Bulakci, Ö., et al. Identifying 5G system enhancements: enabling technologies for multi-service networks, in *2018 IEEE Conference on Standards for Communications and Networking (CSCN), 29-31 October 2018, Paris, France [Proceedings].* IEEE, 2018, 5 p.

DOI: https://doi.org/10.1109/CSCN.2018.8581785

Identifying 5G System Enhancements:

Enabling Technologies for Multi-Service Networks

Ömer Bulakci[1], Qing Wei[1], Aravinthan Gopalasingham[2], Antonio De Domenico[3], Mehrdad Shariat[4], Emmanouil Pateromichelakis[1], Fabrizio Moggio[5], Dimitris Tsolkas[6], Christian Mannweiler[7], Marco Gramaglia[8], Stavros Papadopoulos[9], Marcos Rates Crippa[10], Dimitrios Tzovaras[9]
[1]Huawei GRC, Germany; [2]Nokia Bell Labs, France; [3]CEA LETI, France; [4]Samsung, UK; [5]Telecom Italia, Italy; [6]Mobics, Greece;
[7]Nokia Bell Labs, Germany; [8]UC3M, Spain; [9]CERTH, Greece; [10]TUK, Germany

Abstract—**The fifth generation (5G) of mobile and wireless communications networks aims at addressing a diverse set of use cases, services, and applications with a particular focus on enabling new business cases via network slicing. The development of 5G has thus advanced quickly with research projects and standardization efforts resulting in the 5G baseline architecture. Nevertheless, for the realization of native end-to-end (E2E) network slicing, further features and optimizations shall still be introduced. In this paper, we provide a gap analysis of current 5G system (5GS) with respect to some specific enhancements and detail our insights on the enabling innovations that can fill the identified gaps. We will then discuss the essential building blocks and design principles of an evolved 5G baseline architecture capitalizing on the innovations that are being developed.**

Keywords—5G; 5G Enhancements; 5G-MoNArch; Architecture; System Design

I. INTRODUCTION

Since the early research phase of the fifth generation (5G) starting in 2012 [1], the development of concepts for the 5G system (5GS) has progressed at a rapid pace. Within the 5GS, end-to-end (E2E) network slicing spanning over network domains where multiple logical networks corresponding to different business operations, aka verticals, are sharing a common infrastructure, is seen as the fundamental pillar. Diverse and continuously emerging new communication services driven by the verticals require the mobile communication industry to support multiple telecommunications services with heterogeneous key performance indicators (KPIs) in a cost-efficient way. 5G, powered by network virtualization and network slicing, shall give mobile network operators unique opportunities to offer new business models to consumers, enterprises, verticals, and third-party tenants and address such various requirements. Third generation partnership project (3GPP) has already completed the early-drop "non-standalone (NSA)" release of 5G by the end of 2017 [2].

Although all these aforementioned efforts have provided a solid baseline architecture, there is still room for 5G system (5GS) enhancements (5GEs) to better fulfil the 5G vision of supporting diverse service requirements while enabling new business sectors from vertical industries. To this aim, we have performed a thorough 5GS gap analysis, in order to identify the features and optimizations that can be included in the future refinements of the 5G architecture. Such refinements and enhancements can be considered, for example, in the ongoing and future study items and work items of standardization developing organizations (SDOs). In particular, we have found that current baseline architectural work, although understanding the importance of proper slice-coordination schemes, is still not addressing them at full steam.

In this paper, we present an overview of the 5GS gap analysis and how 5G-MoNArch enabling innovations are addressing these gaps. Building upon these innovations, a functional architecture is depicted, as well. Accordingly, this paper aims at providing the main research and development directions that are needed for the realization of 5GS along with its standardization. Further details regarding the overview of the project 5G-MoNArch, including objectives, structure and expected impact can be found in [3].

The rest of the paper is organized as follows. Section II provides an overview of the identified 5GS enhancements along with the associated 5G-MoNArch innovations. Section III describes the overall functional architecture being developed in the 5G-MoNArch project, and Section IV provides a conclusion.

