services shall be available in case of disaster.
- During the recovery phase of disaster, the service continuity of those minimal services that start being provided should be ensured.

Flexible application traffic routing

Example: Service Provider X provides 3D Augmented Reality (AR) service for the users, one user can interact with other user via live 3D Augmented Reality service. Service Provider X has a service agreement with an MNO, and the MNO network may optimize the traffic transfer for the live 3D Augmented Reality service in order to realize good user experience.

Potential system requirements:

- Subject to operator's policy and/or based on application needs, the network shall support efficient user-plane paths between Ues attached to the same network, even if the Ues change their location during communication.

- Subject to operator's policy and/or based on application needs, the network shall support efficient user-plane paths between a UE attached to the mobile network and communication peers outside of the mobile network (e.g. Internet hosts).

Routing path optimization when server changes
Potential system requirements:

- Subject to the service agreement between the operator and the service provider, the network shall enable hosting of services (including both MNO provided services and 3rd party provided services) closer to the end user to improve user experience and save backhaul resources.
- Support routing of data traffic to the entity hosting services closer to the end user for specific services of a UE.
- Enable efficient user-plane paths between a UE and the entity hosting the service closer to the end user even if the UE changes its location during communication.
- Enable charging, QoS, and Lawful Interception (LI) for services hosted closer to the end user.

Provision of essential services for very low-ARPU areas
Potential system requirements:

- Network sharing with capabilities for operators to set parameters for resource sharing both on demand and dynamically.
- Low or no mobility Ues (e.g. up to 50 km/h).
- Efficient use of the control plane (e.g., cooperation between services to minimize overall signalling between a UE and the network).
- Efficient use of the data plane (e.g., packaging data from multiple applications and sending it on a periodic basis rather than an on demand basis).
- APIs that provide network status information to applications (e.g., to allow applications to use network resources efficiently).
- Minimise as much as possible the traffic (Data and signalling) on the interfaces between the access network and the core network in order to reduce the amount of backhaul traffic.
- UEs with minimal functionality (e.g. user experienced data rate of 10 Mbps at DL and 10 Mbps at UL with E2E latency of 50 ms).
- very large cells (e.g.: link budget better than 160 dB, relaxed timing on random access and other procedures to enable very long range beyond 50km).

Network capability exposure
Potential system requirements:

- To provide the network capabilities to the 3rd party ISP/ICP.
- Service providers can be capable of configuring and managing the service via e.g. open API, while operators will have the option to manage and evolve the network.

Broadcasting Support (duplicate with eMBB)
See eMBB.

Ad-Hoc Broadcasting
Example: Setup event based video content broadcasting, using a slice of the local or temporary 3GPP system in the environ of the event.

Potential system requirements:

- Support an interface from an external Ad-Hoc Broadcast management system that manages broadcast requests.
- Reserve groups of resources temporarily for scheduled Ad-Hoc Broadcasts.
- Allow the UE to receive broadcasts selected by the user (from the Ad-Hoc Broadcast management system) in accordance with any appropriate authorisations.

A.4.2 Scalability

Flexibility and scalability
Example: Since traffic varies depending on the time of the day and on the day of the week, network deployment decisions based on peak traffic cause waste of resources. In addition, traffic varies also depending on location. It is understood that traffic moves from a location to another in a way, while the total amount of traffic in a wider area is less changed. Therefore it is important that the system can flexibly scale with various levels of control and user-plane demand in order to avoid localized underutilization of resources [105]. Resiliency against congestion and disasters would be also much enhanced by that.

Potential system requirements:

- Using resources (compute, network and storage resources) in more than one geographic area by the system shall be supported without requiring manual re-configuration of neighbouring nodes, without service disruption, and while avoiding additional signalling due to unnecessary UE's re-attachments (e.g. due to loss of call state information in the network).

Context Awareness to support network elasticity
Example: The exact location information of UE can be used by network node to optimally select the cell that the UE should be connected to. If network nodes can utilize the speed and heading information of UE, which can be computed by using accelerometer or gyroscope, the network node can optimally configure which cells to monitor for handover or when to perform handover.

Potential system requirements:

- Enable elastic configuration of the network based on system information, including:
    - Instantaneous network conditions, such as serving RATs (macro cell, small cell, WiFi), network load information and congestion levels;
    - Application's user characteristics, such as mobility type (high mobility, low mobility, no mobility), expected traffic over time, location)
    - When allowed by a user, UE context information, such as sensor-level information (e.g. direction, speed, power status, display status, other sensor information installed in the UE), application-level information (e.g. foreground applications, running background application, application data, user settings, etc.)

Network enhancements to support scalability and automation
Potential system requirements:

- Guarantee the service experience of the subscribers during the network scaling and automation operation.
- Moreover, the existing mechanisms (e.g. load balancing, network function selection) which are closely related with the effect of the network scalability and automation operation need to be enhanced.

A.4.3 Mobility support

Mobility on demand
Potential system requirements:

- Enable operators to define different levels of mobility support for different UEs.
- Mobility support consists of providing none, any one or any combination of the following:
    - Minimizing packet loss during inter- and/or intra-RAT cell changes,
    - Maintaining the same IP address assigned to a UE across different cells,
    - Minimizing interruption time until a UE can continue to communicate with a potentially different IP address (in case the same IP address is not maintained during a mobility event).
    - Avoiding network congestion and minimizing interference due to handover of multiple users in a high mobility scenario (disconnecting and reconnecting at the same time many users from one cell to another may increase RAN and network congestion).

Service Continuity
Potential system requirements:

- Subject to operator's policy, the system shall be able to change the IP anchoring point for a UE with minimal impact on the user experience.

