Cloud Technologies for Flexible 5G Radio Access Networks

Peter Rost, Carlos J. Bernardos, Antonio De Domenico, Marco Di Girolamo, Massinissa Lalam, Andreas Maeder, Dario Sabella, and Dirk Wübben

ABSTRACT

The evolution toward 5G mobile networks will be characterized by an increasing number of wireless devices, increasing device and service complexity, and the requirement to access mobile services ubiquitously. Two key enablers will allow the realization of the vision of 5G: very dense deployments and centralized processing. This article discusses the challenges and requirements in the design of 5G mobile networks based on these two key enablers. It discusses how cloud technologies and flexible functionality assignment in radio access networks enable network densification and centralized operation of the radio access network over heterogeneous backhaul networks. The article describes the fundamental concepts, shows how to evolve the 3GPP LTE architecture, and outlines the expected benefits.

INTRODUCTION

The fourth generation (4G) is revolutionizing mobile communications by integrating fixed line and mobile services through all-IP networks. This enables the introduction of new mobile broadband services requiring high data rates and provides high connectivity to more devices. Currently, there is a global discussion on the definition of future 5G networks [1, 2], and the general consensus is that in 5G this development will proceed even further by introducing new and more diverse mobile services delivered not only to devices operated by humans, but to fully automated special-purpose devices (machine-to-machine, M2M) as well. These communication devices will be integrated in any imaginable way into daily use objects such as cars, household appliances, textiles, and health-critical appliances. The increasingly complex scenarios make it more challenging for mobile network operators to manage and operate networks efficiently while providing the demanded quality of experience.

It is unlikely that one standard and one model of network deployment will be able to fit all use cases and scenarios in 2020 and beyond. On the contrary, mobile networks and deployed equipment need to be flexible in order to be optimized for individual scenarios, which may be dynamic in various dimensions such as space and time. Hence, flexibility and scalability become fundamental requirements to allow for the required network adaptation to the needs of the individual services. This requisite for flexibility will have a significant impact on the design of new network architectures, which will also need to operate along with legacy systems.

In this article, we present one way to provide this flexibility by leveraging cloud technology and exploiting it to operate radio access networks (RANs). Cloud technology has already received increasing attention for the deployment of mobile core network functionalities. Operators investigate the possibility of commodity hardware implementations in order to exploit the benefits of cloud technology (e.g., by means of network function virtualization (NFV) and software defined networking (SDN) [3]. However, these approaches have not yet been applied and considered for the RAN, which is the focus of this article. This article explains the challenges and opportunities in exploiting cloud technologies for 5G mobile networks, and presents particular technology examples. It focuses thereby on the novel concept of a RAN as a service (RANaaS) that centralizes flexibly RAN functionality through an open information technology (IT) platform based on a cloud infrastructure [4].

We give an overview of the challenges for 5G networks and why cloud technology will be a key enabler for 5G networks. We introduce a flexible RAN design that leverages the flexibility from cloud technologies and delivers the service diversity as required in 5G mobile networks. Finally, we conclude the article.

CHALLENGES AND KEY ENABLERS

In the following, key enablers to satisfy the 5G demands and their associated challenges are briefly outlined.

Peter Rost and Andreas Maeder are with NEC Laboratories Europe.

Carlos J. Bernardos, is with Universidad Carlos III de Madrid

Antonio De Domenico is with CEA LETI.

Marco Di Girolamo is with the HP Italy Innovation Center.

Massinissa Lalam is with Sagemcom Broadband.

Dario Sabella is with Telecom Italia.

Dirk Wübben is with the University of Bremen.

0163-6804/14/$25.00 © 2014 IEEE

REQUIREMENTS AND DEMANDS

5G networks will face an exponential increase in data traffic caused by different factors [5]:

1. More devices access the Internet and broadband services, including M2M devices.
2. Devices, particularly smartphones, become more powerful. While in the early 2010s a smartphone featured 2D videos and web browsing, high-definition 3D video and real-time interaction will be common in 2020.
3. More diverse and bandwidth-hungry services appear and are used more pervasively.
4. Devices are integrated into more areas of life and industry.
5. Smartphones are used primarily as a gateway (also for other devices) to access services performed in the cloud. This implies that per-user storage and processing requirements will further increase, while per-device capabilities will not increase at the same pace. The gap will have to be filled by communication networks.

Along with the increasing service and application variety, the required diversity of radio access technology characteristics will also increase. While 4G's main driver is ubiquitous mobile broadband, 5G will serve many different purposes with respect to reliability, latency, throughput, data volume, and mobility. The integration of all these characteristics implies a complex system that will be difficult to manage, operate, and adapt to changing demands when using current technologies. Therefore, we believe that 5G will be based on two key enablers in order to be flexible and adaptive enough for the described requirements: ultra-dense deployments that are demand-adaptive, combined with flexible centralized processing, which allows efficient management of an ultra-dense mobile network and enables more flexible dedicated software solutions.

ULTRA-DENSE DEPLOYMENTS

Since 1950 the system throughput of cellular networks has risen by a factor of 1600 simply by increased spatial reuse (i.e., denser networks and smaller cells). In contrast, the per-link throughput "only" saw a 25-factor increase due to new physical layer techniques [6]. Therefore, the use of very dense, low-power, small-cell networks appears to be a promising option to allow future data rate demands to be handled. Ultra-dense deployments exploit two fundamental effects. First, the distance between the radio access point (RAP) and the user is reduced, leading to higher achievable data rates. Second, the spectrum is more efficiently exploited due to the reuse of time-frequency resources across multiple cells. Small cells complement existing macrocellular deployments, which are still required to provide coverage for fast-moving users and in areas with low user density.

