5G Architectural Design Patterns

Kostas Katsalis*, Navid Nikaein*, Eryk Schiller[‡], Romain Favraud*[†], Torsten Ingo Braun[‡]

* {kostas.katsalis, romain.favraud,navid.nikaein}@eurecom.fr

* Eurecom 06410 Biot Sophia-Antipolis, France

[†] DCNS, France

[‡] {schiller,braun}@iam.unibe.ch [‡] University of Bern, Switzerland

Abstract—In this work, we present novel Architectural Design Patterns towards open, cloud-based 5G communications. We provide a brief classification of technologies that cannot be ignored in the design process of 5G systems and illustrate how a new technological added value can be created, when current methodologies, design paradigms, as well as design patterns and their extensions are properly exploited in efficient Radio Access Network (RAN) architectures. We believe that in many cases, the required technology is already there; nevertheless the correct approach has to be worked out and placed within an appropriate context, especially in the case of the integration of complex RAN systems. The enhancements in RF optimization, the progress in cloud computing, Software Defined Networks (SDN) and Network Function Virtualization (NFV), new design concepts such as Network Slicing have to become part of the RAN design methodology. Diverse architectural concepts should break existing stereotypes to pave the way towards the true 5G system integration.

Keywords—5G communications, RAN, Cloud Computing, Network Slicing, SDN/NFV, RAN Design Patterns

I. INTRODUCTION

A great part of research activities, which are related to network and system integration towards a holistic 5G system, is still ahead of us. The reason is that despite the unprecedented advancements in the wireless link capacity, the actual 5G ecosystem contains numerous diverse software and hardware technologies including a multitude of components for different radio access networks. Moreover, the combination of various services requires complex functionality of the system. All these factors have immediate impact on the final performance and actual future modifications of 5G production systems. From the Telecom provider perspective, extreme pressure is put on the existing infrastructure. Traffic demand has dramatically increased, therefore preserving appropriate communications quality, responding to massive traffic volumes, and supporting a number of diverse use cases is a great challenge for future communication networks.

There are numerous aspects (especially in network access) that an efficient 5G design needs to consider towards true system integration. As a first step, particularities of all the major radio access technologies have to be clearly understood (e.g., GSM/GPRS/EDGE, WCDMA/CDMA2000, LTE, LTE-A, Wi-Fi, WiMAX, or generic TDMA/FDMA systems). Radio performance aspects, RF optimization, IP radio access network (RAN) engineering aspects, coverage, wireless capacity/spectrum management, physical and MAC layer functionalities with a fine-grained description and low-level models are required for multi-dimensional analysis of the RAN system design [1], [2]. Although skepticism have been expressed about the efficiency of interference-affected heterogeneous wireless networks (HetNets) [3], a great research attention is focused on the enhancement of each particular technology. Simultaneously, a shift from macro deployments towards wireless heterogeneous networks of macro and small cells and the adoption of Mobile Edge Computing (MEC) [4], seem to aggressively change the structure of new 5G designs and architectures.

There is another 5G access network design challenge related to control plane and user plane operations that appears independently of physical layer characteristics or technology specific optimizations. It appears in classical RAN; Cloud-RAN [5], [6], in which cloud-based BBU is "decoupled" from Remote Radio Heads (RRH); and new eNodeB designs integrating RAN with MEC concepts. We truly believe that a technological breakthrough can be materialized through the deployment of cloud technologies, SDN, NFV and a number of Design Patterns within the Access Network design process (c.f., Fig.1). This is probably the best remedy for various issues related to resource limitations, scalability, rapid service development and deployment, efficient control and enhanced service plane functionality towards the vision of an integrated 5G ecosystem as described by the Telecom operators [7]. In this work, we focus on the control and user plane of future cloudbased, SDN/NFV-enabled RAN designs. We exploit knowledge obtained from existing RAN designs, cloud computing methodologies, experience gained from SDN/NFV application, and state of the art mechanisms, while also from software engineering designs paradigms, like Service Oriented Architecture (SOA) and micro-services, to answer the following questions. How cloud computing technologies can be utilized in the mobile edge? How SDN and NFV design paradigms can be applied in the RAN? How Service Level Agreement driven SLA-Driven RAN methodologies can be applied, when diverge technologies are exploited? What is a Network Slice and how can it be implemented? What are the appropriate communication patterns in today's RAN implementations? Shall we focus on stateful or stateless communication paradigms? These are a few critical issues addressed and covered by this study. Please note that the physical and MAC layer of RAN will be described by means of data-plane functionality and will be abstracted in order to hide technology specific details.