II. 5GS GAP ANALYSIS AND ENABLING TECHNOLOGIES

In [4], we consolidated the views coming from the work of the relevant fora, consortia, SDOs (such as, 3GPP and ETSI), 5G Public Private Partnership (5G PPP) Phase 1 projects along with 5G PPP working groups (WGs) into an overall baseline architecture. Then, after a thorough gap analysis, we identified the possible enhancements (5GEs listed in Table 1) that may be exploited to achieve the 5G objectives faster. The possible enhancements we have identified mostly lie in three main areas: i) a clearly defined set of mechanisms that tackle the problem of coordinating different network slices that are possibly instantiated in the same infrastructure (i.e., 5GEs #2, #3, #6, #8), ii) the lack of a specific protocol stack that natively addresses the requirements and exploits the characteristics of a *cloudified* environment (i.e., 5GEs #1, #2, #3, #4, #5)., and iii) an hands-on experimentation of such protocol stack, to better understand how virtual network functions (VNFs) work in the wild and tailor e.g. orchestration algorithm accordingly (i.e., 5GE #7). In what follow, we present a concise view on enabling innovations and how the technologies within the enabling innovations are utilized to enhance the 5GS·

Table 1. Identified 5GS Enhancements

Enhancement	Description
5GE #1	Inter-dependencies consideration between Network Functions (NFs) co-located in the same node
5GE #2	Orchestration-driven elasticity
5GE #3	Paradigm change from fixed functional operation of small cells toward slice-adaptive operation
5GE #4	Support for computational offloading
5GE #5	Support for telco grade performance (e.g., low latency, high performance, and scalability)
5GE #6	Full support of E2E cross-slice optimisation
5GE #7	Experiment-based E2E resource management for virtual NFs (VNFs)
5GE #8	Enhancement on (radio) resource sharing strategy for network slices

A. *Cloud-enabled Protocol Stack*

One of the key concepts of the 5G-MoNArch architecture is the flexible function decomposition and allocation. This concept builds on the work carried out by the 5G PPP Phase 1 projects [2] which decouples mobile network functions (NFs) from the underlying hardware infrastructure and enables their flexible placement within the physical network. The entire network, comprising edge and core nodes in different locations, thus becomes a large "telco cloud," where NFs can be appropriately located depending on the requirements of the associated service. Building on this concept, 5G-MoNArch will implement a flexible execution platform that builds on ETSI Management and Orchestration (MANO) and Network Function Virtualization (NFV) as enablers for flexible function allocation.

Traditional protocol stacks are not well adapted to allow such a flexible placement of NFs (**5GE #1**). Indeed, placing certain NFs with heavy inter-dependencies in different nodes may incur very high overheads or may simply not be possible. This poses significant constraints on the flexibility of placing NFs, which compromises the overall gains obtained from the flexible function allocation.

The 5G-MoNArch *cloud-enabled protocol stack* will be based on two innovation elements, which we briefly present in the following, the **Telco cloud-aware protocol design** and **Terminal-aware protocol design**.

In 5G-MoNArch, the mapping of NFs to nodes follows a different paradigm: As we will discuss in the following sections, the 5G-MoNArch architecture provides the flexibility to shift NFs to the nodes that better fit the specific requirements of each service (**5GE #2**). As a result of this, NFs that were traditionally co-located in the same node (**5GE #3**) may now be placed on different nodes. More specifically, we aim to design a flexible network that simultaneously supports different protocol splits (for instance, per-user equipment (UE), per-bearer, and per-slice). In addition, the NF deployment can be adapted in time and space depending on the momentary availability of network resources and service requirements.

One example of logical dependencies within the stack is the recursive dependencies between Modulation Coding Scheme (MCS), Segmentation, Scheduling, and radio resource control (RRC). These functions depend on each other and require close synchronization among each other for their operation.

To overcome the problems related to the inter-dependencies between NFs in the protocol stack, one of the key innovations of 5G-MoNArch is the redesign of the RAN protocol stack with the goal of leveraging the benefits of the flexible function decomposition and allocation. Specifically, 5G-MoNArch will focus on the design of **Telco cloud-aware protocol design** adapted to its execution in a cloud environment. With the **Telco cloud-aware protocol design,** the aim is to relax and (as much as possible) remove the logical and temporal dependencies between NFs, with the goal of providing a higher flexibility in their placement. To enable a fully flexible functional split within a cloud-enabled protocol stack, also the role of user terminal needs to be revisited. Such terminal-aware protocol design enables highly distributed deployments that can provide a faster adaptation and reconfiguration of the NFs.

Recently, user terminal has transitioned into more prominent roles in line with new developments in standards. This has enabled device-to-device (D2D) relaying solutions since Release 13 [5] and is currently being further enhanced (feD2D) in Release 15 specifications [6], via emerging concepts like Group Handover (GHO), to improve remote UE reachability, support efficient traffic differentiation, signaling, and service continuity while limiting device complexity and power consumption. In this manner, the contexts of remote and relay UEs can remain collocated in the network. This will increase the time available for handover execution, reduce the risk of handover failure, and result in more accurate resource allocation at the target cell.