The higher the deployment density, the more spatial and temporal load fluctuation can be observed at each RAP. Hence, the probability increases that an individual RAP does not carry any traffic or only low traffic load. In a conventional small cell deployment, a considerable number of sites would consume energy and computational resources under such conditions. This opens the opportunity for more targeted provisioning of data rates, leading to more efficient use of spectral and energy resources.

Ultra-dense deployments exploit two fundamental effects: the distance between the radio access point (RAP) and the user is reduced, leading to higher achievable data rates; and the spectrum is more efficiently exploited due to the reuse of time-frequency resources across multiple cells.

CENTRALIZED PROCESSING

As networks become denser, interference scenarios become more complex due to multi-cell interference. Centralized processing permits the implementation of efficient radio resource management (RRM) algorithms, which allow for radio resource coordination across multiple cells. It also allows optimization of the radio access performance at the signal level, for example, through joint multi-cell processing and intercell interference coordination (ICIC). RRM and ICIC algorithms improve RAN performance by avoiding, cancelling, or exploiting interference between adjacent cells. At the network level, centralized processing is required to orchestrate and optimize ultra-dense networks (e.g., to dynamically adapt to spatial and temporal fluctuations by turning on/off RAPs) by adding spectrum resources and configuring the network to fine tune user data traffic delivery. Furthermore, central resource pools may allow for flexible software deployment. Depending on the actual scenario, different algorithms can be used that are optimized for particular use cases (e.g., based on traffic characteristics, intercell dependencies, or RAN deployments). This also enables the operator to deploy most recent algorithms on a large scale.

Centralized RAN (C-RAN) recently attracted a great deal of attention as one possible way to efficiently centralize computational resources [7]. In C-RAN, multiple sites are connected to a central data center where all the baseband (BB) processing is performed. Radio signals are exchanged over dedicated transmission lines (called fronthaul) between remote radio heads (RRHs) and the data center. At present, only fiber inks are capable of supporting the data rates (e.g., about 10 Gb/s for TD-LTE with 20 MHz bandwidth and eight receive antennas). This need for a high-capacity fronthaul link constitutes the main drawback of C-RAN. Due to the necessity for optical fiber, current C-RAN deployments are characterized by poor flexibility and scalability because only spots with existing fiber access may be chosen, or costly fiber access must be deployed. Hence, there is a trade-off between centralized processing requiring high-capacity fronthaul links, and decentralized processing using traditional backhaul to transport the user and control data to/from the RAPs. In addition, current C-RAN deployments are based on pools of baseband processors, which do not allow flexible and adaptive software deployment, and therefore leave the enormous potential of cloud computing unused.

FLEXIBLE DESIGN FOR ADAPTIVE OPERATION

This section introduces concepts and technologies for a 5G mobile network satisfying the previously discussed requirements. As most 3.5G/4G

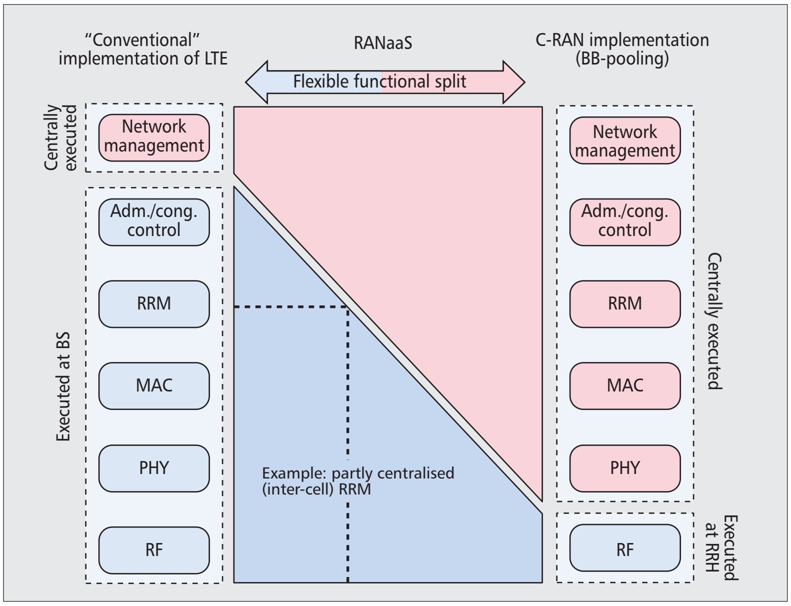

Figure 1. Flexible functional split.

mobile networks are based on Third Generation Partnership Project (3GPP) standards, we use the Long Term Evolution (LTE) technology as our baseline for both network architecture and radio access, and outline an evolutionary path from it.

KEY CONCEPTS

Radio Access Network as a Service — The centralization of processing and management in 5G mobile networks will need to be flexible and adapted to the actual service requirements. This will lead to a trade-off between full centralization, as in C-RAN, and decentralization, as in today's networks. This trade-off is addressed by the novel RAN as a service (RANaaS) concept, which partially centralizes functionalities of the RAN depending on the actual needs as well as network characteristics. RANaaS is an application of the XaaS paradigm [8], stating that any kind of function may be packaged and delivered in the form of a service, possibly centralized inside a cloud platform. This allows exploitation of the increasing data storage and processing capabilities provided by a cloud platform hosted in data centers. The cloud-based design of RANaaS enables flexibility and adaptability from different perspectives:

- Depending on the network connectivity, the RAN is centralized, and the appropriate software functionality is used.
- The actual use cases and current traffic characteristics determine the algorithms that are used and were designed for these use cases.
- The latest software implementations and sophisticated algorithms may be used, which exploit the available resources in a data center more efficiently.

