The METIS 5G Architecture

A Summary of METIS work on 5G Architectures

Heinz Droste[1], Gerd Zimmermann[1], Makis Stamatelatos[2], Neiva Lindqvist[3], Ömer Bulakci[4], Josef Eichinger[4],
Venkatkumar Venkatasubramanian[5], Uwe Dötsch[6], Hugo Tullberg[3]
[1]Deutsche Telekom, [2]National and Kapodistrian University of Athens, [3]Ericsson Research, [4]Huawei ERC, [5]Nokia Networks,
[6]Alcatel-Lucent

***Abstract*—During the last two years, the METIS project ("Mobile and wireless communications Enablers for the Twenty-twenty Information Society") has been conducting research on 5G-enabling technology components. This paper provides a summary of METIS work on 5G architectures. The architecture description is presented from different viewpoints. First, a functional architecture is presented that may lay a foundation for development of first novel 5G network functions. It is based on functional decomposition of most relevant 5G technology components provided by METIS. The logical orchestration & control architecture depicts the realization of flexibility, scalability and service orientation needed to fulfil diverse 5G requirements. Finally, a third viewpoint reveals deployment aspects and function placement options for 5G.**

Keywords—5G, METIS, Multi-RAT, Ultra-Dense Networks, Massive MIMO, Dynamic RAN, Functional Architecture, Orchestration & Control Architecture, Deployment Architecture

I. INTRODUCTION

The main objective of METIS [1] is to respond to societal challenges beyond 2020 by providing the basis for the all-communicating world, and by laying the foundation for a future mobile and wireless communications system.

A set of specific technical goals has been derived from these main objectives including [2]:

- 1000 times higher mobile data volume per area,
- 10 to 100 times higher typical user data rate,
- 10 to 100 times higher number of connected devices,
- 10 times longer battery life for low-power devices and
- 5 times reduced End-to-End (E2E) latency.

It can be deduced that these goals are partly contradicting and, hence, an integrated system concept has to be highly flexible and scalable in order to fulfil such requirements in a cost-efficient and affordable way.

In order to decompose the complex task of system development into clearly arranged tasks specific concepts have been developed for each of the METIS Horizontal Topics (HTs) [3]: Direct Device-to-Device Communication (D2D), Massive Machine Communication (MMC), Moving Networks (MN), Ultra-Dense Networks (UDN), and Ultra-Reliable Communication (URC).

The D2D concept addresses the utility of the local exchange of information among devices and creates a framework for solving the associated technological challenges. The MMC concept incorporates the radio access technologies enabling support of an unprecedented number of devices. The MN concept introduces innovative directions for the future relationship between vehicles and wireless communications. UDN is a specific concept optimized for the potential stand-alone operation of ultra-densely deployed small cells. The URC system concept targets operational modes that are not present in today's systems. In particular, whereas URC-L (Long-term URC) targets at moderate minimum rates provided to all users, URC-S (Short-term URC) aims to guarantee latency as needed for time-critical services. URC-E (URC for Emergency) aims to provide minimal guaranteed connectivity upon infrastructure damage.

Taking commonalities of the HT-specific concepts into account, three generic services have been identified [4]:

A. *Extreme Mobile Broadband (xMBB)*

xMBB addresses the increasing traffic volume and data rates required by new applications, such as virtual or augmented reality. Improved user experience in terms of guaranteed minimum data rates and smart content delivery will also be necessary to be provided by 5G systems.

B. *Massive MTC (mMTC)*

mMTC concerns massive deployments of, e.g., low-cost battery-powered sensors and actuators, remote-controlled and -readable utility meters. 5G systems must provide up- and down-scaling connectivity solutions for tens of billions of devices since it is expected that there will be 10-100 more connected devices per one human user of communications systems (for human interaction, connected machines owned by the user and devices owned, e.g., by the city the user lives in).

C. *Ultra-reliable MTC (uMTC)*

uMTC relates to the capability to provide a given service level with very high probability. It also includes applications where low delay is a critical factor, such as remote driving, industrial control, and haptic communication enabling remote work in, e.g., hazardous environments or remote surgery.

The various kinds of MTC will enable the wireless Internet of Things (IoT) encompassing tens of billions connected devices with diverging service and traffic demands.