The above D2D framework combined with modular NFs provides a platform to further push costly NFs beyond the network edge to save energy or to off-load resource demanding tasks (**5GE #4**). The key is to find optimal balance between centralized NFs and distributed ones towards the edge (based on **Terminal-aware protocol design** as above). This, in turn, may enable faster adaptation and reconfiguration of certain NFs (including mobility) in agreement with cyber foraging concept [7] already established in the literature in the area of pervasive computing (**5GE #5**). In this manner, relay nodes (as network surrogates) act as the last stand of "local" computing beyond the edge. This leads to a truly organic network with self-sufficient sub-networks connected to the umbrella network via single anchor UEs.

B. *Inter-slice Control and Management*

The inter-slice control and management framework defined in 5G-MoNArch is driven by four main enablers, i.e., **Inter-slice control and resource management**, **Inter-slice context management**, **Terminal analytics-driven slice selection/control**, and **inter-slice management and**

orchestration. The main objective of this framework is to support simultaneous operation of multiple network slices, each with tailored access functions and functional placements to meet their target key performance indicators (KPIs). The **inter-slice resource management** is required for the improvement of overall system efficiency, especially on shared infrastructure resources, which is a means for cross-slice optimization. In addition to the slice-adaptive radio resource allocation, slice awareness can be extended to the so-called hard network resources, namely, wireless access nodes, particularly self-backhauled dynamic small cells along with fixed relay nodes for flexibility. A fixed relay can be typically deployed only as fixed radio frequency (RF) amplify & forward (AF) /repeater or layer 3 (L3) decode & forward-(DF) node [8]. In this context, the functional operation of small cell networks is fixed (**5GE #3**) and does not change relative to service requirements or the location of the small cell. Furthermore, each slice can have its own RRC functions and configurations, the uplink/downlink (UL/DL) configuration of the operation of small cells (e.g., time-division duplex, TDD, pattern) may depend on the slice requirements and the deployment of small cells. In this direction, in [9] the development of dynamic adaptation of TDD patterns was proposed; however, this focused mainly on fixed deployments and pre-defined allocation of resources to slices (related to **5GE #8**). Also, with respect to **5GE #3** and **5GE #6**, one main disadvantage of fixed small cells is, thus, the aforementioned lack of flexibility which is essential in 5G systems, where slice-awareness and 5G tight KPIs can necessitate on-demand flexible small cell operation. Hence, 5G-MoNArch considers **slice-aware functional operation within 5GE #3 and 5GE #6** so that 5G RAN can adapt to network changes while fulfilling the requirements of different network slices**.** Accordingly, **inter-slice resource management addressing 5GE #8** can be considered as one component of slice-aware functional operation, e.g., to allocate radio resources to meet slice-specific SLAs.

Moving beyond the current developments in NextGen specifications, 5G-MoNArch also addresses **5GE #6** via **terminal data analytics-driven slice selection or quality of experience (QoS) control** involving UE. This is also in line with utilization of Network Data Analytics (NWDA) as introduced in recent standard developments to optimize mobility decisions. Currently, with regards to **5GE #6**, 3GPP SA2 Release 15 only defines the context processing/sharing between the NFs in the same network slice. The utilization of common context information among network slices depends on the isolation level and the customized NFs per slice. In case of high logical separation parallel operation of control functionalities in these network slices associated to a UE can increase signaling cost along with CP latency. Accordingly, **inter-slice context sharing and management shall be enabled** where the context management mechanisms for cross network slices shall aim: (i) the identification of context information that can be commonly utilized by network slice instances, (ii) the utilization of context information available in one network slice instance to improve the functionalities in another, (iii) the identification and storage of parameters for network slice reconfiguration, management and troubleshooting, (iv) the provision of various context required for enabling slice-specific and inter-slice specific features.

The focus of 5G-MoNArch with respect to **inter-slice management and orchestration** involves:

- The identification of NFs that can be shared between slices
- The identification of functionalities that can be deployed in the (inter slice) control layer (ISC) to monitor and re-configure shared NFs to maximize the overall resource utilization, and thus improving QoS as part of **5GE #2** and **5GE #8** within inter-slice resource management,
- Analyzing and improving the current state-of-the-art software-defined networking (SDN)/NFV management frameworks (e.g., ETSI MANO and ONOS).