This allows the theoretical limits with respect to system throughput, energy efficiency, or backhaul capability to be reached. This increased degree of flexibility and adaptability will be a key enabler for future 5G networks.

The central element of RANaaS is the flexible functional split of the radio protocol stack between the central RANaaS platform and the local RAPs. This functional split introduces more degrees of freedom in processing design and flexibility in the actual execution of functions, as shown in Fig. 1. The left side exemplifies a traditional LTE implementation where all functionalities up to admission/congestion control are locally implemented at the RAP, that is, at the base station (BS). The right side illustrates the C-RAN approach, where only the radio front-end is locally implemented, and all other functionality is centralized (in this case a RAP is reduced, e.g., to an RRH). In contrast, RANaaS does not fully centralize all RAN functionalities, but centralizes only some of them.

Implementing such a functional split constitutes a serious challenge for the RAN. In theory, such a split may happen on each protocol layer or on the interface between each layer. However, 3GPP LTE implies certain constraints on timing as well as feedback loops between individual protocol layers. Hence, in a deployment with a constrained backhaul, most of the radio protocol stack and RRM are executed locally, while functions with less stringent requirements such as bearer management and load balancing are placed in the RANaaS platform. If a high-capacity backhaul is available, a higher degree of centralization is achieved by shifting lower-layer functions (e.g., parts of the physical, PHY, and medium access control, MAC, layers or scheduling) into the RANaaS platform. Another major challenge is the exploitation of virtualized resources on commodity hardware, which does not provide the same real-time characteristics as currently deployed hardware. This will introduce an additional computational latency and jitter, which needs to be considered in the protocol design. On the other hand, this poses an opportunity as well because algorithms may exploit the possibly large amount of resources efficiently (e.g., through stronger parallelization, and exploiting temporal and spatial fluctuations in ultra-dense 5G networks), which implies an enormous potential to computational diversity.

The following list summarizes major characteristics of a RANaaS implementation similar to the basic characteristics of a cloud-computing platform:

On-demand provisioning of wireless capacity, to deliver mobile communication services more closely adapted to the actual needs of operators and subscribers, which significantly vary in time and space in 5G mobile networks.

Virtualization of RAN resources and functions for optimized usage, management, and scalability with the actual mobile network.

Resource pooling allowing for more advanced network sharing scenarios in which virtual operators offer dedicated services enabling more diverse business opportunities. This is of particular interest in very dense 5G network deployments where the number of deployment options may be limited.

Elasticity by scaling network resources at the central processing entity as well as by scaling the number of active RAPs.

Service metering, allowing operators to sell

RAN operation services (i.e., the central coordination and processing entity as well as usage of RAPs) and to charge the usage of these on a measurable and controllable basis. This will allow for more diverse usage of radio network resources and virtual operator scenarios.

Multi-tenancy, enabling isolation, policy enforcement, and charging of different users of the RANaaS platform (i.e., different service providers). This is of particular interest to ensure security in a 5G mobile networks.

Joint RAN-Backhaul Operation — 5G mobile networks will rely on a very dense small cell layer that needs to be connected to the RANaaS platform. However, small cells may need to be deployed where it is either difficult or too expensive to deploy fixed broadband access or line-of-sight-based microwave solutions for backhaul. Therefore, the backhaul network becomes an even more critical infrastructure part as it needs to connect small cells at different locations. This requires heterogeneous backhaul technologies suitable for different scenarios and use cases. Therefore, limited backhaul resources must be considered when operating the RAN. This will drive the need for co-designing and co-optimizing the RAN and backhaul network through standardized interfaces.

In particular, flexible centralization as implemented through RANaaS will require dynamic adaptation of network routes and the degree of RAN centralization depending on available backhaul resources. Among others, this implies the need for a sophisticated transport network design that can deliver the data toward the central entity independent of the degree of centralization. This is a key requirement in order to allow for maximum flexibility when introducing new functionalities to the network. However, this also complicates routing as well as classification of data packets according to their quality of service. Classical distributed routing algorithms cannot provide this degree of flexibility. In contrast, the use of SDN [3] allows faster reaction to link/node failures, higher utilization of the available resources, and easier and faster deployment of new functionalities or updates, and elastic computation. These advantages mainly result from a centralized control instance that simplifies the configuration and management, and allows for increased computational efforts as individual routing devices no longer constrain the algorithmic complexity.

