IEEE COMSOC MMTC E-Letter

Building a new multi-facial Architecture of 5G

Josef Eichinger, Huawei European Research Center, Germany, joseph.eichinger@huawei.com
Ömer Bulakci, Huawei European Research Center, Germany, oemer.bulakci@huawei.com
Gerd Zimmermann, Deutsche Telekom Innovation Laboratories, zimmermanng@telekom.de
Patrick Marsch, Nokia Networks Research, Poland, patrick.marsch@nsn.com
Hugo Tullberg, Ericsson, hugo.tullberg@ericsson.com

1. Introduction

Observing the current market trends, the Information and Communications Technologies (ICT) landscape is going through a radical and accelerating transformation, with a significant impact on economic and societal growth. This transformation is enabled by new trends and emerging concepts, such as fully integrated group communication, “virtual zero latency”, "zero distance” and "fully immersive" services, ubiquitous access to globally connected knowledge and social media, etc. The vision of wireless connectivity or communication anywhere, anytime and between every-*body* and every-*thing* (smart houses, cars, cities, offices etc.) is gaining momentum, as it is expected to render our daily lives easier and more efficient. This momentum will continue to grow, resulting in an increasing demand on wireless connectivity between people, machines, communities, physical objects, processes, content etc., anytime, in flexible, reliable and secure ways. There is an increasing consensus that this will be the beginning of the Fifth Generation (5G) wireless and mobile communications system.

In the decade beyond 2020, it will be necessary to support 1000 times higher mobile data volume per area [1] [2] [3] together with a broad range of diverse requirements imposed by widespread adoption of high-end devices, such as smart phones and tablets, and various wireless communication services (e.g., UHD video streaming or live video chats). Mobile computing based on both end user devices and cloud application computing platforms will further push that trend which is to be addressed within the 5G framework of **Extreme Mobile Broadband (xMBB)** [4]. This clearly requires certain disruptive features with respect to legacy technologies [5], which will go beyond the natural evolution of IMT-Advanced technologies w.r.t. requirements like higher data rates per user and capacity per area as well as a reduced latency

Beside xMBB, the Internet of Things (IoT) with an envisioned massive connectivity of billions of smart devices and the integration of **massive machine type communication (M-MTC or MMC)** [2] requires new solutions to strongly reduce the power dissipation and cost on the device side and to minimize the signaling effort for low data chunks in the wireless network. M-MTC concerns massive deployments of, e.g., low-cost battery-powered sensors and actuators, remote-controlled and remote-read utility meters being placed anywhere in the landscape. In addition to massive numbers of simple devices, 5G will also need to accommodate for the more complex and bandwidth-extensive interactions between smarter devices in private and industrial households, e.g. in the context of remote video processing and object recognition e.g. by robots, which may in sum consume a substantial portion of the overall radio resources. 5G systems must provide vast up- and down-scaling capability, since it is expected that there will be 10-100 times more connected devices per one human user of communications systems [4], either owned by the user himself or by a third party.

A further essential challenge is the requirement to support highly reliable and latency-critical services [3]. **Ultra-reliable MTC** (**U-MTC**) [4] relates to the capability to provide a given service level with a very high probability under a guaranteed end-to-end (E2E) latency. Example applications would be autonomous driving or vehicular communications, where for instance information on a sudden road hazard needs to be propagated to many cars with only a few ms E2E latency, or industrial automation, where in some cases an E2E latency of 0.5ms is needed at a reliability in terms of block error rate of 10^{-9} [6]. Similar requirements apply in remote control, robotics or surgery, where haptic feedback requires E2E latencies on the order of 1 ms. While existing wireless standards may be able to be extended to higher reliabilities at the price of spectral efficiency (e.g. through redundant transmission), it is clear that the E2E latencies stated above require a completely new system architecture with a native support of UR-MTC from the beginning.

This paper is organized as follows. Section 2 provides the background, drivers and the vision of a new 5G Architecture which is flexible and scalable to accommodate the stringent and conflicting requirements of the evolutionary (xMBB) and the emerging services M-MTC and U-MTC as illustrated in Figure 1. Section 3 is devoted to our vision of the new architecture from three different points of view, namely, the **topology**, **functional** and **logical** views. Finally, Section 4 will conclude the paper.