According to 3GPP SA5 [10], NFs are grouped into Network Slice Subnets (NSIs) that can be shared among network slices to optimize network deployment and management. Orchestrating NSIs that have shared NFs, according, e.g., to elasticity requirements, implies that the automated management takes care of the requirements of all the involved NSIs that shares those NFs.

5G-MoNArch E2E M&O is thus designed with the capability to be aware of all the involved NSI network requirements when managing a NF that serves different NSIs (within **5GE #2**).

C. Experiment-driven Optimization

Current efforts towards realizing 5GS are lacking valuable information regarding the resource demands and the performance requirements of the VNFs that compose the network (or network slice). This information, acquired through measurement campaigns (i.e., network monitoring processes), can enable the modelling of the VNF behavior regarding their computational, storage and networking resource demands. This modelling may facilitate the overall resource optimization of the cloud-enabled network infrastructure. Algorithms and functions that apply upon the 5G protocol stack can improve their performance by exploiting experiment-driven insights and, thus, taking more intelligent decisions regarding network scaling and elasticity.

Experiment-driven modelling and optimization is a key innovation enabler for the 5G-MoNArch project. The key innovation element of this enabler is the **E2E management of computational, storage and networking resources consumed by VNFs.** This innovation element primarily places the focus on developing experiment-based E2E resource management for VNFs (**5GE #7**). It is worth noting that the experiment-driven optimization is inter-related to all other identified enhancements depicted in Table 1, since it may feed with implementation-related inputs to two other enablers of the project, namely the cloud-enabled protocol stack and inter-slice control, as well as the functional innovations of the project.

The potential of this approach, is better understood if we examine the problem of orchestrating VNFs within a telco cloud. Some of the nodes in the telco cloud (particularly at the edge) may be equipped with limited resources. Thus, the placement of VNFs in nodes needs to consider the availability of computational, storage and networking resources in addition to other criteria such as service requirements, slice awareness, and functional split options. Let's assume that the target is to perform the placement of VNFs of a slice based on the availability of the computational resources. Traditional approaches consider that a fixed amount of computational resources is required for each network function. However, this model is very coarse and clearly insufficient to understand the performance of a real environment in which the computational and traffic load fluctuates significantly over time. To overcome this, accurate models of the computational behavior can be used to determine the performance impact resulting from a given VNF allocation. In particular, new and precise models, extracted from experiments, are needed, to characterize the available resources at each node and the utilization profile of VNFs. Additionally, E2E measurement campaigns are required to feed an analysis on the time-variant behavior of the VNFs, as well as the statistical correlation resulting from logical dependencies between VNFs. Currently, Open Air Interface (OAI) and srsLTE have been recognized as the major open source tools for reliable measurement campaigns. The main experimental results so far are related to the C-RAN architecture. C-RAN demands very high capacity fronthaul solutions (~50 times higher than backhaul) [11] due to the digital transmission of the I/Q samples between Radio Remote Head (RRH) and Baseband Unit (BBU) pool. These results have to be extended beyond the RAN part of the network. E2E software implementations that utilize, in a dynamic manner, behavioral patterns of VNFs' resource demands are needed.

III. 5G-MONARCH SYSTEM ARCHITECTURE

Starting from the need of including the enhancements described in Section II, we next present the initial 5G-MoNArch system architecture. This overall functional architecture also considers the requirements from the project's use cases and the ones initially defined in [12]. As discussed before, the architecture presented in this section has its roots in the results of 5G PPP Phase 1 projects (especially 5G NORMA and METIS-II).

A. 5G-MoNArch overall functional architecture

Fig. 1 depicts the four fundamental layers of the architecture. For each of these layers, a set of architectural elements that deliver the system's functionality, including the key network functions, their responsibilities, the interfaces exposed, and the interactions between them, are defined. The Service layer comprises Business Support Systems (BSS), business-level Policy and Decision functions, and further applications and services operated by a tenant or other external entities. The M&O layer is composed of the M&O functions from different network and technology domains, including, but not limited to, 3GPP network management, virtualization management & orchestration (e.g., ETSI NFV MANO or container management functions), management functions of transport networks (TNs) and private networks. The Network layer comprises the virtualized and physical NFs of both control and user plane, e.g., 5G RAN and CN network functions defined in 3GPP Release 15. The (optional) Controller layer accommodates two controller types: (1) the Cross-slice Controller (XSC) for cross-slice NFs and (2) Intra-slice Controller (ISC) for Intra-slice NFs. In the following subsections, we describe a concise overview of each layer, while for details the readers are referred to [13].

