Mobile Network Architecture Evolution toward 5G

Peter Rost, Albert Banchs, Ignacio Berberana, Markus Breitbach, Mark Doll, Heinz Droste, Christian Mannweiler, Miguel A. Puente, Konstantinos Samdanis, and Bessem Sayadi

The authors discuss 3GPP EPS mobile network evolution as a whole, analyzing specific architecture properties that are critical in future 3GPP EPS releases. In particular, they discuss the evolution toward a "network of functions," network slicing, and software-defined mobile network control, management, and orchestration.

Abstract

As a chain is as strong as its weakest element, so are the efficiency, flexibility, and robustness of a mobile network, which relies on a range of different functional elements and mechanisms. Indeed, the mobile network architecture needs particular attention when discussing the evolution of 3GPP EPS because it is the architecture that integrates the many different future technologies into one mobile network. This article discusses 3GPP EPS mobile network evolution as a whole, analyzing specific architecture properties that are critical in future 3GPP EPS releases. In particular, this article discusses the evolution toward a "network of functions," network slicing, and software-defined mobile network control, management, and orchestration. Furthermore, the roadmap for the future evolution of 3GPP EPS and its technology components is detailed and relevant standards defining organizations are listed.

Introduction

The Third Generation Partnership Project (3GPP) evolved packet system (EPS) of Long Term Evolution (LTE) refers to the logical architecture composed of the radio access network (RAN), called the evolved universal terrestrial radio access network (E-UTRAN) in the case of LTE, and the evolved packet core (EPC) as defined in [1, 2] and illustrated in Fig. 1. The objective of this logical architecture is to enable a flat IP-based network and provide a standardized set of network elements and network interfaces. Standardized elements and interfaces enable operators to integrate equipment and implementations from different vendors into a single system, while ensuring interoperability. The design of a logical architecture satisfies requirements originating from use cases that are expected to be of particular interest for 3GPP EPS. So far, the aim of 3GPP EPS has been mainly the provision of mobile broadband service, for which the system makes very efficient use of available spectrum.

So far, past releases (i.e., Rel-11, Re-12, and Rel-13) studied and specified how to integrate further services such as small data services as well as machine type communication (MTC) services. Meanwhile, cloud computing technologies and cloud concepts have gained momentum not only from the information technology (IT) perspective, but also within the telecom world. Integrating cloud concepts into 3GPP EPS allows support for novel and emerging services. On the other hand, it requires novel architectural concepts, which natively support cloud technologies. However, the static assignment of functionality to network elements and the strong functional dependencies within each network element make it difficult to support the required flexibility of future 3GPP EPS deployments.

The following sections detail concepts that could contribute to the evolution of 3GPP EPS in order to provide the required flexibility for supporting network services with diverse requirements, to enable diverse mobile networks deployments, and to provide a higher degree of context awareness. Specifically, the next section introduces relevant concepts such as flexible function composition, network slicing, and software-defined network control. After that we provide an overview of the standardization roadmap, and the article concludes in the final section.

Mobile Network Evolution

In order to support diverse services such as eHealth, the Internet of Things (IoT), and vehicular-to-everything (V2X) in future mobile networks, we see a need for enhancing the EPS toward a flexible mobile network accommodating novel architectural principles while maintaining backward compatibility. Such an evolved EPS architecture must support legacy radio technologies as well as novel radio access interfaces such as millimeter-wave (mmWave) or centimeter-wave transmission. It should accommodate emerging processing paradigms such as mobile edge computing (MEC) and cloud-RAN (C-RAN), while enabling flexible deployment patterns based on small, micro, and macrocells and allowing programmability to support very different requirements in terms of latency, robustness, and throughput.

Based on this, we see two main objectives

Peter Rost and Christian Mannweiler are with Nokia Networks; Albert Banchs is with IMDEA Networks; Ignacio Berberana is with Telefonia I+D; Markus Breitbach and Heinz Droste are with Deutsche Telekom; Mark Doll was with Alcatel Lucent and is now with Nokia; Miguel A. Puente is with ATOS; Konstantinos Samdanis is with NEC Europe Labs, UK; Bessem Sayadi was with Alcatel Lucent, France, and is now with Nokia.

0163-6804/16/$25.00 © 2016 IEEE

that must be addressed by an evolved 3GPP EPS architecture.

