A Flexible and Reconfigurable 5G Networking Architecture Based on Context and Content Information

D. Sabella[1], P.Serrano[2], G.Stea[3], A.Virdis[3], I. Tinnirello[4], F. Giuliano[4], D. Garlisi[4],
P. Vlacheas[5], P. Demestichas[8], V. Foteinos[5], N. Bartzoudis[6], M. Payaró[6], A. Medela[7]

[1]Intel, Germany; [2] Universidad Carlos III de Madrid, Spain; [3] Università di Pisa, Italy; [4] CNIT, Italy;
[5] WINGS ICT solutions, Greece; [6] CTTC/CERCA, Spain; [7] TST Sistemas, Spain; [8] University of Piraeus, Greece

***Abstract* — The need for massive content delivery is a consolidated trend in mobile communications, and will even increase for next years. Moreover, while 4G maturity and evolution is driven by video contents, next generation (5G) networks will be dominated by heterogeneous data and additional massive diffusion of Internet of Things (IoT). The current network architecture is not sufficient to cope with such traffic, which is heterogeneous in terms of latency and QoS requirements, and variable in space and time. This paper proposes architectural advances to endow the network with the necessary flexibility helping to adapt to these varying traffic needs by providing content and communication services where and when actually needed. Our functional hardware/software (HW/SW) architecture aims at influencing future system standardization and leverage the benefits of some key 5G networking enablers described in the paper. Preliminary results demonstrate the potential of these key technologies to support the evolution toward content-centric and context-aware 5G systems.**

Keywords— flexible HW/SW platforms, 5G networking, context-awareness; content-centric networks; reconfigurability.

I. Introduction

Current mobile networks, from 2G to recent LTE-Advanced systems, have been designed and deployed with a connection-centric mindset, and static coverage and capacity. Mobile operators have been constrained by the inflexibility of their Radio Access Network (RAN) equipment, limiting the communication and computation capabilities as well as their reconfigurability and flexibility. In the last few years we all have witnessed an explosion of the traffic demand (dominated by video contents [1]) and a paradigm shift in terms of service usage, where content sharing and social behavior are redefining the way the network is used. The next years will also witness a dramatic increase of Machine-to-machine (M2M) connections [2] due to the progressive introduction of Internet of Things (IoT) traffic and services, that will be dominated by several new vertical business segments [3] (e.g. automotive and mobility, factories of the future, healthcare, media and entertainment, energy).

As a direct consequence of the connection-centric characteristics discussed out above, today's mobile networks are not effective in providing the required flexibility to adapt to varying traffic needs and to provide content and communication services where and when actually needed, and should therefore evolve in order to meet all the above challenges. To migrate toward content-centric networks, substantial changes in the network architecture are needed, adding more flexibility and adaptiveness, while preserving the communication efficiency and effectiveness. According to our vision, this evolution can only be achieved through technological improvements enabled by new hardware (HW) and software (SW) platforms. Moreover, their adaptiveness should stem from the ability to manage context information from measurements of network equipment and sensors, in order to gain insight over traffic, mobility, users, conditions and crowds, etc. Thus, the migration should be not only towards content-centric networks, but also towards context-centric networks.

In this paper, we propose significant architectural advances for the evolution toward context and content-centric networks, with the explicit aim to influence future 5G systems. We argue that this evolution should capitalize on the availability of reconfigurable radio platforms for the Radio Access Network (RAN) and take advantage of virtualization technologies, such as Cloud-RAN and the standardized Network Function Virtualization (NFV) framework [4], including virtual RAN (vRAN). On top of that it should account/incorporate/adopt Software Defined Network (SDN) technologies and the recent Multi-Access Edge Computing (MEC) [5]. The above are commonly considered as key ingredients of future 5G systems [3] that will provide operators with the required flexibility and reconfigurability to satisfy this increasing traffic demand.