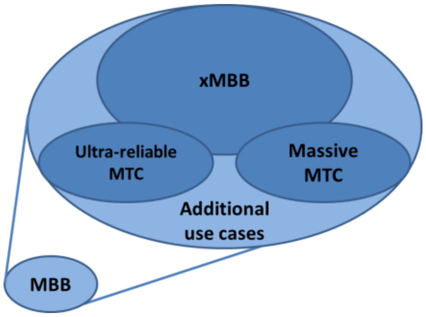

Figure 1: Evolutionary and emerging path of present MBB in 4G to 5G use cases

2. Background and drivers for a new architecture

The traditional definitions of network nodes and assignment of certain functionalities to network nodes limit the achievable flexibility and scalability of the current wireless and mobile communications systems. Accordingly, this conventional approach will disappear in future generations, where the association of network to functionality will be updated or removed based on instantaneous service requirements. Furthermore, user devices will become active network elements. As an active network element, a user device may be the end-point for a service but also may act as an intermediate service point for other devices. This implies that user devices will support a unified mechanism supporting D2D and relaying or self-backhauling with the same approach. Hence, a more generic approach to support all kinds of forwarding of data via wireless interfaces is needed to meet the diverse application requirements while minimize the overall cost, energy dissipation and complexity. The generalized concept of a data forwarding functionality will be independent of network node association, i.e., whether it is deployed in a radio infrastructure node or in a mobile device. For instance, with aforementioned enhanced definitions, radio nodes in vehicles will enable on-demand densification of the radio networks [7].

5G systems will see new connectivity approaches emerge, rethinking the conventional cellular operation. Wireless network elements can have multiple connections to multiple network nodes on a multi-generation and multi-layer perspective. As opposed to current LTE, multiple connections will be supported natively. The connections and the associated functionalities will be activated on-demand based on instantaneous service needs. For example, a minimum connection mode will be tailored for energy optimized operation, whereas, additional connections will be enabled for sporadic usage of high-data rate applications. Such an approach will not only efficiently support the service needs but also enable efficient utilization of the network resources in accordance with device capabilities

There is a clear trend to move typical core network functions closer to the radio edge, e.g., gateway functions and caching, e.g. driven by very tight latency or high reliability requirements. Management of localized traffic flows will further contribute to the reduction in latency. It can be inferred that localized services e.g. for MTC can be managed e.g. by Mobility Management Functions (MMFs) deployed close to the area where the involved devices are in operation. Note that in 5G we may not only see a handover of radio functionality among network nodes, but ultimately also the application layer will be moving along with the entities actually requiring a localized service. It is worth noting that thanks to separation of network node to functionality association (i.e., flexible function split), mobility management functions do not need to be clustered under a certain network node, e.g., MME in LTE, and distributed MMFs can be utilized.

Flexible spectrum usage becomes more important to reuse existing bands and to react *flexibly* to new operator demand and regulatory requirements and finally to apply 5G radio *flexibly* in licensed and unlicensed bands. Rethinking of the carrier aggregation concept of LTE is necessary in combination with flexible spectrum sharing concepts. Thus, it is a must for 5G to deliver mechanisms for a flexible usage of the fragmented spectrum with respect to the different regulatory and regional requirements. The current practice of using dedicated licensed spectrum will remain the main stream, but it will be extended by new regulatory tools and approaches of sharing the spectrum and optimizing its use [4].

3. Dynamic RAN providing a new generation of agile Radio Access Network

As explained above, the 5G system needs to efficiently handle diverse requirements, multiple layers and a variety of air interface parameterizations in the access and the backhaul domains. It further has to control and cope with the dynamics of traffic, user behavior, and active groups of nodes in all levels of deployment. In the sequel, we describe the generic METIS network architecture fulfilling these requirements from the topology, functional and logical point of view [4].

Topological View of 5G Architecture

From the topological point of view, as sketched in Figure 2, the system must accommodate a number of technical enablers and changes in communication paradigms. First of all, the network topology will comprise various flavors of decentralized, centralized or more localized, small scale Cloud-RANs (C-RAN) [8], where the latter for instance resemble multiple small cells connected via fibre or wirelessly to a macro site. It is expected that C-RAN deployments will dominate in the future, but complex stand-alone access nodes will also exist in 5G in areas where infrastructure cost or other constraints do not allow for or justify C-RAN deployments. It is also a trend to distribute the radio processing functions between the centralized and the decentralized processing based on the requirements given by the delay and the bandwidth of the network. Further, the 5G system topology will consist of traditional access nodes as well as new virtual access nodes where classical cell concepts are replaced by more device-centric communication paradigms. Clearly, 5G network topology will be scenario and use-case specific, e.g., depending on whether human-based, M-MTC or U-MTC dominates in a certain area.