II. MULTI-FACIAL ARCHITECTURE DESCRIPTIONS

A. *Functional Architecture*

Fig. 1 illustrates the main building blocks (BBs) identified within METIS from functional architecture point of view.

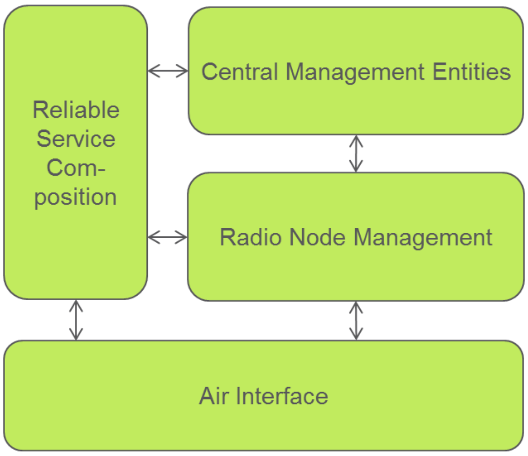

Fig 1. Main building blocks of METIS 5G architecture.

Each main BB can be hierarchically split into a number of sub-BBs. These sub-BBs can be “common BBs”, containing functionalities required for more than one HT concept, and “HT-specific BBs”, which are essential for enabling a single HT concept. Each sub-BB is finally described through a set of more fine granular Functional Elements (FEs) with each FE performing an inherently consistent logical task. FEs have been derived by functional decomposition of prioritized technology components (TeCs) developed in METIS [4]. Notably, a TeC comprises a specific methodology, algorithm, module or protocol enabling certain system features and contributing to the fulfillment of specific technical requirements. It has to be also noted that METIS is primarily focusing on the radio access network (RAN) part, so not all components required to finally build and operate a 5G system are covered with those BBs.

The functional architecture can be decomposed into:

- Central Management Entities (CMEs), containing BBs that cover network overarching functionalities which are not specific for certain HTs and use cases/scenarios. Typical examples are Context Management and Spectrum Management. These BBs are usually more centrally arranged. However, depending on the use case, a partially distributed realization might be possible, as well.
- Radio Node Management (RNM) containing BBs that provide radio functionalities that usually affect more than one radio node and that are not HT-specific. Exemplary functions are Long-/Short-Term Radio Resource & Interference Management, Mobility Management, Radio Node Clustering & (De-) Activation, and D2D Device Discovery & Mode Selection. In principle those functions will be deployed at medium network layers (e.g., at dedicated Cloud-RAN (C-RAN) nodes [5]). The interface requirements between FEs that are mapped to those BBs (especially the air interface sub-BBs) have strong impact on the function placement.
- Air Interface (AI) including BBs that are directly related to air interface functionalities of radio nodes and devices. It comprises HT-specific as well as common BBs. Examples are AI enablers for UDN or for different types of MMC applications.
- Reliable Service Composition represents a central C-Plane functionality with interfaces to all other main BBs. It is used for availability evaluation and/or provisioning of ultra-reliable radio links which can be applied for novel service types requiring extremely high reliabilities in message data transfer and/or extreme low latencies (e.g.. industrial environments, eHealth, or V2X (vehicle-to-anything) communication).

B. *Logical Orchestration & Control Architecture*

The METIS 5G architecture development is driven by three key aspects; flexibility, scalability, and service-oriented management. The envisioned logical orchestration & control architecture (see Fig. 2) is based on usage of upcoming architectural trends, such as Software Defined Networking (SDN) [6][7] and Network Function Virtualization (NFV) [8][9]. It will provide the necessary flexibility for realizing efficient integration and cooperation of FEs according to the individual service needs as well as future evolution of existing cellular and wireless networks [10][11][12].

Network functions (NFs) derived from FEs are flexibly deployed and instantiated by the 5G Orchestrator (consisting of NFV Orchestrator, Virtual NF (VNF) Manager and Virtualized Infrastructure Manager [8] as well as their extensions Service-oriented function Processing Manager and Service-oriented Topology Manager). It is responsible for managing all VNFs of the 5G network including radio, core and service layer by mapping logical topologies of C-/U-Planes to physical resources in the deployment architecture dependent on corresponding logical topologies for each service.