The key enablers for the proposed architecture are thus:

1) context-aware reconfigurable radio platforms, where a flexible, efficient and reconfigurable HW/SW adapts the access network characteristics based on data and context information;

2) content-centric resource allocation strategies, especially implemented in C-RAN/vRAN environments, whereby intelligent baseband (BB) processing within vRAN exploits the presence of information related to the aggregation point of several base stations;

These enablers can be realized by exploiting radio network information gathering through MEC framework and APIs (possibly co-located with the same virtualized infrastructure hosting vRAN and other virtualized technologies), where context information from the RAN can be provided to user level applications or other services for network performance and Quality of Experience (QoE) improvements.

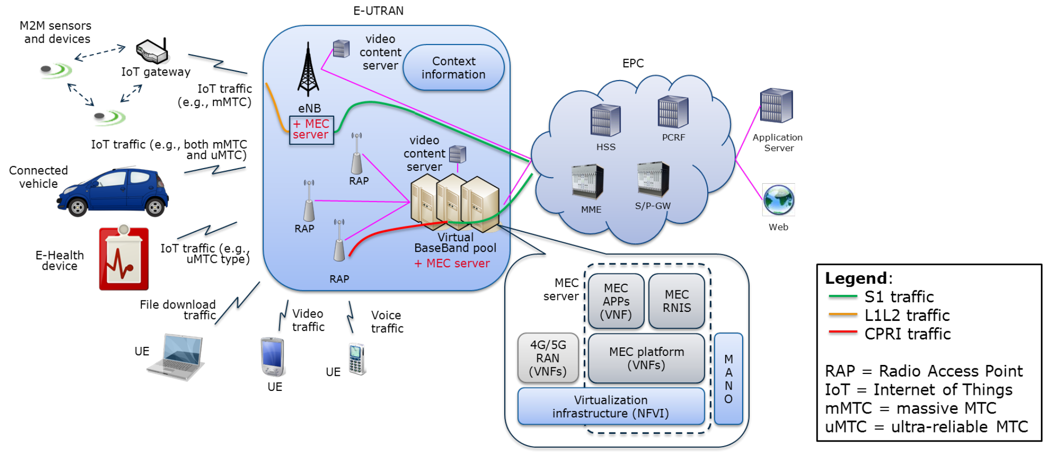

Fig. 1. *Graphical representation of the reference system considered in this paper*

These enablers together with the reference system considered in this paper have been depicted in Fig. 1.

The remainder of this paper is organized as follows: we present in Section II an overview of proposed functional architecture; Section III describes the key technology enablers and Section IV shows some relevant numerical results. Finally, Section V discusses implementation aspects and standardization impacts, while Section VI concludes the paper.

II. Functional Architecture

The functional architecture that we envision, developed within the Flex5Gware 5G-PPP project [17], is illustrated in Fig. 2 and relies upon three key concepts: flexibility, re-configurability and monitoring. It could be argued that our approach falls between some other 5G initiatives, aiming at the integration of SDN/NFV [18] or Radio Access Network [19] in a common architecture. Flex5Gware focuses on integrating the above three key concepts, traditionally associated with cognitive radio [20], into a 5G reference architecture. The functional modules are classified into three layers, depending on their focus, namely: the design of a single device, the coordination of devices, and the control and optimization of the network.

Considering **node operation**, the most relevant concept is flexibility: in order to cope with the high variability of 5G application requirements and network topologies, 5G technologies will support advanced reconfiguration capabilities at both the PHY and medium-access levels. Our architecture moves from the traditional approach of "one-size-fits-all" protocol MAC/PHY stack to an innovative paradigm of on-the-fly configuration of context-specific stacks.

The **intelligent programs** layer consists on modules that optimize network performance, based on an information-centric operation and exploiting the re-configurability and monitoring features. The monitoring library is an API that provides access to the data collected by sensors and monitoring agents throughout the network, and feeds two modules that extract network-wide performance and context estimations, which in turn trigger different types of optimizations. For instance, performance degradations or context variations may trigger the activation of additional network elements or switching to a Radio Access Technique that is more robust. Moreover, a global scheduler coordinates the operation of the base stations, enabling the C-RAN and vRAN vision as described next, and a service scheduler that coordinates the optimization of all these elements, to enable a smooth operation precluding conflicts.