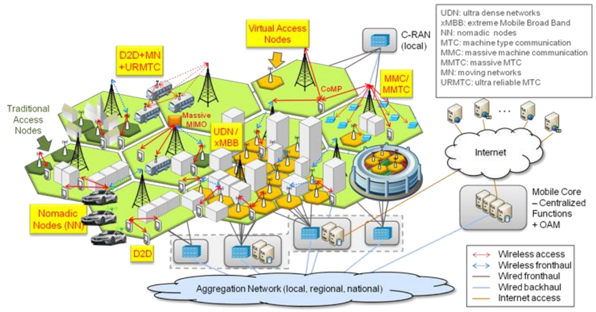

Figure 2: Generic 5G network architecture (high-level topological view)

Further topological features of the agile RAN will be the on-demand activation of radio cells based on fixed or even nomadic nodes [7] on different layers with different coverage areas Also, short range communication between groups of nodes with or without permanent infrastructure support must be supported.

Functional View of 5G Architecture

From the functional point of view, 5G must support a huge set of different radio processing and control functions to manage the requirements of xMBB, M-MTC and U-MTC. There are strong dependencies between the processing steps with respect to delay and required bandwidth for the data exchange for a certain service like U-MTC. Not all of these functions are required anytime and everywhere. Most of these functions will be purely software defined functions. Some others will be implemented in HW accelerators or even distributed among the network nodes. 5G must define an open framework to add new processing or control functions in a cost efficient way. An adoption of SDN [9] and NFV [10] to the demands of RAN will simplify the mapping of the diverse requirements to the available logical and physical distributed resources. It also increases flexibility with respect to integration of decentralized core functions in C-RAN processing units like local MMF, local breakouts as well as Content Delivery Networks with caching capabilities.

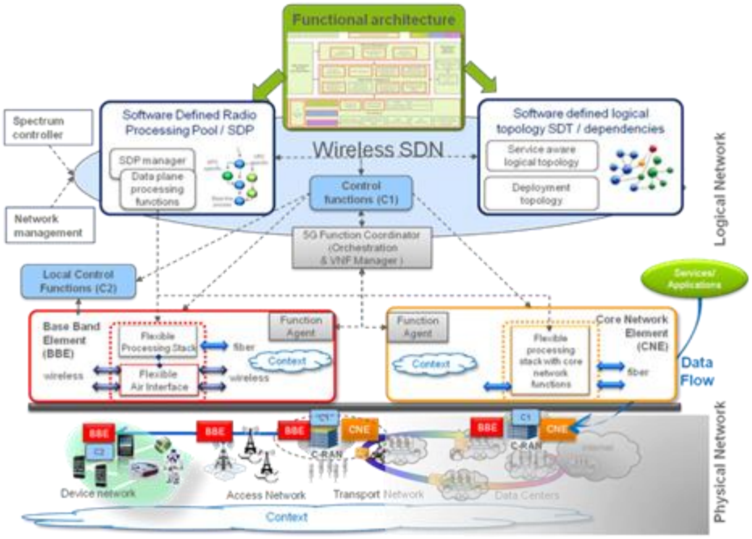

Figure 3: Generic 5G network architecture (high-level functional/logical view)

Logical View of 5G Architecture

From the logical architecture point of view, the METIS architecture foresees only a few different types of nodes. Data processing and forwarding nodes are in charge to process the data with respect to the base band and core network processing schemes applied in order to deliver the data packets with the required latency and reliability to the next node. As shown in **Error! Reference source not found.**, one here differentiates between Base Band processing Elements (BBE) that provide all user plane processing functions with respect to the radio functions and include wireless interfaces, and Core Network Elements (CNE) tailored to core network data processing and forwarding functions. Both logical nodes are connected to the wireless SDN controller, which consists of four main functionalities:

The Software Defined Processing (SDP) manager defines how data plane forwarding and processing functions, all defined in the SDP database, are linked to realize a desired service.

The Software Defined Topology manager (SDT) uses network status information to determine which network elements participate in the data plane operation, their virtual topology and finally the logical data plane topology and the topology of the selected nodes for each service. It also takes into account changes of the physical topology of the networks, for instance in the case of newly activated nomadic nodes [7], involving dynamic self-backhauling, or in the case of newly formed D2D groups.