The rest of the paper is organized as follows. In Section II, we describe the Network Slicing concept and a reference 5G architecture. In Section III, we present and analyze the proposed design patterns. Finally, we conclude in Section IV.



Fig. 1. RAN design patterns and evolution towards 5Gs.

II. REFERENCE ARCHITECTURE AND THE CONCEPT OF NETWORK SLICES

Before we proceed with the description of design patterns, we describe the reference 5G architecture and the concept of *Network Slicing*. In the proposed approach, we consider a strong relationship existing between the proposed design patterns and the efficient *Network Slices* operation of cloud-native RANs. Similarly to [8], [9], a *Network Slice* is understood as bundle of network services, functions, network applications, resources, accouterments, etc. (virtual or not) required to cater the the innovation of 5G use cases.

When RAN becomes a cloud-native application, its complexity from the management and control perspectives will be significantly increased. As a consequence, the setup of an endto-end system spanning the whole protocol stack may become very challenging. However, from our point of view, the application of our design patterns will not only improve the quality of RAN systems (incl. their reliability), but will also facilitate the *Network Slices* operation on top of physical infrastructures with virtual resource isolation and virtual network performance guaranties. With the delivery of the Network as a Service and the flexibility needed to provision network resources ondemand, we will be able to tailor network slices to particular business needs.

Note that with the proliferation of cloud-based technologies towards integrated 5G communications, numerous architectures have been already proposed. Such efforts are presented for example in [2], [5], [6], [10] with respect to the SDN/NFV and the cloud computing design paradigms. In our approach, individual elements of other ideas (e.g., the design of the SDN controller, Cloud Orchestrator), can become part of our architecture, as long as they comply with the proposed design patterns and support the *Network Slicing* concept. The design elements of a Network Slices-related architecture were described in [8], [9] and an illustration of the proposed architecture can be found in Fig. 2. For the sake of completeness, we summarize here the proposed layers and their functionality:

Business Layer: The Business Layer supports use-cases that can be provisioned thanks to a marketplace of Virtual

Network Functions (VNF) and Virtual Network Applications (VNAs). An orchestrator creates a Network Slice Manifest that encodes all the details required by the service layer and deploy the service bundle. The business layer interacts with the Service layer. We believe that NFV will be increasingly important for the Telecom Provider operations, nevertheless issues like "What Network Functions are required?", and "How and where do we deploy them?" are open and the field is in a constant state of mutation. Essentially, using the Network Store, one can deliver customized network slice templates tailored to particular use-cases. For example, an LTE network slice can result from several templates, which dynamically install, program, and configure all the LTE network-specific elements that correspond to specific business use-cases (for instance public-safety or low-latency mobile networks).

Service Layer: The Service Layer supports the configuration of the service bundle, while it provides the Slice orchestrator and the Slice Service Manager. It has direct access to real-time network information required by the VNA, and provides network life-cycle service management. In particular, it bundles and chains the VNF under the control of slice service orchestrator, which provides a direct access to real-time VNF information required by the VNA in order to support a number of use cases, with specific quality guarantees.

Infrastructure layer: The infrastructure layer supports the real-time re-configurable cloud ecosystem and virtualization for fast and ultra-fast services (placement, deployment, provisioning). In more detail, the Infrastructure Layer includes the bare metal layer and the virtualization layer that is responsible for the physical resource abstraction to facilitate the infrastructure management. The relevant design elements include, Programmable Computing, Network and Storage Hardware, Programmable RF Hardware and is also related to the Radio Fronthaul Architecture. It also includes the main parts of the Network and Cloud controller systems. Network controller is a SDN-based system whose role is to establish communication between VNFs on-demand. The required communication should also take into account the SLAs of the composed VNAs. For example, if the real-time system requires a link of a given delay, only such a link has to be considered as a communication platform (operating system, socket-based, virtual infrastructure, etc. delays should be taken into consideration as well). Finally, the cloud controller is responsible for matching VNFs to the processing pool. It also needs to interact with the Service Layer and is main responsibility is to take care that all the deployed VNFs meet their required SLAs. For example, the real-time VNFs should be scheduled in the realtime regions of the cloud. The controller should make sure that upon instantiation, the newly deployed and currently existing VNFs will not violate their SLA agreements.