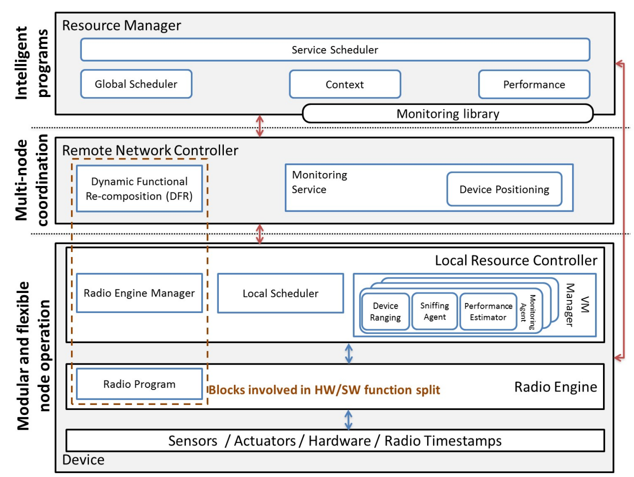

Fig. 2. *Functional architecture for context and content-centric 5G platforms*

Connecting these two layers is achieved by the **multi-node coordination**, composed of: i) a monitoring service, which gathers data from the local monitoring agents of the multiple devices (including sensors), processes their information and forwards it to the intelligent programs; and ii) a Dynamic Functional Re-composition, which stitches together software and hardware functions and abstracts the changing of the operation of multiple devices to the performance optimisers, hence providing a technology-agnostic reconfigurability service to the intelligent programs.

As an example of field-use of the proposed architecture, content-based system requirements could trigger a network-wide partitioning and reconfiguration of hardware-accelerated and software baseband functions, tailored for different traffic service delivery and QoE needs. Out of the different available options, the flexible partitioning could result in placing i) the eNB layer-2 and above together with the EPC functions in the Cloud (e.g., S1 traffic), ii) the eNB layer-2 and above functions at a local MEC server (L1L2 traffic) or iii) move the entire eNB stack in the Cloud (CPRI traffic) and transform the eNB to a Remote Radio Head (RRH). This dynamic split and reconfiguration of baseband functions is complementary to the partitioning of the eNB protocol stack, either at stack or algorithm level (e.g., MAC-PHY, RLC, PDCP) that was promoted relatively recently by key industry actors [11].

III. Key Technology enablers for Content-centric and Context Aware 5G Networks

In this section we describe some key technology enablers that stem from the functional architecture, together with a clear comparison against the state of the art methods. For each introduced technology enabler, next section will describe respectively their preliminary performance evaluation results coming from simulations and experimental works, conducted in the framework of Flex5Gware project.

Context-aware reconfigurable radio platform

Device flexibility requires the (re)configuration of context-specific stacks on the fly, based on context-related triggers. The proposed approach consists of a radio engine that runs radio programs over a set of HW primitives controlling the behavior of HW devices, e.g., analog front-ends, converters, mixers, filters, base-band modulators, transmission queues, and sensors (like accelerometers or cameras). Radio programs are built similarly to computer programs, sent over the air to the radio engine manager, stored in an internal memory, and changed over time. Changes can be triggered not only by traditional performance-related metrics such as delay or bandwidth, or PHY parameters such as signal strength, but also by context estimators (e.g., weather information, accurate localization including indoor/outdoor information, relative position of the mobile to the user) and other, new performance-related metrics, such as energy consumption of HW elements and CPU usage.

The feasibility of this approach has been recently demonstrated by some preliminary work on programmable radio platforms, such as the Wireless MAC Processor architecture for Wi-Fi commercial cards [6] and the TAISC architecture for ZigBee sensor nodes [7].

In this paper, we focus on dynamic stack adaptations for context and content-centric applications by measuring the performance of D2D links based on a contention-access protocol, such as the WiFi Distributed Coordination Function (DCF), that can be activated in areas where sensors are densely deployed and generate periodic traffic. Sensors transmit data to a base station by means of a schedule-based protocol, but channels are allocated in unlicensed bands to offload the cellular network. We also assume that, according to the rate of the sensor data, the frame allocations used for sensors transmissions can be interleaved with empty (blank) frames, as considered in [10].