The 5G Function Coordinator has the task to map the logical data plane topology to physical resources, given the logical data plane topology for each service.
Last but not least, there is a controller connected to the BBE and CNE required to perform e.g. RRM functions. This is foreseen to be split in a logical centralized controller and a second controller responsible for the wireless specific control tasks. As a unique property of the 5G system, the latter edge controller will be able to take over the control of its associated subnet in case of missing or weak network connections.

5. Conclusion

The 5G architecture has to support the delivery of service flows with strongly diverging requirements.
To cope with that issue the architecture will be based on a software-defined networking principle going beyond SDN approaches applied so far for fixed networks (wireless SDN). Only in that way a future-proven network infrastructure can be realized providing sufficient efficiency, scalability and versatility to handle the variety of existing services, to allow fast introduction of new ones, and to bring deployment and operational cost to a low level.
The architecture described in this article is based on few logical nodes for processing of radio-, core- and transport-related functions. These nodes can be flexibly adjusted via NFV orchestration and controlled via SDN layer which can be separated in centralized and localized units with the latter ones for dedicated functions required only in certain wireless environments.

References

[1] Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast, 2013–2018”, February 2014.
[2] Ericsson, “Ericsson Mobility Report: On the pulse of the networked society”, June 2014
[3] ICT-317669 METIS project, “Scenarios, requirements and KPIs for 5G mobile and wireless system”, Deliverable D1.1, May 2013, https://www.metis2020.com/documents/deliverables/
[4] ICT-317669 METIS project, “Intermediate system evaluation results”, Deliverable D6.3, August 2014, https://www.metis2020.com/documents/deliverables/
[5] Stefania Sesia, Issam Toufik, and Matthew Baker, "LTE – The UMTS Long Term Evolution – From Theory to Practice", Second Edition including Release 10 for LTE-Advanced, John Wiley & Sons, 2011, ISBN 978-0-470-66025-6
[6] A. Frotzscher et al., “Requirements and current solutions of wireless communication in industrial automation”, 5G Technologies workshop, ICC 2014, June 2014
[7] Oe. Bulakci, J. Eichinger, Z. Ren, C. Zhou, P. Fertl. S Stanczak “Dynamic Nomadic Node Selection for Performance Enhancement in Composite Fading/Shadowing Environments”, VTC2014-Spring
[8] NGMN Alliance, “Suggestions on potential solutions for C-RAN,” White Paper, January 2013.
[9] D. Kreutz, et al., “Software-Defined Networking: A Comprehensive Survey”, June 2014, http://arxiv.org/abs/1406.0440
[10] ETSI ISG NFV (Operator Group), “Network Functions Virtualisation – Network Operator Perspectives on Industry Progress,” Updated White Paper, October 2013.

Josef Eichinger is Principal Researcher at Huawei's European Research Center in Munich, Germany and received his Dipl.-Ing. degree in electrical engineering from the applied university of Augsburg, and has worked more than 20 years on research of mobile radio systems and the product development. Currently his main focus is on developing solutions for the next generation of wireless communication systems and is member of the EU Project METIS.

Ömer Bulakci received the M.Sc. degree in communications engineering from the Technical University of Munich, Munich, Germany in 2008, and the doctoral degree in communications engineering from the Aalto University School of Electrical Engineering, Espoo, Finland, in 2013. He has contributed to the standardization of relaying in 3GPP LTE Release 10. Since October 2012, he has been conducting research toward 5G mobile and wireless communications networks and contributing to EU project METIS at Huawei European Research Center, Munich, Germany. He is an author or a coauthor of 33 technical publications.

Patrick Marsch is heading a radio research department in Nokia Networks, Wroclaw, Poland. He holds a M.Sc. and Ph.D. degree from Technische Universität Dresden, Germany. He has published 50+ conference and journal papers, received three best paper awards and received the Johann Philipp Reis Prize for pioneering research in the field of Coordinated Multipoint.

Gerd Zimmerman is senior expert in Deutsche Telekom Innovation Laboratories (T-Labs). He received the Dipl.-Ing. and Dr.-Ing. (Ph.D.) degree from University of Kaiserslautern (Germany). Since 1994 he is involved in DT Group in diverse R&D projects with focus on wireless technologies and networks. He is also participating in regulatory and standardization bodies like ITU-R and 3GPP. Since 2009 he is with T-Labs with main activities on LTE and beyond. He is member of IEEE and FP7 Project METIS.

Hugo Tullberg is Senior Researcher at Ericsson Research, Stockholm, Sweden. He holds an M.Sc.E.E degree from Lund University and a Ph.D. degree from University of California at San Diego. He is the technical coordinator of the EU FP7 project METIS.