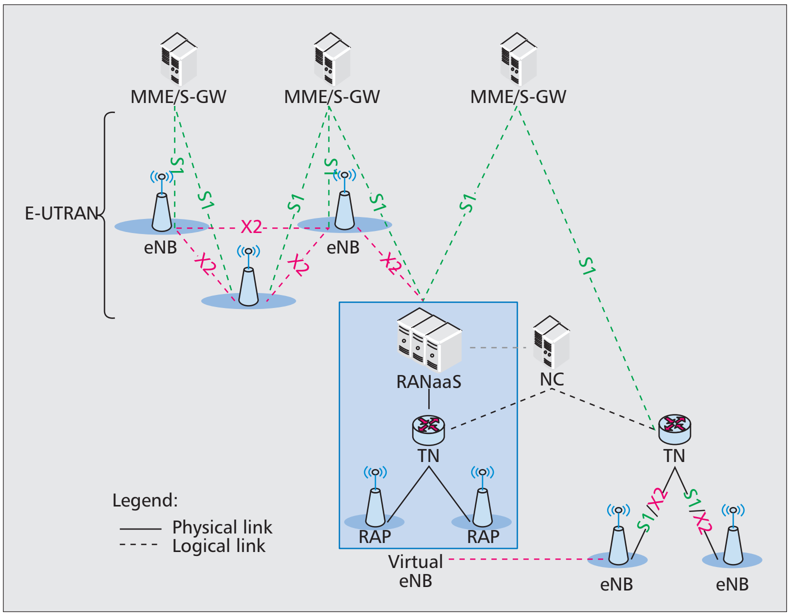

Figure 2. Architecture evolution toward a 5G mobile network.

EVOLUTION TOWARD A FLEXIBLE MOBILE NETWORK ARCHITECTURE

The previously introduced RANaaS concept and joint RAN/backhaul design will affect the mobile network architecture. Nevertheless, for economic reasons, 5G mobile network architectures will most likely be developed as an evolution of LTE Release 12 and beyond. Hence, the introduced concepts need to be as transparent and compatible as possible with the 3GPP network architecture [9] while satisfying operational and customer demands on performance. The mobile network architecture needs:

- To support the (potentially dynamic) flexible centralization of RAN functionality
- To consider criteria such as backhaul and hardware capabilities, traffic demand, and energy efficiency in order to choose an optimal functional split
- To offer a network controlling function that orchestrates and monitors the interaction of functions distributed on different network entities

Figure 2 illustrates the logical network architecture we envision to enable the previously introduced concepts. The combination of RANaaS and one or several RAPs forms a virtual eNB (veNB), which is the functional equivalent of an eNB in the 3GPP architecture (the LTE terminology for a base station) [9]. A *veNB controller function* (veCF) located in the RANaaS platform is responsible for function placement, coherent execution of the distributed functionalities, and the management and configuration of veNB components. The veNB is transparent to the 3GPP architecture because the standard 3GPP interfaces (S1-U, S1-MME, X2) are maintained toward the core network and other (v)eNBs. Data transfer within the veNB domain has to take into account the requirements of different functions and the capabilities of the backhaul linking RANaaS and RAPs. This allows for flexible centralization of RAN functionality dependent on deployment and use cases without affecting 3GPP interfaces or exposing the actual degree of centralization to other network entities.

The SDN-capable backhaul *transport node* (TN) must provide interfaces for exchange of information about backhaul capabilities and available bandwidth that can be used to choose an optimal degree of centralization. TNs are controlled by a *network controller* (NC) for on-demand reconfiguration and path control of the backhaul network in cooperation with the veCF within the RANaaS platform. In order to not expose this information, SDN functionalities will

Centralized functionality	Centralization requirements	Centralization benefits	Challenges
Detection and decoding/ modulation and encoding	Depends on control overhead in UL/DL; Latency req. depends on timing req. in DL; Strong reliability	• Cooperative Tx/Rx • Advanced pre-coding • High computational diversity	• Pre-detection at RAP to reduce backhaul overhead • Separate pre-coding decision and execution at RAP and RANaaS • Optimal quantization of signals and exchange over backhaul
Link reliability protocols (e.g., HARQ)	Depends on entity which performs re-transmission decision	Simplified centralization of scheduling and decoding	• Pre-defined timing of (N)ACK messages • Separation of retransmission decision and packet combining • Strong interaction with other functions, e.g., scheduler, en-/decoder
Scheduling and intercell RRM	Flexible requirements	• Multi-cell gains • Computationally expensive algorithms • Gains depend on backhaul quality	• Scalable latency requirements must be supported • ICIC based on changing quality of channel state information • Variable computational complexity
Segmentation/reassembly	Flexible latency requirements	Medium processing gains	Flexible transport formats required due to possible mismatch with link adaptation
RRC connection handling	Flexible latency requirements	Load balancing in RAN and backhaul	User/data plane split across different RAPs, e.g., macro and small-cells, requiring SDN capabilities
QoS management	Depending on granularity	• Joint QoS management for RAN and backhaul • Multi-cell/user diversity	• Application of QoS management across cells • Application of QoS management based on RAN and backhaul information • Flexible QoS management for backhaul traffic prioritization depending on functional split
Ciphering	Low	Centralized security improves per-RAP security	Real-time requirements need to be satisfied by cloud processor

Table 1. Overview of selected 3GPP LTE radio protocol functionality that may be considered for flexible centralization.

be used to set up the corresponding network route that directs veNB internal interfaces to the actual network entity where it needs to be handled.

FLEXIBLE RADIO ACCESS

The flexible centralization of RAN functionality will impact the operation of the 3GPP LTE RAN protocol stack and may be limited by dependencies within the protocol stack. Table 1 provides an overview of promising functions of the 3GPP LTE radio protocol stack, comprising PHY, MAC, and RRC, which may be considered for partial centralization. In general, the lower we place the functional split within the protocol stack, the higher the overhead and the more stringent the backhaul requirements. Centralizing functionality on the PHY allows for computational diversity, which depends directly on the number of users per RAP. Due to temporal and spatial fluctuations, the computational load can be balanced across cells. Central processing also allows multi-cell algorithms to be implemented to avoid or exploit interference.