Multi-service and context-aware adaptation of the mobile network, which implies that the mobile network needs to adopt its operation based on the actual service requirements and the related context. The context includes deployment properties, transport network properties, and service properties, as well as available RAN technologies.

Mobile network multi-tenancy, which aims to reduce capital and operational costs by allowing infrastructure providers to make the best use of available resources, including spectrum and infrastructure. Hence, multiple tenants may share resources within the mobile network while offering diverse services.

In order to achieve these objectives, the following main functionalities should be supported and will be further detailed in the following sections.

Network of functions: Traditionally, mobile network functions are readily grouped into network entities, each responsible for a predefined set of functions, and interfaces connecting these entities. Using a flexible "network of functions" allows adaptation to diverse services, and optimization using different software rather than using different parameterizations. Each block may be replaceable and could be individually instantiated for each logical network running on the same infrastructure. However, it must not imply a multitude of interfaces, as detailed later.

Network slicing allows the same mobile network infrastructure to be used by multiple different operators, including vertical market players, each implementing its own logical network, for example, a logical network for mobile broadband with very high throughput, a logical network connecting a massive amount of sensor nodes (including indoors), or a logical network providing critical infrastructure connectivity for traffic management or energy control. Hence, each network slice fulfills different requirements and serves very different purposes.

Software-defined mobile network control is required to flexibly control both a flexible network of functions as well as a set of network slices. This control must be programmable in order to adapt the network behavior to the current requirements. This functionality goes beyond the separation of the control and data planes, including the control of RAN functionality as well as the mobile network control plane.

Network of Functions

The objective of a mobile network architecture is to allow for integrating different technologies and enabling different use cases. Due to the partly conflicting requirements, it is necessary to use the right functionality at the right place and time within the network. In order to provide this flexibility, it has recently been discussed whether the network functions virtualization (NFV) paradigm should be adopted in the mobile access network domain, that is, enabling mobile network functionality to be decomposed into smaller function blocks that are flexibly instantiated.

So far, the degrees of freedom for assigning network functionality to network entities is very limited. For instance, it is possible to collocate EPC elements, such as gateways, with a base station in 3GPP EPS. However, it is not possible to only place parts of the functionality of a gateway or mobility management entity (MME) with a base station. Similarly, it is possible to fully centralize RAN functionality using the common public radio interface (CPRI) and central baseband units. However, such deployments use non-virtualized baseband units at the central location; hence, it is rather relocating functionality that does not exploit all characteristics of cloud computing. It is further not possible to only move parts of the RAN functionality except in a proprietary way [3, 4].

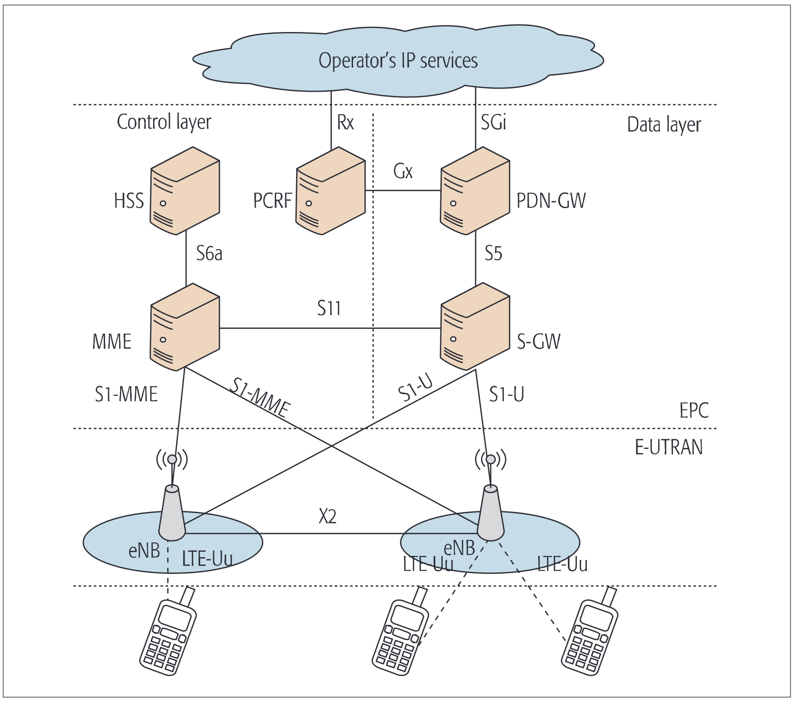

Figure 1. The (basic) 3GPP evolved packet system.