The Service Flow Management is analyzing the customer-demanded services and outlining their requirements for data flows through the network infrastructure. These requirements are communicated to 5G Orchestrator and 5G-SDN Controller. Application/service requirements (e.g., from a 3rd party service provider), like maximum delay and/or minimum bandwidth on data flow path, can be taken into account through dedicated Application Programming Interfaces (APIs). The architecture enables on-demand set-up of customized virtual networks (VNs) using shared resource pools and allowing effective service-adaptive decoupling of C- and U-Plane plane in order to optimize routing and mobility management across the whole service transport chain.

Radio Network Elements (RNEs) and Core Network Elements (CNEs) in the orchestration & control architecture are logical nodes that are specified having in mind the possibility to be implemented on different software/hardware (SW/HW) platforms (both virtualized and non-virtualized). The 5G Orchestrator is interfacing with RNEs and CNEs via the Function Agent (FuAg) by which it performs the

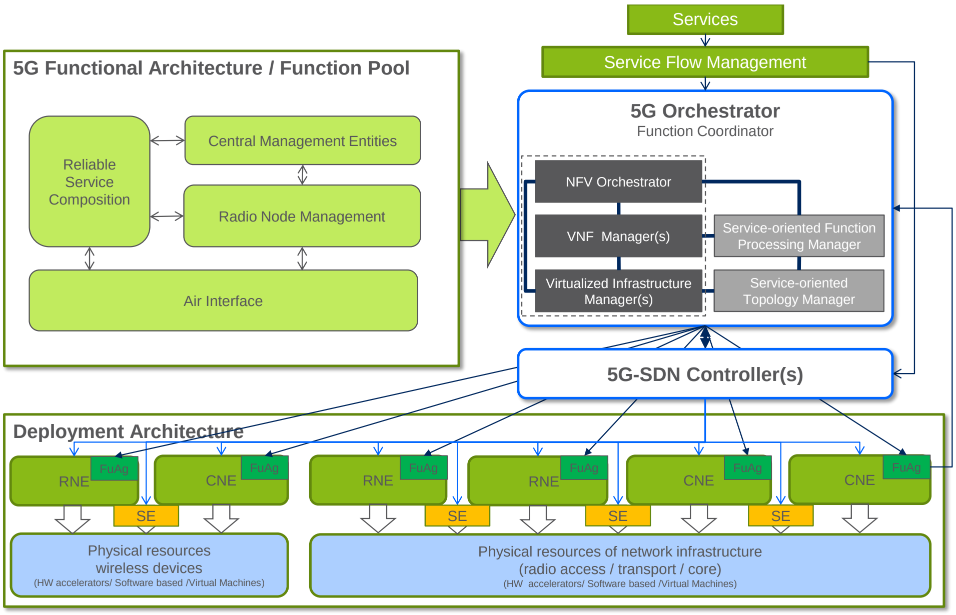

Fig. 2. Logical orchestration & control architecture of METIS 5G system

configuration according to service requirements, also known as service orchestration. In terms of the communication between different RNEs (including wireless devices), flexible protocol functionalities and properly configured AI variants will co-exist by applying varying radio-related NFs customized within VN slices according to requirements of target services and associated 5G use cases.

It is expected that, increasingly, the HW platforms designed to run RNEs are capable of supporting NFV to a certain extent, but especially low-cost equipment – such as small cell nodes for UDN – will probably be realized without or still limited NFV capabilities due to cost reasons. In contrast to that, CNE-related computing platforms allow fully flexible deployment of NFs based on virtualization concepts, which is already happening today in 4G systems [9]. Thanks to the anticipated 5G flexibility, certain core-related functionalities can also be moved to same physical nodes, where RNEs are implemented (e.g., deploying AAA and mobility management functions in a C-RAN environment to reduce the service latency).

The 5G-SDN Controller is setting up the service chain on the physical network infrastructure taking into account the configurations orchestrated by the 5G Orchestrator. The 5G-SDN Controller (implementation as VNF is also possible) then constructs the U-Plane processing for the data flow, i.e., it builds up the connections for the service chain of CNEs and RNEs in the physical network. Accordingly, it configures the switching elements (SEs) (e.g., utilizing OpenFlow [13]) taking care of more radio-related functionalities, e.g., of the mobility management.