III. RAN DESIGN PATTERNS TOWARDS 5G

The architectural pattern that we propose exposes *Network slicing* as the core concept around which a RAN unified solution will nurture technologies, ideas and mechanisms. These span from virtualization technologies and cloud computing services, to SDN/NFV implementations. In our approach, all the design patterns share the common objective of efficient *Network Slices* operation in cloud-based RANs, while our goal



Fig. 2. The proposed RAN Layered Architecture based on the concept of Network Slicing.

is that their consistent application will further improve the quality and efficiency of existing RAN systems.

In our approach, the Virtual Network Functions (VNFs) and their connectivity are selected based on the unified controlplane knowledge of the target services (based on use-case requirements) and the supported technology features. In more detail, in this work we propose as essential and we analyze the following design patterns for cloud-based 5G communications:

- Distributed Shared Memory
- Dedicated Data Plane Principle
- Shared Control-Plane context
- Agent-based VNF/VNA software
- SLA-driven RAN Designs.

In the following subsection, we describe a set of *Design Principles* that every *Design Pattern* needs to consider. According to these principles every pattern must obey to the service oriented principles, while also must consider best practices in cloud computing design. In addition, it also needs to consider the most effective use (depending on the design) of the SDN/NFV design paradigm with the separation of the Data/Control/Application/Management planes and the use of VNFs to allow independence of the function from the actual hardware.

A. RAN Design Principles

In order to facilitate a modular approach to system-building and omitting the RAN technology specific details and particularities, we believe that the following three *Design Principles* must be carefully utilized by each Access Network design pattern:

1) SOA Design Principles: As it was pointed out by Nadue and Cray in [11], it is not only the decoupling of the Data plane by the Control Plane that led to the SDN wide adoption or the advances in virtualization technology that led to the cloud computing bloom. It is the perpetual application of SOA principles in the network and server segments that drastically changed the way technology experts looked at technology in order to facilitate services offering. Thus in the same way we discuss in the field of SOA about Loose coupling, RAN network Application Programming Interfaces (APIs) must be designed with the minimum dependencies. The same holds for other SOA related design principles like Autonomy for the control over the logic network services and network functions encapsulate and Abstraction, meaning that the services hide implementation logic from the outside world. Indeed, it is the SOA design principles that even transformed the "write once, run anywhere" software development principle to the "configure once-deploy rapidly and run everywhere" NFV principle, with the decoupling of the network functions from proprietary hardware appliances, so they can run in software.

Another SOA design principle that is critical for Access Network designs is Statelessness, meaning that the minimum retaining information specific to an activity must be preserved. Stateless RAN services design is even more important because of the distributed nature of the network architectures now build. In the RAN, as the responsibility for data integrity and continuity resides on the server side (e.g., inside the EPC), by incorporating stateless NFV-based techniques we can facilitate state management and promote agility. On the other hand, in the case where stateful designs are utilized, the benefits from applying the NFV/SDN paradigm are rather blur. An extensive description and classification of SOA design principles can be found in [12]. We believe that the SOA design principles must be evident in every modern RAN design methodology and facilitate service and Virtual Functions lifecycle management, service composition, service chaining and

orchestration procedures.

Micro-services based architecture: In principle, Microservices based architectures are considered as a special "implementation" of SOA, in cases for example where there is no need for Service contracts. Language-agnostic APIs, intelligence in the endpoints, and decentralized control of languages and data are some key characteristics of the approach. Furthermore the principle of fully automated deployment machinery, is of great interest towards facilitating the use of NFV in the RAN.