The decomposition of the mobile network functionality would imply a stronger decoupling of logical and physical architecture than in 3GPP EPS as illustrated in Fig. 2, that is, physical network functions (PNFs) may be executed on bare metal, while virtual network functions (VNFs) may be executed on local or remote data centers (referred to as edge and central cloud in Fig. 2). Bare metal refers in this case to the non-virtualized access to radio access resources, for example, through digital signal processors (DSPs), rather than on cloud computing platforms. Hence, depending on the use case, requirements, and the physical properties of the existing deployment, mobile network functionality is executed at different entities within the network. This imposes a number of challenges; for example, the system itself must not become more complex, and the introduction of new interfaces should be avoided as much as possible. Hence, the VNF assignment should exploit an efficient control and orchestration plane as further described below. Furthermore, the coexistence of different use cases and services would imply the need to use different VNF allocations within the network. This is further elaborated later

Network slicing is centered on the concept of deploying multiple dedicated logical mobile networks with varying levels of mutual isolation on top of the same infrastructure. A network slice is a collection of mobile network functions and a specific set of radio access technologies necessary to operate an end-to-end logical mobile network.

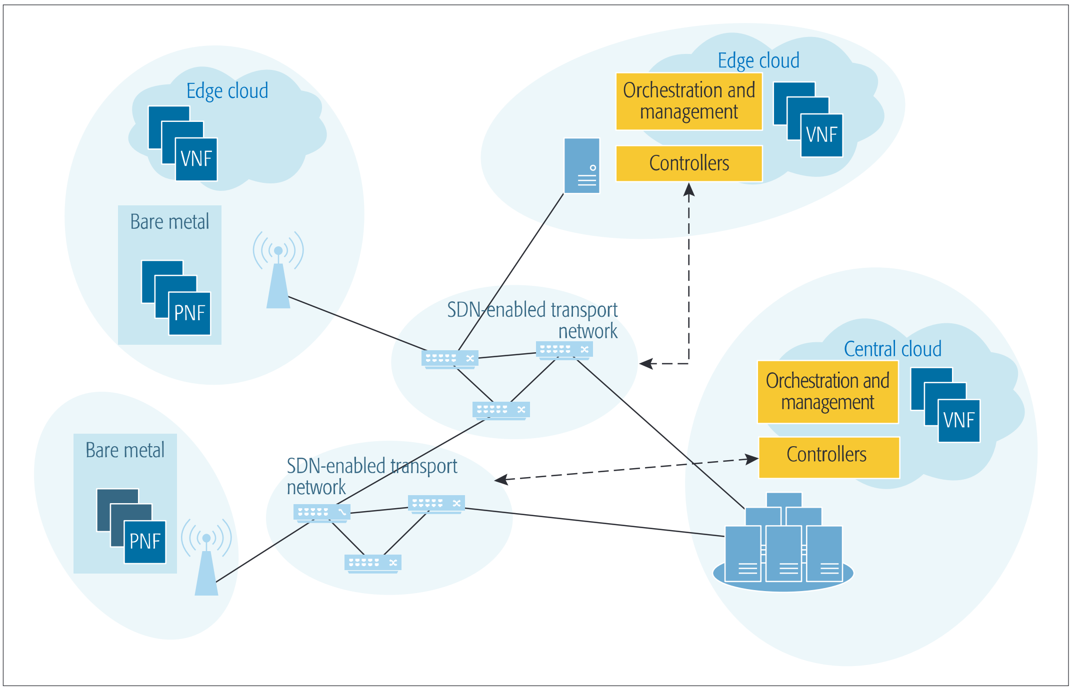

Figure 2. Relationship of functional assignment and physical architecture.

using the network slicing model. The challenge of avoiding many additional interfaces may be addressed by a flexible container protocol on the user [5] and control planes. The mobile network must further integrate legacy technologies as well to guarantee that it can operate with existing networks.