On the PHY layer, detection and decoding in the uplink may provide the most significant gains through centralized operation by exploiting global network knowledge and the increased computational resources [10]. Consider joint multi-user detection (MUD), which jointly processes the received signals of several users (UEs) at more than one RAP. Joint MUD can be partitioned into local preprocessing at the RAP, cooperative processing across RAPs, and central processing in the RANaaS platform. An exemplary algorithm for MUD is Multi-Point Turbo Detection (MPTD) [10]. In MPTD, the idea is to schedule edge users attached to different RAPs on the same resources and exploit the interference in each RAP as a source of information through an multi-user turbo detection process [11]. While MPTD fully centralizes the detection, an alternative option is in-network processing (INP), which follows the approach of distributed consensus-based detection by exchanging local variables between neighboring RAPs [12]. As soon as consensus among the RAPs is achieved or a predefined stopping criterion is met, decoding is performed either at

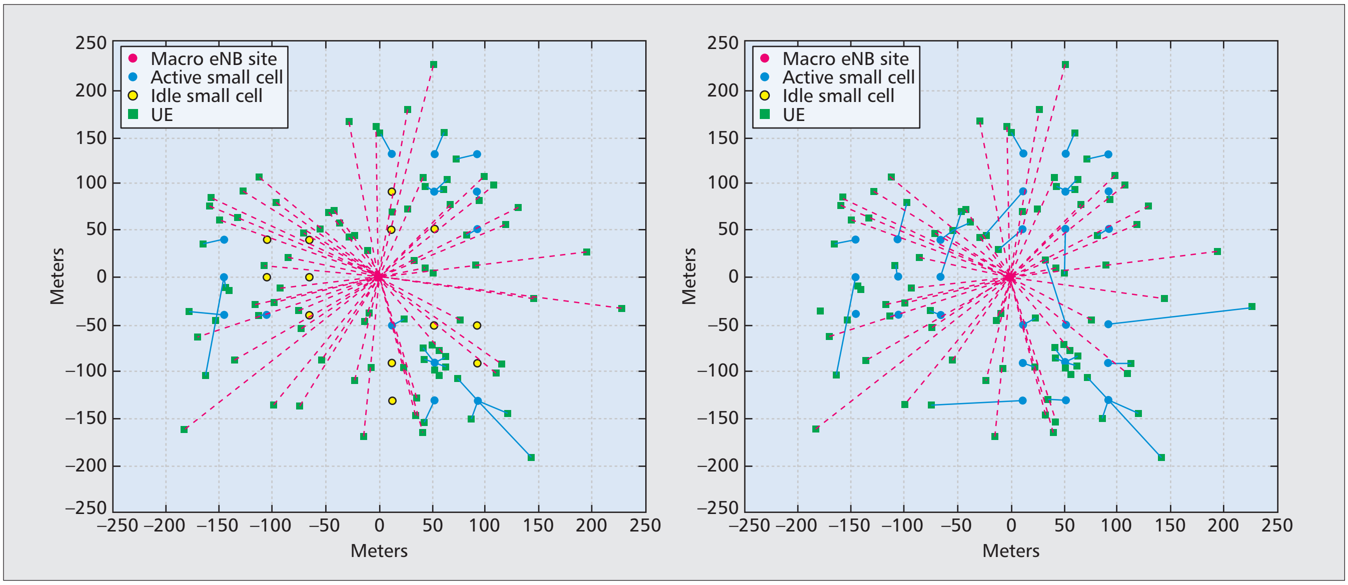

Figure 3. A snapshot of the association pattern when using the strength of the downlink signal (left) and the aggregated network load (right) as the association metric.

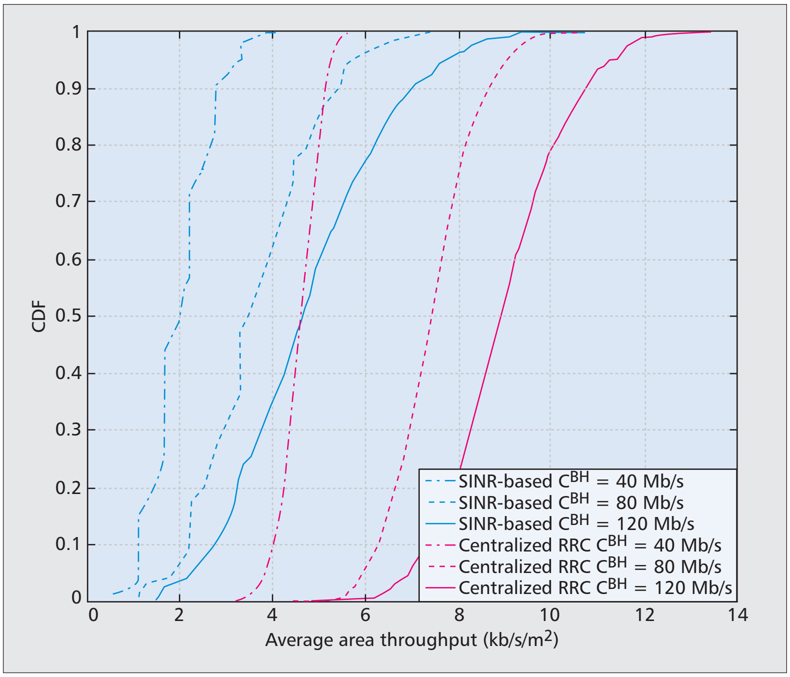

Figure 4. Cumulative distribution function of the average area throughput achieved with the classic signal-to-interference-plus-noise ratio (SINR)-based approach and the proposed centralized RRC solution.

one RAP or within the RANaaS platform. An inherent advantage of this approach is fault tolerance, as broken backhaul links will only affect the number of iterations but not the quality of the final estimate.