2) Cloud Design Principles: From our point of view, both the cloud technologies and the relevant cloud patterns must be nurture inside the RAN design methodology, properly adapted to support concepts like Mobile Edge Computing (MEC) [4] and handle issues like user mobility. In [13] a detailed list of available cloud patterns is provided and issues relevant to the RAN design like *Cloud Resource Access Control* or *Dynamic Scalability* are presented.

3) SDN and NFV Design Principles: While programmability and network agility is promoted by applying the SDN paradigm (with new techniques like software Defined MAC, SDN switching inside the EPC, SDN for VNF connectivity etc.) an extra level of complexity is introduced. Furthermore the use of SDN methodologies in programmable wireless networking, in contrast with the wired domain it involves also the User Plane interaction in addition to the programmable wireless data planes and programmable wireless control planes. Nevertheless the tradeoff between increased complexity and programmability when applying SDN principles, significantly tilted the balance in favor of agility that programmability delivers. The same goes for the NFV design paradigm. Although SDN/NFV approaches are not yet standardized, best practices in SDN/NFV methodologies and experience obtained by the application in the wired world is of paramount importance at this stage of developments toward 5G communications.

B. RAN Design Patterns: Analysis

While describing the proposed approach, we provide the necessary components along with current open issues for every design pattern.

1) Distributed Shared Memory: Wireless Heterogeneous Networks (HetNets) and flexible deployments will be a key for future architectures and 5G networks and will rely on several different RANs that use specific data- and controlplane protocols. These RANs will need a very high level of intra- and inter-RAN communication and cooperation to cope with the endless increase of data rates and user mobility. However, current RAN designs are monolithic or organized around monolithic entities communicating with each other in a connection oriented point-to-point manner. Scalability and flexibility are limited, while failure management can be harsh due to the state-fullness and distributed location of these entities. This increase of communication and cooperation needs between the control- and data-planes calls for new design principles of intra- and inter-RAN communications. With applications moving to the cloud, protocols can be re-design, and heavy connection oriented message passing mechanisms should move to lighter solutions. In order to overcome these restrictions and in order to boost performance, the use one-tomany pub-sub type protocols in conjunction with the use of distributed shared memory technologies can be a valuable gear. This kind of protocol, associated with the split of RAN entities into stateless micro-services, will allow for more flexibility and extremely fast data availability.

2) Dedicated Data Plane Principle: The complex 5G ecosystem has to be decomposed into so called micro-services be it Virtual Network Functions (VNFs) or Virtual Network Applications (VNAs) that communicate among each other through implementation-independent interfaces. The granularity of this decomposition may depend on the service (1:1, 1:N, N:1, etc.). Such an organization of the 5G system allows for natural instantiation of 5G within Infrastructure as a Service (IaaS) or Platform as a Service (PaaS) clouds, because in typical cloud orchestrating frameworks (such as OpenStack), the controller installs appropriate virtualized components on physical computing resources. Micro-services can be more easily distributed/migrated, because as tiny components of the 5G infrastructure, they do not require massive processing power or memory consumption, which might not be the case for big mobile networking components of the monolithic architecture.

On the data plane (be it real-time or not-real-time) SDNbased VNF chaining will be maintained to appropriately distribute traffic over VNFs of the 5G system. Network resources isolation (either physical or virtual) is of paramount importance in order to support the concept of Network Slicing. Furthermore, SDN will allow us to efficiently change the configuration to respond on-demand to changing traffic conditions. As the number of VNFs that build the 5G system will be significant, the SDN controller is necessary to deal with a large number of elements. New abstraction models related to 5G mobile functions have to be derived on the SDN controller part. Also, the control SDN-compatible logic is required to be introduced in VNFs on the control plane. In the case of real-time elements, the SDN controller would have to distribute traffic to take into account real-time SLAs of the real-time VNAs. In typical VNFs, however, only standard metrics such as the link capacity will be considered.