The main benefit of the described architecture is the possibility to exploit centralization gains where possible, to optimize the network operation to the actual network topology and its structural properties, and to use algorithms optimized for particular services, that is, optimize through dedicated implementations instead of parameters.

Table 1 lists examples where the operation may be optimized through different VNFs. For instance, it may be possible to use a flexible air interface numerology and, depending on the network terminal, different coding strategies, multiple-input multiple-output (MIMO) modes, and framing structures, which are optimized for throughput, delay, or reliability. However, the upper layer packetization may still be the same for all use cases, which allows the same software implementation to be reused. Another example includes cooperative transmission, where gains are highly dependent on the environment; for example, if the system is not operating at full load, cooperative scheduling may perform as efficiently as cooperative multipoint transmission, whereas at full load the gains depend highly on the number of interferers and channel knowledge.

Network Slicing

Network slicing is centered on the concept of deploying multiple dedicated logical mobile networks with varying levels of mutual isolation on top of the same infrastructure. A network slice is a collection of mobile network functions (or groups of functions) and a specific set of radio access technologies (RATs) (or specific RAT configurations) necessary to operate an end-to-end (self-contained) logical mobile network. This set of network functions and configurations may be combined such that slice-specific data and control plane functionality is tailored to the requirements of considerably different use cases, network customers, or business models. Consequently, network slicing is a technology that enables both multi-tenancy and service-tailored composition of mobile networks.

Network slicing leverages the economies of scale to be expected when running multiple logical mobile networks on top of a common infrastructure. In this sense, network slicing is an evolution of network sharing, which has been a key business model for mobile network operators to reduce deployment and operational costs. In 3GPP, the System Architecture 1 working group (WG SA1) conducted a study on actively sharing RAN resources while maintaining sharing policies and providing flexibility for on-demand resource sharing within shorter time periods [6]. Architecture and operations that enable different mobile operators with a separate core network (multi-operator core network, MOCN) to share the RAN are specified by WG SA2 [7]. In general, sharing of resources can be divided into three categories: static [8], dynamic (e.g., spectrum sharing [9]), and mixed resource allocation (spectrum sharing and virtualized resource block sharing [10]). While passive and active sharing solutions, for example, for network elements or medium access control (MAC) schedulers, are partially used and standardized today, these sharing concepts are based on fixed contractual agreements with mobile virtual network operators (MVNOs) on a coarse granularity basis (monthly/yearly) [11].

NFV, and software-defined mobile network control and orchestration enable a new level of sharing by decoupling infrastructure resources from application software, and by a split of the control and data planes. This significantly simplifies the partitioning of network infrastructure resources among different operators (or tenants). Further, slices can be isolated from each other to allow for an adaptation of security measures according to service-specific requirements (flexible security) and for securing parallel operation of multiple services or tenants. While isolation between network slices is highly important, it finds its limits where available resources need a common control (e.g., the radio scheduler): If the required isolation level cannot be preserved, a security weakness in one slice can be exploited to attack another slice. Strong security measures to maintain the isolation between multiple services and tenants operating on a shared infrastructure platform must be mandatory for all services and tenants.

Mobile core network elements rapidly evolve toward "cloud readiness" (i.e., deployment in data center environments). Consequently, each network slice can be composed from dedicated, customized instances of required network functions (NFs) and network elements (NEs). Alternatively, slices can share function instances in particular cases (e.g., for storage-intensive components like subscriber databases). In the RAN domain, extended sharing concepts facilitate the exploitation and management of radio resources offered by the owner of the network infrastructure to tenants. In this multi-tenant ecosystem, classic tenants such as mobile network operators (MNOs) and mobile virtual network operators (MVNOs) coexist with vertical businesses, for example, utility companies, automotive and manufacturing companies, and over-the-top (OTT) service providers such as YouTube and Netflix. These tenants relate to network slicing in the sense that a tenant may instantiate and make use of one or more slices. Figure 3 shows how the different NFs may be instantiated on different network elements depending on the network slice (service), that is, physical NFs would be deployed on non-virtualized hardware, different levels of edge cloud instances would provide virtualized resources (e.g., closer to the access point or exploiting points of presence) in addition to a central cloud. It further shows the virtualization layer, which is responsible for multiplexing requests from different slices operating on virtualized resources toward physical resources.