In these conditions, DCF devices will suffer from severe collisions, because they will initiate transmissions that will be disrupted by (non-channel sensing) sensors later on. The higher is the frequency of non-empty frames, the higher is the collision probability. DCF devices react to collisions by increasing the contention windows, thus reducing the channel access rate. If DCF devices employ rate adaptation, they will also aim for lower rates, thus increasing the transmission times and, accordingly, the number of collisions.

While DCF devices cannot detect the cause of performance impairments, let alone act on it, 5G context-aware networks can easily predict the interference sources in a given geographical area (e.g. by means of positioning data of the sensors) and modify accordingly the PHY/MAC protocol to be used for D2D links; for instance, exponential backoff and rate adaptation can be disabled, or customized protocols can be downloaded to the devices to further optimize their performance.

One such protocol, called Future NAV (FNAV), forbids DCF devices from starting a transmission that cannot be completed and acknowledged before the start of sensor transmissions, which take place at predictable times. DCF devices will simply consider the medium as busy during these intervals, thus avoiding collisions.

Although the proposed protocol is not the only possible solution, the main novelty of our approach is the possibility to implement cooperation mechanisms between heterogeneous network solutions, based on radio flexibility. This is quite different from current approaches, for example in the coexistence scenario between LTE in unlicensed bands and WiFi, where the utilization of bandwidth reduction [9], duty-cycling muting or blank subframes [10] is unilaterally taken by LTE, while the implementation of listen-before talk before transmitting LTE frames requires to build new LTE radio [9][8].

Content-centric resource allocation strategies

A fundamental research challenge for 5G networks is to devise optimal resource allocation strategies that consider both content and context information. The Cloud-RAN (C-RAN) paradigm [14] is extremely promising in this area, as it allows the dynamic management of computational resources. For example, in C-RAN, computing power can be allocated to cells on-demand and according to their load. Thus, it is suitable to support the 5G requirement of extreme flexibility. Moreover, in C-RAN environments, issues such as spectral efficiency optimization and interference reduction, traditionally addressed through hardware solutions, can be tackled via a new, entirely software-based approach, using efficient algorithms which leverage the power of cloud computing.

For instance, dynamic interference coordination at fast timescales is now made challenging by the need to transfer state information among geographically sparse entities (i.e., BSs and a possibly co-located controller), due to the latencies involved in inter-cell communication [12]. The same information would travel considerably faster as inter-process communication among Baseband Units (BBUs) in a cloud, making fast-paced algorithms practicable. On the other hand, C-RAN aggregation points in the network are natural candidates to hosting MEC platforms, which can leverage a huge amount of information related to traffic flows, network status and radio conditions, so that MEC applications can manage all the contents to be delivered in the network effectively.

In addition, content-centric algorithms may lead to significant performance benefits. For example, in our studies, we examined a content-centric Coordinated Multi Point (CoMP) algorithm. CoMP has been studied in the literature as an interference-mitigation technique. Among CoMP techniques, Coordinated Scheduling (CS) focuses on coordinating the allocation of the physical resource blocks (PRBs) among eNBs [21][22]. CS can be done statically, ignoring traffic variations [23], or dynamically, by having a Coordinator arbitrate resource requests by the coordinated eNB based on their current traffic [24]. The coordination policy affects the performance of specific sets of users differently; hence the ability to select and change this policy flexibly becomes a key point to improve the user QoE. Context information, such as the position of user or the characteristics of data traffic for a particular hotspot (or group thereof) can help improving the performance in this respect. None of the works we are aware of considers context information within their algorithms.

Our content-aware CoMP (CoCo) algorithm exploits knowledge of the traffic content at a hotspot. On each period every eNB sends its PRB requests to the Coordinator: the latter partitions resources proportionally to the initial request, but giving a boost to those coming from eNBs serving a specific content type. To clarify this, let us assume a scenario where a crowd is moving towards a stadium, requesting a real-time augmented reality service, i.e., low bandwidth and strict delay requirements. In this case, the stadium hotspot is surrounded by macro cells generating high-bandwidth traffic. A content-unaware CoMP would base its decisions on traffic volumes only, thus penalizing the hotspot. A CoCo approach instead recognizes the type of content requested by the hotspot and protects its performance. Using this approach, the interference among eNBs will be effectively managed, and a low-delay content access will be provided to users under the hotspot at the same time.