3) Shared Control-Plane Context: In the network control plane, although the SDN design approach tries to solve fundamental problems related to network configuration, promote agility and facilitate network consolidation, there are still many challenges to be solved. For example strain still arises from all the complexity created by data generated from northbound, southbound or even west-east interfaces. We believe that the relevant communication models need to be revised in order to reduce the communication overhead and signaling, without sacrificing programmability. We support the concept of Shared Control-Plane Context meaning that the data generated regarding network configuration and operation state must be ad-hoc available without taking into the account the source of data generation (e.g., southbound or northbound interface). The need for domain-specific communication models and ubiquitous context information related to network data is now evident.

The relevant data management actions and data transfer can be further improved taking into the account the way we treat data depending on the context (e.g., mutable data that can change in s or immutable data, access rights etc.). Pubsub communications and distributed share memory concepts can be utilized and compensate the huge investment in SDN in wireless domain. We believe that the "sharing the context" via distributed shared memory approach is an ideal candidate in order to reduce the communication overhead (e.g., LTE control plane communications) and clusters of controllers must be successfully build around the new concept. To get a broad view of the available information, the controller should list running services, available information sources, and information consumers. In the case VNFs operating at the network edge (e.g., MEC applications), very fast information sharing techniques will be required (e.g., RAM memory mapping, etc.), but the final composition of VNFs and their interfaces will be produced by the controller according to the SLA requirements.

4) Agent-based VNF/VNA software: The previously described architecture has to go through a careful testing process that allows us to measure the overall overhead of the microservice architecture of tiny micro-services coupled to communication services binding the 5G ecosystem all-together. In the Network Slicing approach, we put a particular focus on the deployment of real-time cloud regions that will enable the existence of VNAs composed of real-time and non-real-time services. It has a huge impact on the 5G network operators as they can maintain applications, such as RAN, directly in the cloud, allowing for cheap multi-RAT deployments instantiated and optimized for a momentary network use. The ecosystem, which builds upon a multitude of connected micro-services has to maintain a logic to carefully select the placement of VNFs and the choice of communication technologies to meet the required SLAs and provide the connected 5G system structure of good performance. To accomplish this, the abstraction of communicating entities and communication technologies has to be provided.

As we described in the begging of this study, cloudification of the RAN is not just putting a function on a VM or a container. From our point of view a centralized entity that controls all the VNFs is not enough. Following this reasoning, the architecture of a cloud-based 5G system has to introduce a high level of redundancy to overcome failures of the cloud infrastructure. To facilitate the redundant operation, we propose the introduction of a local agent-based stateless VNF. The stateless VNF is materialized through shared memory at the back-end to maintain states. A state can be shared by a bundle of identical replicated VNFs synchronized through shared memory. When one service instance fails, others can immediately replace the failed module to immensely improve the service availability. A local point of control must be embedded inside every VNFs for the RAN (local agents on top of VNFs that are utilized per slice, see Fig. 2), in order to handle time-critical functionalities and facilitate uninterrupted operation.

5) SLA-driven RAN Designs: In the 5G ecosystem, there exist two essential service categories, i.e., real-time and nonreal time. The first category is mostly associated with Radio Access Networks (RANs), in which signal processing has to finish before a given deadline. For example in LTE-FDD, a BBU has to accomplish sub-frame signal processing within 2.3 ms; other functions (EPC, HSS –related) are not deadline critical. Typical clouds, however, do not recognize deadline critical operations and therefore cannot provide realtime support. To properly cope with this problem, the cloud has to be organized into real-time and non-real-time regions. We believe that a deadline critical module, such as an LTE BBU, has to be further decomposed into a real-time (e.g., radio processing chain), and non-real-time (e.g., S1-U GTP encapsulation) VNAs. The real-time part has to be described in terms of appropriate deadline critical SLAs. The cloud scheduler has to take into account the processing delay of every VNF composing the VNA as well as the propagation delay overhead for communication between chained/communicating VNFs (e.g., point-to-point communication such as a BBU-RRH link). When the SLAs can be accomplished on the current infrastructure, the scheduler has to adequately place/chain the VNA elements. Otherwise, an error has to be returned that the service SLAs are not met and the VNA cannot run on this infrastructure. Also the scheduler has to be aware of real-time functions already deployed on real-time cloud servers, because overbooking can severely degrade performance of real-time VNFs and should be avoided. This can be materialized through VNF CPU pining or more sophisticated deadline scheduling with the help of Operating System-based virtualization.