The flexibility is restricted by limitations of physical network elements (NEs), but also by pre-coded accelerators implemented in certain nodes, e.g., hard-coded physical layer procedures in order to minimize processing delay and energy consumption. Those node capabilities are reported by the FuAg at RNEs and CNEs and are taken into account by the 5G Orchestrator.

C. Deployment Architecture

Fig. 3 introduces an E2E reference network that is used when functional placement within the network topology is discussed. This reference network shows how the different types of sites are located along the access, aggregation and core networks within a typical telecom operator system.

The model includes devices, e.g., terminals and D2D groups, antenna sites, e.g., small cells, relay nodes, cluster nodes, as well as radio base station (RBS) sites. In addition, data centers with data processing and storage capabilities at access and aggregation level are depicted.

In principle, NFs can be deployed at all those sites in a flexible architecture, but finally, it strongly depends on the underlying service/use case requirements. Important requirements include latency and throughput on the input and output interfaces, time synchronization (e.g., on radio time slot level) and scaling of processing (e.g., relation to U-Plane throughput).

In order to enable positioning within the network topology two types of NFs are distinguished:

- Synchronous NFs for which processing is time-synchronous to the 5G AI (slots/frames). They typically require high throughput on the interfaces, which scales with traffic load, overall radio bandwidth, and number of antennas. Potential for centralization at certain sites is limited (e.g., in a range in the order of 10-20 km), since it would only work in case of an efficient transport network with very low latencies. Potential for virtualization is limited because of the timing, real-time processing requirements and since synchronous functions benefit a lot from local HW acceleration.

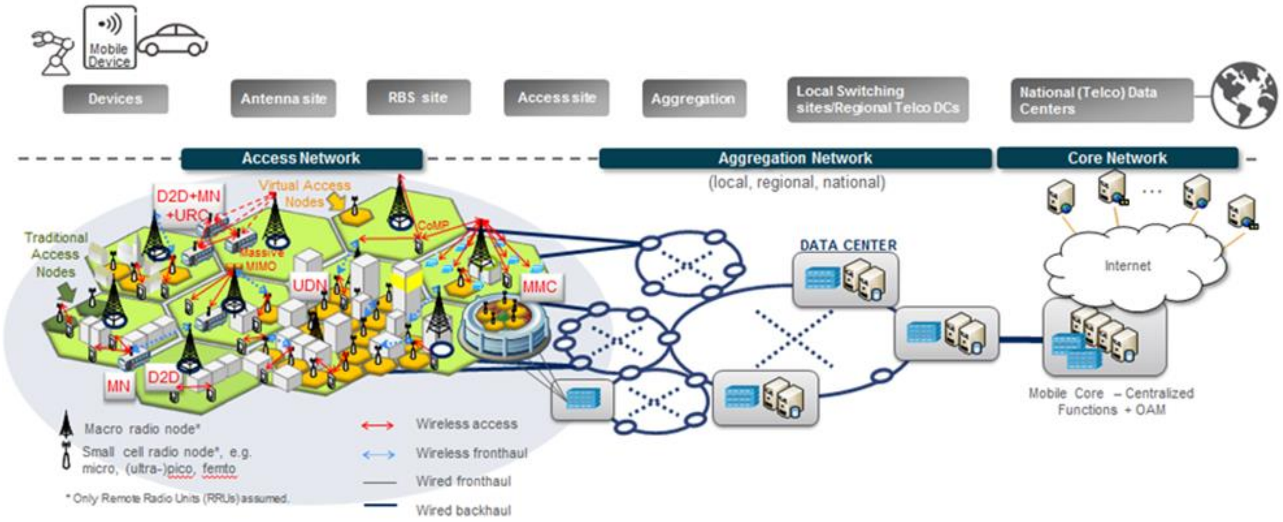

Fig. 3. METIS E2E Reference Network

- Asynchronous NFs for which processing is time-asynchronous to the 5G AI (slots/frames). They typically require low throughput on the interfaces, and the processing requirements scale with the number of users, but not with the overall traffic load. These functions can typically cope with tenths of milliseconds of latency and are therefore good candidates to be deployed on more centralized and virtualized platforms.