IV. Experimental Work, Setup and Performance Evaluation

For each introduced technology enabler, this section describes respectively their preliminary performance evaluation results coming from simulations and experimental works, conducted in the framework of Flex5Gware project.

Context-aware reconfigurable radio platform

We implemented the above mechanism for the configurable radio platform called Wireless MAC Processor built for legacy WiFi cards [6]. In the experiment, the interference period is 20 ms, while the interference duration is 10 ms. In case of legacy access rules, the device would experience deterministically a successful data transmission and subsequent collision in each interference-free interval. Instead, as shown in Fig. 3, FNAV is able to prevent these collisions (thus saving energy and channel time).

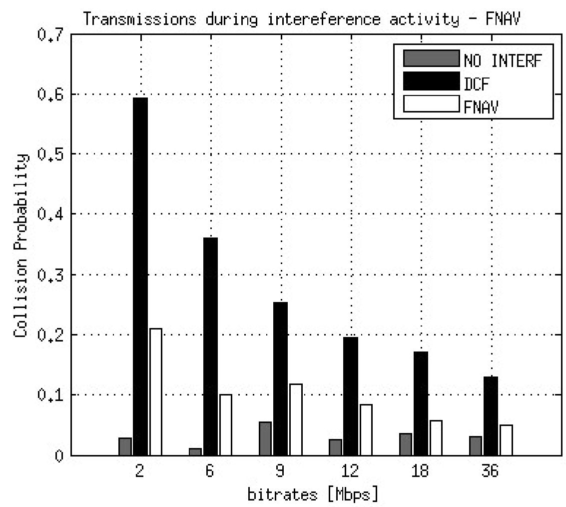

Fig. 3. Collision probability for different values of the bit rate

Content-centric resource allocation strategies

The simulation scenario for this key enabler includes three eNBs at a distance of 500 meters connected via X2 interface. During the simulation, each eNB serves 30 UEs. Real-time and best-effort traffic is generated at a remote server, where applications alternate active and silence periods. In active periods, constant-size packets are generated periodically. A summary of the main simulation parameters is provided in Table 1. Three load levels, namely low, medium and high, are obtained by setting the packets size of best-effort traffic, respectively to 500, 1000 and 1500 bytes.

Table 1 – Main simulation parameters

Parameter	Value
RBs	15
# independent replicas	3
Simulation duration	100s
Warmup period	10s
UEs	30 for each eNB
eNBs	3
Real-time traffic: Packet size Inter-packet time Active period duration Silence period duration	 50 bytes 20 ms (during active periods) 1s 5s
Best Effort traffic Packet size: Inter-packet time Active period duration Silence period duration	 {500, 1000, 1500} bytes 100 ms (during active periods) Weibull distributed, mean 0.8 Weibull distributed, mean 1.0

We simulated the above scenario using SimuLTE [13], and in Fig. 4 we compare the application delay experienced in the hotspot, obtained with the CoCo approach against a content-unaware Dynamic CoMP-CS, for three load levels. CoCo keeps the extra delay caused by the load increase reasonably low, ensuring a seamless responsiveness of the application, regardless of how crowded the hotspot is. The throughput reduction experienced in the surrounding macro cells is in the order of 1%, hence negligible.