Beyond multi-tenancy, network slicing additionally serves as a means to deploy multiple service-tailored mobile network instances within a single MNO, each addressing a particular use case with a specific set of requirements (e.g., mobile broadband or IoT). In that context, the aforementioned "network of functions" concept enables the joint optimization of mobile access and core network functions. Each network slice is composed of functions according to service needs; for example, low-latency services require the allocation of most network functions at the edge.

Orchestration and Management

As mentioned before, an essential component of the mobile network is the efficient orchestration and management of mobile network functions through a low-complexity interface. In that context, software-defined network (SDN) functionality has recently gained momentum as a new approach to performing network operations. With traditional SDN, control functions are decoupled from the data plane through a well defined interface and are implemented in software. This simplifies networking, provides a higher degree of flexibility and enhanced scalability, while reducing cost. Indeed, by simply modifying the software of the control functions, SDN allows the behavior of the network to be flexibly changed, considering specific services and applications.

Following the paradigm of SDN, the control of the mobile network architecture adopts the software-defined mobile network control (SDMC) concept focusing on wireless-specific functions. Our SDMC approach resembles SDN by splitting wireless functionality into those functions that are being controlled and remain relatively stable, and those functions that control the overall network and are executed at the controller. However, our SDMC concept is specifically devised to control mobile network functionality, and it is not limited to data plane functions, but includes control plane functions of the mobile network, both of which can be placed arbitrarily in the edge cloud or the central cloud, as shown in Fig. 2.

Network functions	Relevant parameters
Cell discovery	Highly depends on carrier frequency (e.g., sub-6 GHz or mmWave), MIMO technologies (e.g., beamforming).
Mobility	Mobility may not be required by some services (metering), or only very locally (enterprises), in groups (trains), or at very high speed (cars).
Carrier aggregation	Carrier aggregation may not be needed in each scenario as it also impacts battery consumption; it could further include very distinct spectrum.
Multi-connectivity	Multi-connectivity could include different network layers (micro/macro), different technologies (WiFi/LTE), and different spectrum (sub-6 GHz/mmWave). It may further be implemented at very different layers (e.g., among others) depending on deployments.
Connectivity model	The actual connectivity may be based on bearers (high throughput) or connectionless (IoT). In the connectionless case, many non-access stratum (NAS) functions are not needed.
Coding	Coding techniques may vary depending on the use case, for example, block codes for short (sensor) transmissions or turbo codes for high throughput.
Multi-cell cooperation	Depending on the current load, deployment, and channels, tighter cooperation (joint Tx/Rx) or looser cooperation (ICIC) is possible.
Spectrum access	Depending on the use case requirements and available spectrum, possibly different spectrum access strategies may be required (e.g., licensed, unlicensed, license-assisted).
Authentication, authorization, accounting (AAA)	Depending on the applicable access control and accounting/charging policies, AAA functionality is different and may be placed/instantiated in different locations.
Parental control	Depending on the user context (children) and the requested service, the parental control function becomes part of the service chain for according service flows.

Table 1. Examples for functional optimization.

Adopting a logically centralized control unifies heterogeneous network technologies and provides efficient network control of heterogeneously deployed networks. In particular, the network control must consider evolving traffic demands, enhanced mobility management, and dynamic radio characteristics.