In order to address architectural aspects of different technical enablers, architecture options including related functional deployment combinations have been investigated. These architecture options are discussed in more detail in Section III.

III. ARCHITECTURE OPTIONS

Architecture options refer to dedicated functional architectures that correspond to METIS technical enablers. In the following, most important options are discussed.

A. *Centralized vs. Distributed Deployment of NFs*

From economical point of view, a maximized central deployment of NFs would be desirable. But, due to challenging latency and/or throughput requirements, the degree of freedom for deployment of NFs is limited. Synchronous radio functions must be deployed in the vicinity of the baseband processing. In that context, direct localization of radio functions at antenna sites is called Distributed RAN (D-RAN). The centralization of baseband processing in C-RANs stands for the alternative deployment option. Furthermore, asynchronous NFs may be placed more freely in the network topology according to operational needs.

B. *Massive MIMO (M-MIMO)*

M-MIMO [14] provides means for significant improvement of spectral efficiency in access, but in particular also for wireless backhaul. Large antenna dimensions will force a distributed deployment of M-MIMO arrays at lower frequencies. At high frequencies M-MIMO will be a powerful means to increase link budgets and decrease interference. Due to high number of RF chains and, therefore, high fronthaul capacity requirements in a C-RAN environment, radio functions including baseband processing will have to be deployed directly at the antenna.

C. *Ultra-Dense Networks*

The UDN concept is targeting at local ultra-dense deployment of small cells. Tight node collaboration and utilization of spectrum at higher frequencies with significant increased bandwidth compared to today’s legacy small cells are the most important differentiators. There will be a fast interaction between coverage layers (Multi-RAT), whilst fast activation/deactivation of radio nodes will be vital to ensure energy efficient operation. Especially in cm Wave (cmW) frequency bands, where spectrum is scarce, tools for flexible spectrum management will be required for UDNs.

D. *Multi-RAT*

Tighter integration of multiple RANs in 5G allowing for direct interaction will enable multi-connectivity for data rate aggregation, control plane diversity as well as flexible uplink and downlink split. Multi-RAT will play an important role in realizing coverage improvements without high operator investments.

E. *Device Networks and Dynamic RAN*

Within the framework of Dynamic RAN [4], we consider the network as a whole which takes into account any connected or connectable NE. In such a dynamic RAN environment, the autonomous moderation of nomadic nodes and moving relays as well as context-aware cluster heads and device networks are envisaged.

F. *Ultra-Reliable Machine Type Communication*

The automation industry sector can use wireless networks to improve operations, product quality, productivity, and reliability by, e.g., gathering more data from production processes and predicting equipment maintenance. Wireless technology can generally be easier and faster implemented compared to wired networks in an industry environment.

IV. ASSESSMENT OF ARCHITECTURE OPTIONS

Opportunities for centralized or distributed deployment of NFs are mainly driven by scenarios, site constraints and HW requirements. Synchronous radio NFs will have to be placed in the vicinity of the radio baseband processing, whereas, asynchronous NFs may be placed more freely in the network topology. Combining NFV and SDN will enable increased flexibility and, hence, new opportunities for load optimization and resilience by function repositioning according to network load, service requirements or operational reasons. Introduction of flexible functional split between baseband processing and RF units enables adequate adjustment of fronthaul requirements to scenario-specific boundary conditions and, hence, opens more space for C-RAN applications. Nevertheless, it is envisaged that wide area (WA) D-RAN RBS sites will be locations for so-called local centralization of synchronous NFs that are needed for UDNs, D2D, V2X and Dynamic RAN. In that sense, the meaning of D-RAN and C-RAN will change over the time. WA networks (WANs) will evolve in an evolutional manner as LTE and its future enhancements may play an important role. They can be integrated into 5G by means of Multi-RAT enablers. Driver for 5G WAN components could be a ubiquitous need for native integration of D2D and MMC into the WAN air interface.