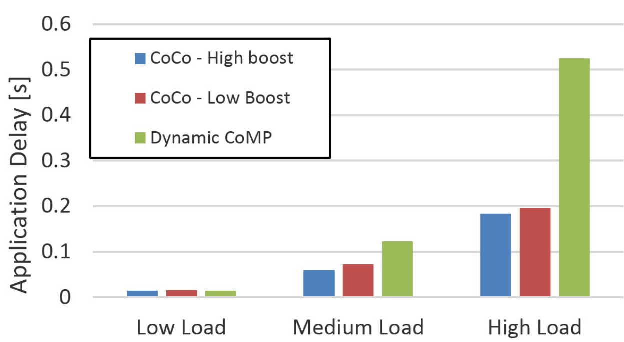

Fig. 4. Application delay with CoCO and dynamic CoMP

V. Technology Implementation aspects and Standardization Impacts

The two technologies described in the previous sections can be realized in 5G networks by means of customized implementations (depending on vendors choices) or in alternative by exploiting content and context measures coming from standardized MEC platforms and MEC APIs.

In particular, the Radio Network Information (RNI) API currently being standardized by ETSI MEC ISG have the following characteristics:

- expose up-to-date radio network information regarding the current radio network conditions;
- this radio network information is provided at the relevant granularity (e.g. per User Equipment (UE) or per cell, per period of time);
- include measurement and statistics information related to the user plane, based on information defined by 3GPP specifications;
- include measurement and statistics information related to UEs connected to the radio node(s) associated with the mobile edge host, their UE context and the related radio access bearers, and changes in it related to those UEs.

All these features, when implemented in a 5G system, will feed suitable MEC applications especially dedicated for the network performance and QoE improvements (as defined by ETSI MEC GS 002 [15]), where network operators and application developers will benefits from MEC RNI API (that can be seen as "middleware services" aimed at improving network performance via application-specific or generic improvements).

The Flex5Gware project, being focused on implementation and experimental activities, is actively contributing in ETSI MEC standard, especially for the testing and performance assessment of prototypes and advanced radio platforms [16]. In fact, this is the starting point for the evaluation of E2E network performances in presence of a MEC platform and related APIs. The next step, necessary to gain a wider market acceptance of MEC technology, will be the harmonization of ETSI MEC architecture with 3GPP standard, especially in the view the definition of a interoperable interface between the MEC RNI service and 5G radio base stations.

VI. Conclusions and Future Work

Current mobile networks (designed and deployed with a connection-centric mindset) are not effective in providing the required flexibility to adapt to varying traffic needs and to provide content and communication services where and when actually needed.

In this paper we proposed significant architectural advances for the evolution toward context and content-centric networks, with the explicit aim to influence future 5G standardized systems. In particular we highlighted the importance of the following key technology enablers:

- context-aware reconfigurable radio platforms, triggered by proper sensors and able to easily predict the interference sources in a given geographical area (e.g. by means of positioning data of the sensors) and to modify accordingly the PHY/MAC protocol to be used for D2D links (for instance, exponential backoff and rate adaptation can be disabled, or customized protocols can be downloaded to the devices to further optimize their performance);
- content-aware CoMP (CoCo) algorithms, able to exploit knowledge of the traffic content at a hotspot in order to improve performances with respect to traditional CoMP algorithms (unaware of the traffic content);

These enablers can be implemented by exploiting real-time radio network information gathering through MEC RNI API (recently introduced by ETSI MEC), as an essential tool to improve network performance and QoE up to the user application level. The next step, necessary to gain a wider market acceptance of MEC technology, will be the harmonization of ETSI MEC architecture with 3GPP standard for 5G networks. Future work may include also the implementation of a single prototype able to combine all these key technologies, as an integrated proof-of-concept of future flexible and reconfigurable 5G network platforms.

Acknowledgment

This work was supported by the European Commission in the framework of the H2020-ICT-2014-2 project Flex5Gware (Grant agreement no. 671563).

References

[1] Cisco White Paper "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021", Feb. 2017, Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf [Accessed Feb. 2017]

[2] Cisco White Paper, "Cisco Visual Networking Index: Forecast and Methodology, 2015-2020", June 2016, Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf [Accessed Feb. 2017].