Another example is the interface between the PHY and MAC layers, and in particular hybrid automatic repeat request (HARQ), which poses strong timing requirements [13]. In 3GPP LTE frequency-division duplex (FDD), HARQ feedback needs to be sent within 3 ms after receiving the corresponding frame. In a scenario with centralized decoding, this implies that the round-trip delay on the backhaul, including computational latency, needs to be less than 3 ms. In addition, this delay may not be guaranteed due to computational jitter in the RANaaS platform. Therefore, for the considered use cases and deployment scenarios in 5G mobile networks, new signal processing algorithms are required that handle HARQ more efficiently and allow for higher backhaul latency as well as computational jitter.

Further above, on the MAC layer, scheduling and segmentation will particularly introduce challenges to the system design. Scheduling can benefit from centralization through implementing advanced ICIC algorithms with high complexity. However, scheduling is sensitive to imperfect and outdated channel state information, which needs to be taken into account. Both scheduling and segmentation may introduce further constraints on the system as the actual modulation and coding scheme (link adaptation) is selected at the RAP and therefore not known perfectly. Hence, new and adaptive packetizing mechanisms would be required.

Centralized radio resource control (RRC) would enable coordinated traffic steering mechanisms. One example is mobility load balancing, where UEs from overloaded cells are re-assigned to neighboring cells with available resources. Corresponding functions and messages are already defined in 3GPP LTE. Nevertheless, finding the optimal association of UEs and eNBs is difficult because of the large number of possible assignments and the side-effects on resource management. In addition, exploiting knowledge about available backhaul capacity may have a significant impact on overall performance [14]. As shown on the left side of Fig. 3, when using only the strength of the downlink signal as a decision parameter to associate UEs and eNBs, most of the UEs get connected to the central macrocell, and many small cells and related backhaul facilities remain unused (12/27 small cells are idle in this example). On the other

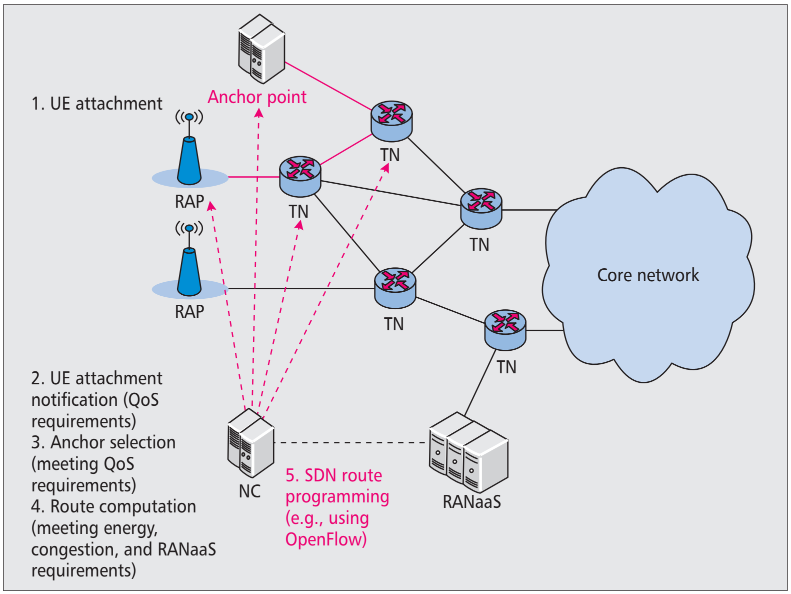

Figure 5. SDN-based backhaul management.

hand, exploiting the available backhaul information increases macrocell offloading by activating all small cells (Fig. 3, right). Figure 4 shows the cumulative distribution function (CDF) of the average area throughput achieved with the SINR-based solution and with the described enhancements with respect to different backhaul capacity constraints (C^{BH}). These results indicate that centralized RRC improves the median value by up to 130 percent compared to solutions that only consider information from the RAN.

Finally, centralization of ciphering offers the possibility to implement more advanced security algorithms and avoid security breaches locally at the RAP, which places a serious security risk in 5G networks where RAPs will be deployed more densely.

The degree of centralization may affect control algorithms and algorithms applied to actual user data differently. Control algorithms such as scheduling, HARQ, and RRC are sensitive to imperfect channel state information and latency on the backhaul. However, many of these algorithms may be divided into time-critical and less time-critical parts. The former part may be decentralized, while the latter is centralized and exploits global network knowledge. Furthermore, the complexity of the former may be rather low, while it is much higher for the latter; for example, a scheduler could be divided into a link-adaptive part, executed locally, and a central more coarse-grained intercell-interference-aware part. This requires algorithms that are not just ported from current deployments to RANaaS, but rather designed for this new network architecture. Furthermore, algorithms operating on user data may not operate under the same stringent timing requirements and could make use of the massive computational resources. This, however, requires algorithms that are dedicated to cloud computing platforms (e.g., exploiting massive parallelization and tolerating computational jitter).