M-MIMO or UDN, at first glance, may be considered as alternative architecture options aiming at capacity extension. Indeed, they are highly complementary as outdoor UDNs at higher frequencies can apply M-MIMO as preferred means for link budget improvement and interference mitigation. In this context, Multi-RAT behaves highly complementary as it supports mobility and coverage in UDN and M-MIMO scenarios. For outdoor UDNs M-MIMO will play a significant role enabling high spectral efficiency supporting wireless backhauling. As first outdoor UDNs are likely to be realized in cmW bands around 10 GHz, the usage of flexible spectrum management is significant to satisfy operator's spectrum demand.

In 5G networks, Dynamic RAN will play an important role for enhancing the flexibility in terms of adapting its topology to local needs. Even if this architecture option comes across some limitations such as limited usability of customer-owned devices and restricted device capabilities for handover of certain NFs, a big economic impact is expected, due to the fact that 5G networks will much less have to be designed for peak performance that is only needed for a short period in time during a day. Energy management for UDNs will be one of the key factors for sustainable operation. D2D communication and UDNs realizing opportunistic caching may be affordable if with progress of time storage becomes more and more viable. In that case, they will not only provide offloading, but also decrease the peak-to-average capacity demand.

uMTC in industrial applications combines the requirements of URC with the appearance of thousands of devices. At the same time, some sensors may require high mobility whereas others are at hidden places where network coverage is difficult to realize. Challenging in those scenarios from architectural point of view is the concurrency of contradicting requirements that can only be fulfilled by an extremely scalable and flexible network that adopts available resources in terms of computational power to timely changing and location dependent requirements. Multi-RAT, UDN, M-MIMO and Dynamic RAN (including D2D and V2X) may enable those extreme applications.

V. CONCLUSIONS

The presented METIS functional architecture, which is described in more detail in [15], will serve as a first source for identification and development of novel 5G NFs. The envisioned logical orchestration & control architecture of the METIS system based on SDN and NFV principles is providing the necessary flexibility to realize efficient integration and cooperation of NFs according to individual service needs as well as future evolution of cellular and wireless networks. The consideration and assessment of architecture options give a first tangible impression on novel trends in 5G.

ACKNOWLEDGMENT

Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

REFERENCES

[1] METIS Web page: https://www.metis2020.com/.

[2] ICT-317669 METIS, Deliverable D1.1 "Scenarios, requirements and KPIs for 5G mobile and wireless system," April 2013, https://www.metis2020.com/documents/deliverables/.

[3] ICT-317669 METIS, Deliverable D6.2 "Initial report on horizontal topics, first results and 5G system concept," April 2014, https://www.metis2020.com/documents/deliverables/.

[4] ICT-317669 METIS, Deliverable D6.3 "Intermediate system evaluation results," August 2014, https://www.metis2020.com/documents/deliverables/.

[5] P. Rost, et al., "Cloud Technologies for Flexible 5G Radio Access Networks," IEEE Commun. Mag., May 2014.

[6] D. Kreutz, et al., "Software-Defined Networking: A Comprehensive Survey," June 2014, http://arxiv.org/abs/1406.0440 .

[7] C. J. Bernados, et al., "An architecture for software defined wireless networking," IEEE Wireless Commun. Mag., June 2014.

[8] ETSI Industry Network Functions Virtualisation Industry Specification Group (NFV ISG), White Papers, Specifications, etc., http://www.etsi.org/technologies-clusters/technologies/nfv.

[9] 4G Americas, "Bringing Network Function Virtualization to LTE", White Paper, November 2014.

[10] J. Eichinger, et al., "Building a New Multi-Facial Architecture of 5G", IEEE MMTC E-Letter , September 2014.

[11] P. Agyapong, et al., "Design Considerations for a 5G Network Architecture," IEEE Commun. Mag., November 2014.

[12] NGMN Alliance, "5G White Paper - Executive Version," White Paper, December 2014.

[13] Open Networking Foundation (ONF). White Papers, Specifications, etc., https://www.opennetworking.org/index.php.

[14] E. G. Larsson, et al., "Massive MIMO for Next Generation Wireless Systems," IEEE Commun. Mag., February 2014.

[15] ICT-317669 METIS, Deliverable D6.4 "Final report on architecture," February 2015, https://www.metis2020.com/documents/deliverables/.