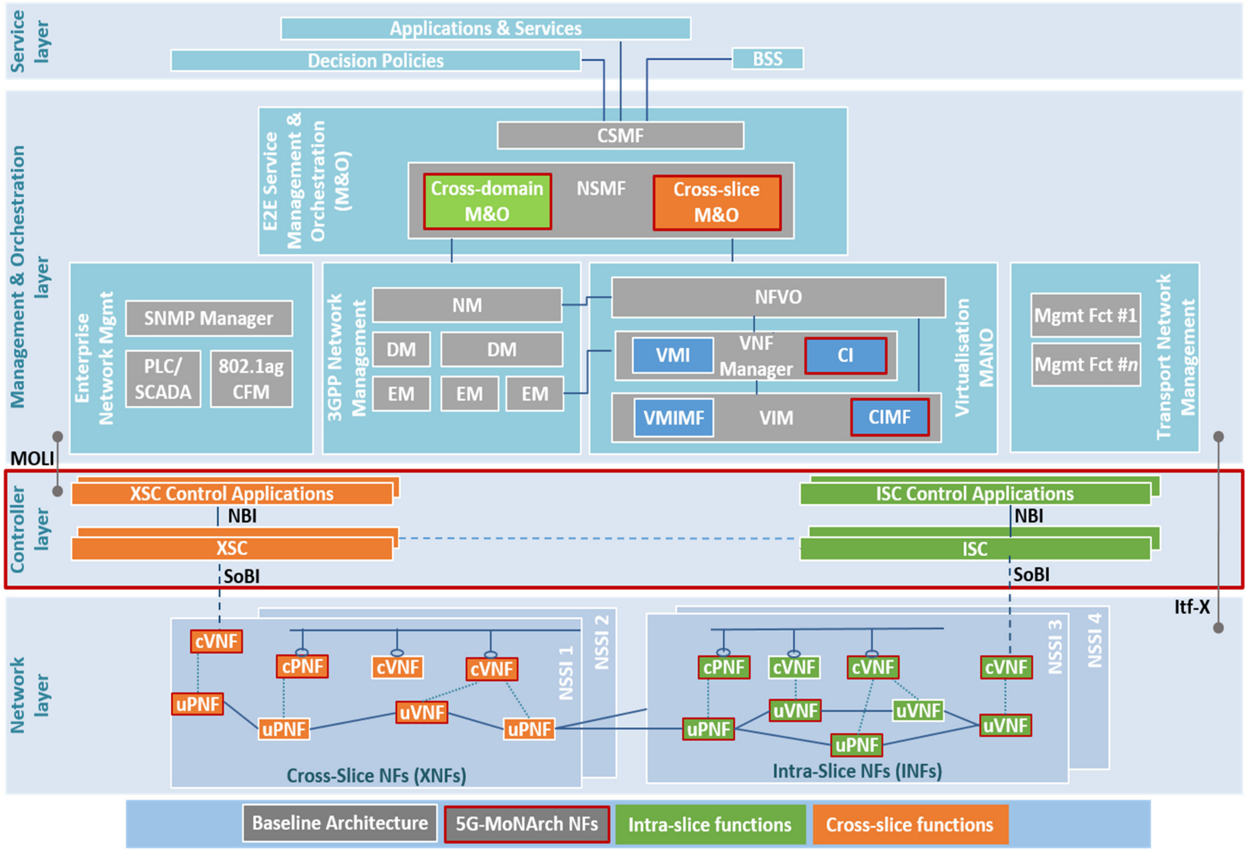

Figure 1. Initial 5G-MoNArch overall functional architecture

B. Management and Orchestration Layer

This layer is substantially the gateway between the different services that could be instantiated by different tenants such as the verticals, and the real network operation. This layer shall support different ways of operation (i.e., Infrastructure as a Service, Platform as a Service or Network Slice as a Service), offering two main functionalities: *i)* the translation of high-level service requirements into real (V)NF deployments by means of end-to-end network slice *blueprinting* and *instantiations*, and *ii)* the efficient, domain-specific lifecycle management and resource orchestration of the network slices running on the shared infrastructure. This layer shall perform two types of slicing management task: Cross-domain MANO for E2E slicing management cross different domains (e.g., 3GPP, non-3GPP, RAN, CN, TN, etc.) and cross-slice MANO for coordinated management across different network slices.