FLEXIBLE RAN AND BACKHAUL NETWORK

As described, 5G backhaul networks need to be more flexible and adaptive to the use cases and actual traffic as well as service characteristics. This triggers the need for efficient network-wide optimizations that offer more degrees of freedom to operate the backhaul depending on RAN parameters, active path management, and topology control in order to provide the correct network for 3GPP interfaces depending on the actual degree of centralization.

Distributed mechanisms struggle with the aforementioned situations because of their convergence time (which is orders of magnitude longer than what is required) and their often lower robustness to identifying a global optimum. The simplified view of the network fabric enabled by SDN simplifies the operation of the network, and allows higher utilization to be achieved by adopting a centralized traffic management approach. Therefore, we adopt a logically centralized architecture following an SDN approach for flexible management of the RAN and backhaul network. This approach comprises an SDN controller, which programs the network entities under its control and dynamically changes the network behavior. It is implemented as part of the NC and provides the required communication metrics for the functional split. Supported by network-wide knowledge at a central entity (the NC in Fig. 2), load can be distributed optimally within small cell networks. The NC has an accurate and up-to-date view of the network status, and is therefore capable of optimally orchestrating the network resources and enabling advanced approaches for:

- **Mobility management:** Denser networks imply more frequent handovers due to the cell size. Hence, mobility management may no longer be exclusively triggered by radio quality, but also by network management decisions. An SDN-based approach allows for shorter service disruption time and switching costs while enabling effective load balancing.
- **Distributed anchoring and local break-out support:** The current centralized 3GPP architectures cause high traffic demands in the operators' core networks. Based on an SDN approach, the user data plane can be distributed to allow local offloading of user data traffic. On the other hand, the control plane remains logically centralized in the NC to allow for globally optimized operation.
- **Energy optimization of the RAN and backhaul:** Depending on user demand and network status, the NC may jointly switch off parts of the RAN and backhaul to reduce energy consumption.

The use of SDN in a 5G network also poses challenges. First, it introduces overhead by flow control programming, which requires careful design of the traffic management algorithms. Second, it is a nontrivial decision to select which functionality is offloaded to the controller and what is still executed on the network devices. Third, multiple controllers should be provisioned, which requires mechanisms to partition the network, and allow controller selection as well as the required scalability and reliability.

Figure 5 shows how an SDN-based approach may operate using the UE attachment process as an example. When a terminal attaches to the network (step 1), it indicates the quality of service (QoS) it requires (or the application that is run, step 2). Based on the required QoS, the NC first needs to select an anchor point that meets the requirements (step 3). Based on the selected anchor point, required QoS, and current network status, the NC determines the optimal route, taking into account energy consumption in the RAN and backhaul, congestion, and requirements of the veNB (step 4). Finally, after the route has been computed, it is programmed within the network through an interface between the NC and the involved TNs, for example, using OpenFlow [15] or extensions of it (step 5).

CONCLUSIONS

This article discusses the novel RANaaS concept, which leverages cloud technologies to implement a flexible functional split in 5G mobile networks enabling optimized usage of spectral, energy, and computational resources in ultra-dense deployments. We discuss an architectural evolution from 3GPP LTE, outline challenges and potential technologies to implement this functional split, and describe the potential gains. Implementing RANaaS will allow for more flexibility of RAN deployments under homogeneous and heterogeneous backhaul. By taking into account the changing service requirements of 5G mobile networks, the RANaaS approach has been defined as a flexible evolution of 4G networks such as 3GPP LTE, which is able to integrate and support a multitude of radio access technologies, services, and deployment strategies.

By taking into account the changing service requirements of 5G mobile networks, the RANaaS approach has been defined as a flexible evolution of 4G networks such as 3GPP LTE, which is able to integrate and support a multitude of radio access technologies, services, and deployment strategies.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 317941 — project iJOIN, http://www.ict-ijoin.eu. The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the view of its authors only. We gratefully recognize the great contributions of many colleagues from iJOIN, who, in fruitful cooperation, contributed with valuable insight, surveys, and vision.

REFERENCES

[1] G. P. Fettweis, "A 5G Wireless Communications Vision," *Microwave J.*, Dec. 2012.

[2] R. Baldemair *et al.*, "Evolving Wireless Communications: Addressing the Challenges and Expectations of the Future," *IEEE Vehic. Tech. Mag.*, vol. 8, no. 1, Mar. 2013, pp. 24–30.

[3] ONF, "Software-Defined Networking: The New Norm for Networks," White Paper, Apr. 2011, https://www.opennetworking.org/images/stories/downloads/white-papers/wp-sdn-newnorm.pdf, accessed Aug. 2013.

[4] D. Sabella *et al.*, "RAN as a Service: Challenges of Designing a Flexible RAN Architecture in a Cloud-Based Heterogeneous Mobile Network," 2013 Future Networks Summit, July 2013, Lisbon, Portugal.

[5] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2012–2017," White Paper, May 2013, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, accessed Aug. 2013.

[6] M. Dohler *et al.*, "Is The PHY Layer Dead?," *IEEE Commun. Mag.*, vol. 49, no. 4, Apr. 2011, pp. 159–65.

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

[8] NIST, P. Mell and T. Grance, "The NIST Definition of Cloud Computing," Sept. 2011, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf, accessed Aug. 2013.

[9] 3GPP TS 36.300 (v11.6.0), "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description; Stage 2 (Release 11)," July 2013.