In real-time composed VNAs, the selection of the underlying communication technology has to be therefore based on deadlines, as both communication components and VNFs will account for the delay of a composed VNA. As an example, let us consider an LTE-FDD eNB VNA composed of an RRH VNF, BBU VNF, and an underlying best effort Ethernet link. Even though BBU and RRH were real-time VNFs appropriately maintained in real-time cloud regions, the Ethernet link of unpredictable delay can still seriously violate LTE-FDD processing requirements and degrade the performance of the eNB [8]. As a BBU-RRH communication technology, synchronous Ethernet with Single Root I/O Virtualization (SRIOV) seems to be more appropriate, because low communication deadlines can be maintained for such a link and the overall VNA deadline can hold.

IV. CONCLUSIONS AND FUTURE WORK

In this work, we present architectural design patterns towards open, cloud-based 5G systems. All the proposed patterns center around *Network slicing* as a core concept for unified RAN solutions. We analyze Distributed memory concepts, the dedicated data-plane principle, the shared control plane context, and VNF/VNA design methodologies as a key components of the future 5G system.

ACKNOWLEDGEMENTS

The work has received funding from the European Framework Programme under H2020 grant agreement no 671639 for COHERENT project, "Coordinated control and spectrum management for 5G heterogeneous radio access networks", and under FP7 grant agreement no 318109 for MCN project, "Mobile Cloud Networking" and no 612050 FLEX project, "FIRE LTE testbeds for open Experimentation".

REFERENCES

- C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 122–130, 2014.
- [2] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *Communications Magazine, IEEE*, vol. 52, no. 11, pp. 65–75, 2014.
- [3] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are We Approaching the Fundamental Limits of Wireless Network Densification?" *ArXiv e-prints*, Dec. 2015.
- [4] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal et al., "Mobile-Edge Computing Introductory Technical White Paper," White Paper, Mobile-edge Computing (MEC) industry initiative, 2014.
- [5] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networksa technology overview," *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 1, pp. 405–426, 2015.
- [6] A. De la Oliva, X. Costa Perez, A. Azcorra, A. Di Giglio, F. Cavaliere, D. Tiegelbekkers, J. Lessmann, T. Haustein, A. Mourad, and P. Iovanna, "Xhaul: toward an integrated fronthaul/backhaul architecture in 5g networks," *Wireless Communications, IEEE*, vol. 22, no. 5, pp. 32–40, 2015.

- [7] Alliance, NGMN, "5G white paper," *Next Generation Mobile Networks*, 2015.
- [8] N. Nikaein, E. Schiller, R. Favraud, K. Katsalis, D. Stavropoulos, I. Alyafawi, Z. Zhao, T. Braun, and T. Korakis, "Network Store: Exploring Slicing in Future 5G Networks," in *Proceedings of the* 10th International Workshop on Mobility in the Evolving Internet Architecture, ser. MobiArch '15. New York, NY, USA: ACM, 2015, pp. 8–13. [Online]. Available: http://doi.acm.org/10.1145/2795381.2795390
- [9] N. Nikaein, R. Knopp, L. Gauthier, E. Schiller, T. Braun, D. Pichon, C. Bonnet, F. Kaltenberger, and D. Nussbaum, "Demo: Closer to Cloud-RAN: RAN As a Service," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 193–195. [Online]. Available: http://doi.acm.org/10.1145/2789168.2789178
- [10] J. Liu, T. Zhao, S. Zhou, Y. Cheng, and Z. Niu, "Concert: a cloud-based architecture for next-generation cellular systems," *Wireless Communications, IEEE*, vol. 21, no. 6, pp. 14–22, 2014.
- [11] T. D. Nadeau and K. Gray, SDN: Software Defined Networks. O'Reilly Media, Inc., 2013.
- [12] T. Erl, SOA: Principles of service design. Prentice Hall Upper Saddle River, 2008, vol. 1.
- [13] "Cloud Computing Patterns." [Online]. Available: http://cloudpatterns.