Low mobility devices (duplicate with Machine-to-machine and sensors)

A.4.4 Efficient content delivery

In-network caching
Potential system requirements:

- Efficiently deliver or forward content from in-network entities controlled by the operator.
- Provide charging, Lawful Interception (LI) and QoS differentiation for content delivered from an in-network caching entity.
- Enable a flexible deployment of content caching entity located at multiple locations within the network (e.g. at various radio sites and local aggregation points).
- Support a content caching entity that is capable of being integrated within a device under the control of the operator.
- Authorized UE shall be able to receive cached content broadcasted by content caching entity.
- Efficient delivery of content from an appropriate caching entity, e.g. a cache located close to the user.

ICN Based Content Retrieval
Example: Information Centric Networks (ICN) and Content Centric Networks (CCN) that enable discovery, routing and caching based on named content.

Wireless Briefcase
Example: Providing a user with Personal Content Management (PeCM) of all of their traditionally stored HDD information in the form of a Flat Distributed Personal Cloud (FDPeC) facilitated over the 3GPP communications network.

Potential system requirements:

- Ability to securely store the personal data information/files of a user in such a way that they are retrievable with no perceptible delay to the user.
- Mechanism to control the upload and download of personal information/files between the 3GPP device and a server in the network (e.g., Flat Distributed Personal Cloud).

A.4.5 Self-backhauling

Wireless Self-Backhauling
Potential system requirements:

- Flexible partitioning of resources between access and backhaul functions when supported in a common band, including quasi-static provisioning of separate access and backhaul resources, and dynamic allocation of access and backhaul resources, e.g., based on current local conditions.
- Autonomous neighbour discovery and link setup, self-configuration of addressing and forwarding plane, and autonomous integration into core/OAM.
- Use of multiple RATs to increase service availability and network resiliency.
- Multihop wireless network topologies.
- Network topologies with redundant connectivity and paths to minimize service disruptions due to network dynamics.
- Dynamic adaptation to topology changes (e.g., due to node additions, node failures, link fluctuations.)

A.4.6 Access

On-demand networking
See eMBB.

Best Connection per Traffic Type
Example: A user has two applications running, one voice and one video streaming application. The two applications have very different requirements, as one is generating low volume, real time traffic that needs to access MNO services, and the second requires much higher data rates and access to the closest Content Distribution Network (CDN). If the user is in the coverage area of multiple cells, the best cell for the given application should be used, so that the traffic is routed in optimal manner.

Potential system requirements:

- Provide a mechanism such that a specific traffic type (from a specific application or service) to/from a UE can be routed via specific RAN nodes, and traffic in one RAN node can be offloaded towards a defined IP network close to the UE's point of attachment to the access network, while other traffic type to/from that same UE is not offloaded.

Multi Access network integration
Potential system requirements:

- Enable the capability to connect to multiple non-3GPP and 3GPP access networks in order to allow the operator to improve the efficiency in the exploitation of the network infrastructure and to provide the best capabilities to end-user.
- Support at least mobility between 3GPP and non-3GPP networks with optional session continuity, capability for the UE based on network control to select the access to connect to.

Multiple RAT connectivity and RAT selection
Potential system requirements:

- Provide data transmission by using both the 5G New RATs and E-UTRA simultaneously.
- Select a radio access (either a 5G New RAT or E-UTRA) to assign each data flow, taking into account e.g. service, traffic characteristics, radio characteristics, and UE’s moving speed.

Temporary Service for Users of Other Operators in Emergency Case
Potential system requirements:

- Support temporary service for users of other than home operators as temporary users in emergency case by serving operator policy.
- Support defining the limited set of necessary communication services and acceptable terminal features for temporary users by serving operator policy.
- Support an appropriate level of communications security for temporary service.

High Accuracy Enhanced Positioning (ePositioning)
See **CriC.**

Use Case for Green Radio
Potential system requirements:
The 3GPP system shall be capable of achieving [1000] times energy efficiency compared to legacy system.

A.4.7 Migration and interworking

Migration of services from earlier generations
Potential system requirements:

- Support the following operational requirement defined in previous releases of EPS: RAN Sharing.

Coexistence with legacy systems

Potential system requirements:

- Support seamless handover and Inter System Mobility between 5G RAT(s) and E-UTRAN.
- Seamless handover between the 5G RAT(s) and GERAN or UTRAN is not required.

A.5 eV2X

Mobile broadband services with seamless wide-area coverage
See eMBB.

Improvement of network capabilities for vehicular case
See CriC.

Connected vehicles
Examples:

- Autonomous driving
- Video information between vehicles and infrastructure is needed to further enhance the efficiency and safety, where higher data rate between vehicles, and between vehicles and infrastructure is required. The required uplink data rate per vehicle is tens of Mbps, while in case of video cameras providing to the car low-latency images of the hidden part of the road, the data rate may be hundreds of Mbps. The image receptor can be a UE connected to a processor controlling the steering system of the self-driving car. The hidden part of a road can be an intersection, the exit of a garage or of a parking lot, a serpentine road.

Potential system requirements:

- Very low latency (e.g., 1 millisecond end-to-end latency).
- Very high reliability (e.g., nearly 100%).
- Support high uplink data rate per vehicle even in a dense environment (e.g., tens of Mbps per device in a dense environment).
- High downlink data rates (e.g. above one hundred Mb/s) for transmitting images taken by the infrastructure in a dense deployment, i.e. starting from every 20-30m.
- Very high mobility (e.g., absolute speed more than 200 km/h while relative speed more than 400 km/h).
- Data transmission from one point to multipoint (e.g. multicast and/or broadcast).
- High positioning accuracy (e.g. 0.1 meters).
- High density of connections for vehicles (e.g. the number of vehicles can exceed 10000 in scenarios with multiple lanes and multiple levels and types of roads).