[10] INFSO-ICT-317941 iJOIN, "D2.1: State-of-the-Art of and Promising Candidates for PHY Layer Approaches on Access and Backhaul Network," http://www.ict-ijoin.eu/wp-content/uploads/2014/01/D2.1.pdf, accessed Jan. 2014.

[11] G. Caire and R. Müller, "The Optimal Received Power Distribution of IC-Based Iterative Multiuser Joint Decoders," *Proc. 39th Annual Allerton Conf. Commun., Control and Computing*, Monticello, IL, Oct. 2001.

[12] H. Paul *et al.*, "In-Network Processing for Small Cell Cooperation in Dense Networks," IEEE VTC-Fall 2013 Wksp. (CLEEN 2013), Las Vegas, NV, Sept. 2013.

[13] 3GPP TS 36.213 (v11.3.0), "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures (Release 11)," June 2013.

[14] A. De Domenico, V. Savin, and D. Ktenas, "A Backhaul-Aware Cell Selection Algorithm for Heterogeneous Cellular Networks," IEEE 24th Int'l. Symp. Personal Indoor and Mobile Radio Commun., London, U.K., Sept. 2013.

[15] ONF, "OpenFlow Switch Specification," Apr. 2013, https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.2.pdf, accessed Aug. 2013.

BIOGRAPHIES

PETER ROST [M] (peter.rost@neclab.eu) received his Ph.D. degree from Technische Universität Dresden, Germany, in 2009 (supervised by Prof. G. Fettweis) and his M.Sc. degree from the University of Stuttgart, Germany, in 2005. Since April 2010 he has been member of the Wireless and Backhaul Networks group at NEC Laboratories Europe, where he is a senior researcher in business unit projects, 3GPP RAN2, and the EU FP7 project iJOIN, which he currently manages as Technical Manager (www.ict-ijoin.eu). He is a member of IEEE ComSoc GITC, IEEE Online GreenComm Steering Committee, and the Executive Editorial Committee of *IEEE Transactions of Wireless Communications*.

CARLOS JE. BERNARDOS (cjbc@it.uc3m.es) received a telecommunication engineering degree in 2003, and a Ph.D. in telematics in 2006, both from Universidad Carlos III de Madrid, where he worked as a research and teaching assistant from 2003 to 2008, and since then has worked as an associate professor. His current work focuses on mobility in heterogeneous wireless networks. He has published over 50 journal and conference papers, and he is an active contributor to the IETF.

ANTONIO DE DOMENICO (antonio.de-domenico@cea.fr) received his M.Sc. and Ph.D. degrees in telecommunication engineering in 2008 and 2012 from the University of Rome "La Sapienza" and the University of Grenoble, respectively. Since 2009, he has worked with the CEA-LETI – MINATEC, Grenoble, France, as a research engineer. Currently, his research topics are cloud-enabled heterogeneous wireless networks, millimetric-wave-based communications, machine learning, and green communications. He is the main inventor or co-inventor of seven patents.

MARCO DI GIROLAMO (marco.digirolamo@hp.com) graduated in electronics engineering summa cum laude in November 1986. He started his professional career as an R&D engineer for defense telecommunication systems; then in 1993 he moved to Hewlett-Packard as a designer and developer of hardcopy devices. Since 2002, he is leading HP Italy activities in collaborative research and innovation programs, and has since then contributed to multiple FP6 and FP7 projects. He is currently involved in iJOIN as a cloud computing domain expert.

MASSINISSA LALAM (massinissa.lalam@sagemcom.com) received his Master's degree from the Institut National Polytechnique de Grenoble (ENSIMAG, INPG) in 2002. He received his Ph.D from Telecom Bretagne (France) in 2006. He is now an Expert in Sagemcom (France), where he works on wireless (WiFi) and cellular (3G/LTE) technologies. His expertise includes link and system-level performance evaluation, heterogeneous networks, network modeling, and radio resource management. He is author or co-author of 15+ international publications.

ANDREAS MAEDER (andreas.maeder@neclab.eu) is a senior researcher at NEC Laboratories Europe, Heidelberg, Germany. He received his Ph.D. from the University of Würzburg, Germany. Currently, his main area of research is the convergence of IT and telecommunication technologies. He has been contributing to the standardization of broadband wireless access technologies in IEEE and 3GPP since 2008. He was Rapporteur of the work item on user plane congestion management in 3GPP SA2. He was Chair of the IEEE BWA workshop 2013, and is an author of numerous scientific articles, conference papers, and patents.

DARIO SABELLA (dario.sabella@telecomitalia.it), after graduating in engineering and a post-degree specialization Master's in telecommunications, joined the R&D center of Telecom Italia, involved in several projects targeting future generation communications. Currently, he is working in the Wireless Innovation Department responsible for some international projects on mobile networks. Since 2012 he has led Telecom Italia energy efficiency activities related to the LTE auction in Italy. He is a co-author of several publications and patents in the field of wireless communications and energy efficiency.

DIRK WÜBBEN (wuebben@ant.uni-bremen.de) received his Dipl.-Ing. (FH) degree in electrical engineering from the University of Applied Science, Münster, Germany, in 1998, and his Dipl.-Ing. (Uni) and Dr.-Ing. degrees in electrical engineering from the University of Bremen, Germany, in 2000 and 2005, respectively. In 2001, he joined the Department of Communications Engineering, University of Bremen, Germany, where he is currently a senior researcher and lecturer. His research interests include wireless communications, signal processing, cooperative communication systems, and channel coding.