C. Controller Layer

This layer bridges the high-level directives mandated by the M&O layer with the specific VNF configuration that finally compose a network slice. The control functionality follows the principles of network programmability: rapid reconfiguration of VNF is achieved by getting information from both the orchestration layer and the network layer. Within 5G-MoNArch, we envision that this role is played by flexible network controllers that are especially useful for the coordination of different tenants on scarce resources, such as radio resources.

D. Network Layer

The flexibility brought by 'pillar' enablers, such as NFV and SDN, allows the fast instantiation, management and configuration of the, formerly monolithic, now modular network functions that compose a slice. This layer includes both control plane VNFs (cVNFs) and user plane VNFs (uVNFs). Besides the enforcement of quality of experience (QoE) and QoS policies mandated by the M&O layer through the Controller layer, the Network layer also proactively reports monitoring data that can be used together with the information provisioned from applications, management plane, UEs by the Big Data -based algorithms for overall network optimization as well as automated network management.

E. Design Principles

The presented 5G-MoNArch system architecture follows three fundamental design principles: (1) Split of control and user planes, (2) support for E2E network slicing, and (3) network programmability. 5G-MoNArch applies a consistent split of control plane and user plane throughout different network domains, including RAN, core network, and transport network. The architecture allows for different levels of slicing support across different layers. The network functions in the network layer are either slice specific or slice common, and can be controlled by ISC and XSC, respectively. In the management and orchestration layer, the cross domain M&O functions take care of per slice management, and cross slice M&O functions perform the management across different slices. A network slice in the network layer can be flexibly configured/ programmable based on the decision from Controller layer. This enables the network to flexibly adjust its behavior according to the requirements of various use cases in high dynamic environment.

IV. CONCLUSION

The first major milestone for the 5GS has been recently achieved, where the baseline architecture is already specified. Nevertheless, there is still room for new features and further enhancements to cover the envisioned full space of use cases and applications beyond enhanced mobile broadband (eMBB) and to realize a native support for E2E network slicing considering vertical industries associated with new business cases. In this paper, we have highlighted the needed enhancements for further development of the 5GS. Coupled with these enhancements, we have outlined enabling innovations as the main research and development directions. An initial functional architecture is then sketched, where the main building blocks and design principles are presented.

ACKNOWLEDGMENT

This work has been performed in the framework of the H2020 project 5G-MoNArch co-funded by the EU. The views expressed are those of the authors and do not necessarily represent the project. The consortium is not liable for any use that may be made of any of the information contained therein.

REFERENCES

[1] A. Osseiran, et al. "Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS project," IEEE Communications Magazine, 52(5), 2014, pp. 26-35.

[2] 3GPP, "System architecture milestone of 5G Phase 1 is achieved," see http://www.3gpp.org/news-events/3gpp-news/1930-sys_architecture

[3] 5G-MoNArch project, see https://5g-monarch.eu/

[4] 5G-MoNArch Deliverable D2.1, "Baseline architecture based on 5G-PPP Phase 1 results and gap analysis," Oct. 2017.

[5] RP-150441, "Revised WI: Enhanced LTE Device to Device Proximity Services", Release 13.

[6] TR 36.746, "Study on further enhancements to LTE Device to Device (D2D), User Equipment (UE) to network relays for Internet of Things (IoT) and wearables", Release 15.

[7] Yang Cao, Shiyong Yang, Tao Jiang and Daiming Qu, "Performance Optimization for Cyber Foraging Network via Dynamic Spectrum Allocation", IEEE INFOCOM 2011 workshop on Cloud Computing, Apr. 2011.

[8] 3GPP TS 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," v13.3.0, April 2016.

[9] V. Sciancalepore, K. Samdanis, R. Shrivastava, A. Ksentini and X. Costa-Perez, "A service-tailored TDD cell-less architecture," *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, 2016, pp. 1-6.

[10] 3GPP TS 28.531, "Management and orchestration of networks and network slicing," v0.3.0, Feb 2018.

[11] Checko, A., Christiansen, H.L., Yan, Y., Scolari, L., Kardaras, G., Berger, M.S. and Dittmann, L. "Cloud RAN for mobile networks—A technology overview", IEEE Communications surveys & tutorials, 17(1), pp.405-426, 2015.

[12] Next Generation Mobile Networks (NGMN) Alliance, "5G White Paper", Feb. 2015.

[13] 5G-MoNArch Deliverable D2.2, "Initial overall architecture and concepts for enabling innovations," July 2018.