[3] 5G Infrastructure Association: Vision White Paper, February 2015. http://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[4] ETSI GS NFV-SWA 001 V1.1.1 (2014-12), Network Functions Virtualisation (NFV); Virtual Network Functions Architecture

[5] ETSI white paper: "Mobile Edge Computing: A key technology towards 5G", September 2015; www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf

[6] Tinnirello, I. et al., "Wireless MAC processors: Programming MAC protocols on commodity Hardware," in Proc. of IEEE INFOCOM 2012, pp.1269-1277, 25-30 March 2012

[7] Jooris, B., Bauwens, J., Ruckebusch, P., De Valck, P., Van Praet, C., Moerman, I., & De Poorter, E. (2016). TAISC: a cross-platform mac protocol compiler and execution engine. Computer Networks. Vol. 107, pp 315–326

[8] Yubing Jian, Chao-Fang Shih, Bhuvana Krishnaswamy, Raghupathy Sivakumar. "Coexistence of Wi-Fi and LAA-LTE: Experimental Evaluation, Analysis and Insights." *IEEE International Conference Communication Workshop (ICCW), 2015*

[9] A. M. Cavalcante et al., Performance "Evaluation of LTE and Wi-Fi Coexistence in Unlicensed Bands" *2013 IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, 2013, pp. 1-6.*

[10] Almeida, E., *et al.* "Enabling LTE/WiFi coexistence by LTE blank subframe allocation," in *Proc. O IEEE ICC 2013*, pp.5083-5088, 9-13 June 2013

[11] "Virtualization for small cells: Overview", Small Cell Forum, Release 7, version SCF106.07.01. URL: http://www.scf.io/en/documents/106__Virtualization_for_small_cells_Overview.php

[12] G. Nardini, A. Virdis, G. Stea, "Modeling X2 backhauling for LTE-Advanced and assessing its effect on CoMP Coordinated Scheduling", IWSLS2 2016, Vienna, July 1st, 2016

[13] A. Virdis, G. Stea, G. Nardini, "SimuLTE: A Modular System-level Simulator for LTE/LTE-A Networks based on OMNeT++", SimulTech 2014, Vienna, AT, August 28-30, 2014

[14] NGMN, "Suggestions on Potential Solutions to C-RAN by NGMN Alliance," Jan 2013. Available: www.ngmn.org/uploads/media/NGMN_CRAN_Suggestions_on_Potential_Solutions_to_CRAN.pdf

[15] ETSI GS MEC 002 V1.1.1 (2016-03): "Mobile Edge Computing (MEC); Technical Requirements"

[16] ETSI GS MEC-IEG 006 V1.1.1 (2017-01): "MEC Metrics Best Practice and Guidelines"

[17] Flex5Gware website: http://www.flex5gware.eu/

[18] P. Rost et al., "Mobile network architecture evolution toward 5G," in IEEE Communications Magazine, vol. 54, no. 5, pp. 84-91, May 2016.

[19] Patrick Marsch, Olav Queseth, Salah-Eddine El Ayoubi, Michał Maternia, Mikko A. Uusitalo, Rauno Ruismaki, Milos Tesanovic, Alexandros Kaloxylos, Icaro Da Silva and Mauro Boldi, "5G Radio Access Network Design – A Brief Overview on the 5G-PPP Project METIS-II", EuCNC Special 5GPP session, 29 Jun – 2 Jul 201

[20] P. Pawelczak, K. Nolan, L. Doyle, S. W. Oh and D. Cabric, "Cognitive radio: Ten years of experimentation and development," in IEEE Communications Magazine, vol. 49, no. 3, pp. 90-100, March 2011.

[21] S.H. Ali, V.C.M. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks", IEEE Transactions on Wireless Communications, vol.8, no.8, pp. 4286-4295, Aug. 2009.

[22] K. Hoon, H. Youngnam, J. Jayong, "Optimal subchannel allocation scheme in multicell OFDMA systems", Proc. of VTC Spring'04 pp.1821-1825 Vol.3, 17-19 May 2004.

[23] L. Fang, X. Zhang, "Optimal Fractional Frequency Reuse in OFDMA Based Wireless Networks", Proc. WiCOM '08, pp.1-4, 12-14 Oct. 2008.

[24] N. Iardella et alii, "Flexible dynamic Coordinated Scheduling in Virtual-RAN deployments", FlexNets 2017, Paris, FR, 20th May 2017