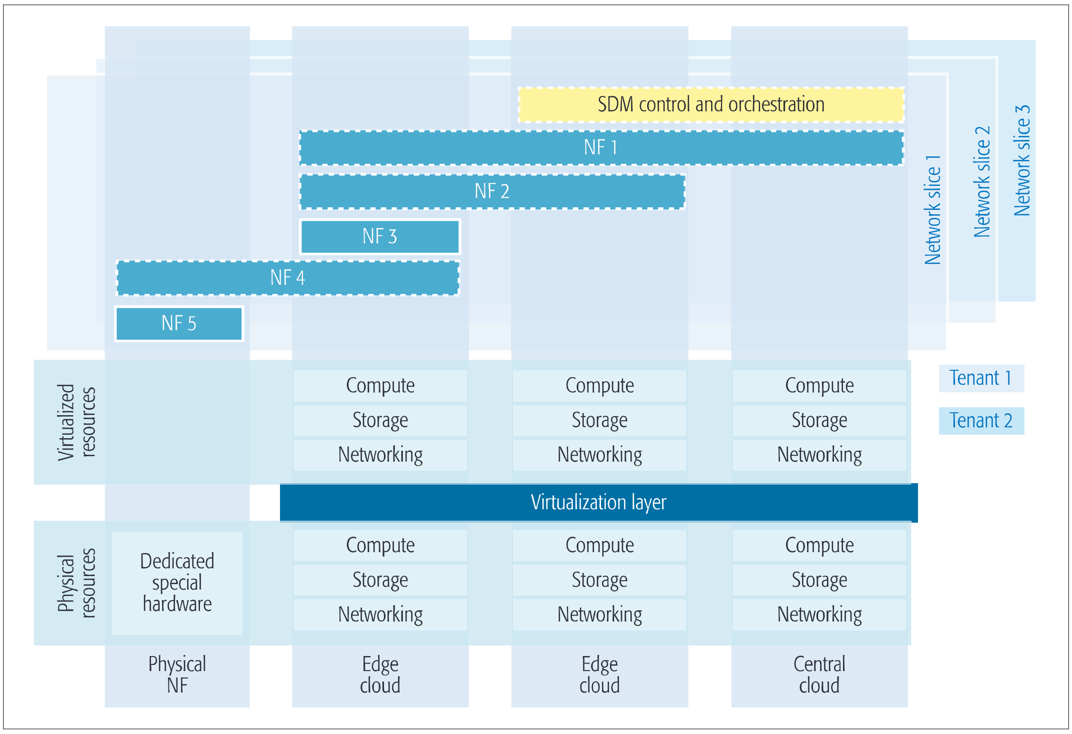

Figure 3. Network slicing concept.

To enable the SDMC paradigm within 3GPP EPS, where wireless functionality is controlled centrally, we collocate the SDMC within the 3GPP network management system. This takes advantage of the legacy performance monitoring, forming a logical global RAN information base that can be used by the SDMC to control various network functions. The control of wireless networks comprises, among others, channel selection, scheduling, modulation and coding scheme selection, and power control. Figure 4 illustrates the SDMC architecture showing the main functional features and operations. With a software-defined approach, all these functions could be performed by a programmable software defined mobile controller, which provides very important *benefits* for the operation of the mobile network.

However, it is essential to enhance the current 3GPP Type 2 interfaces (ItfN) between the network management system and the network equipment to allow the SDMC to provide network programmability and support for multi-tenancy. Those enhancements should reflect SDN capabilities such as network abstraction and control providing sufficient network management flexibility. Interfacing the SDMC with the network management system in such a manner can also enable multi-tenancy support and network programmability taking advantage of the 3GPP Type 5 interface. This allows receiving network sharing requests from MVNOs [12] and offering a means of network resource acquisition to OTT providers and verticals via the SDMC northbound application programming interface (API). In addition, the northbound interface offers the capability of flexible provision of the so-called SDMC Apps. To accommodate the related service requirements of multi-tenancy and SDMC Apps, the infrastructure provider network manager needs to interact with 3GPP policies, that is, the policy and charging rules function (PCRF), via a new network interface called *ItfPolicy*, to enable flexible policy provision for multiple tenants and network innovation.

The key advantages resulting from the proposed approach include the following.

Flexibility: One of the problems that network operators are facing today is that while wireless equipment is quite expensive, this is very rigid and does not adapt to their needs. By using SDMC, operators would be able to fit the equipment to their needs through simply reprogramming the controller and thus reducing costs, while being able to scale up and down virtual functions, also enhancing reliability.

Unified Management: Adopting logically centralized control unifies heterogeneous network technologies and provides efficient network control of heterogeneously deployed networks. In particular, the network control must consider evolving traffic demands, enhanced mobility management, and dynamic radio characteristics.

Simplified Operation of the Wireless Network: With SDMC, network operators only need to control a set of logically centralized entities that run the entire network, which, depending on actual latency requirements, possibly includes heterogeneous radio technologies.

Enabling Network Innovation: By modifying the controller functions (i.e., SDMC Apps), many new services that were not included in the initial architecture design can be enabled by modifying the network behavior to introduce service-specific enhancements within a few hours instead of weeks [13].

Programmability: By adapting the functions such as scheduling or channel selection to the specific needs of the applications or the scenario, significant performance gains can be achieved.

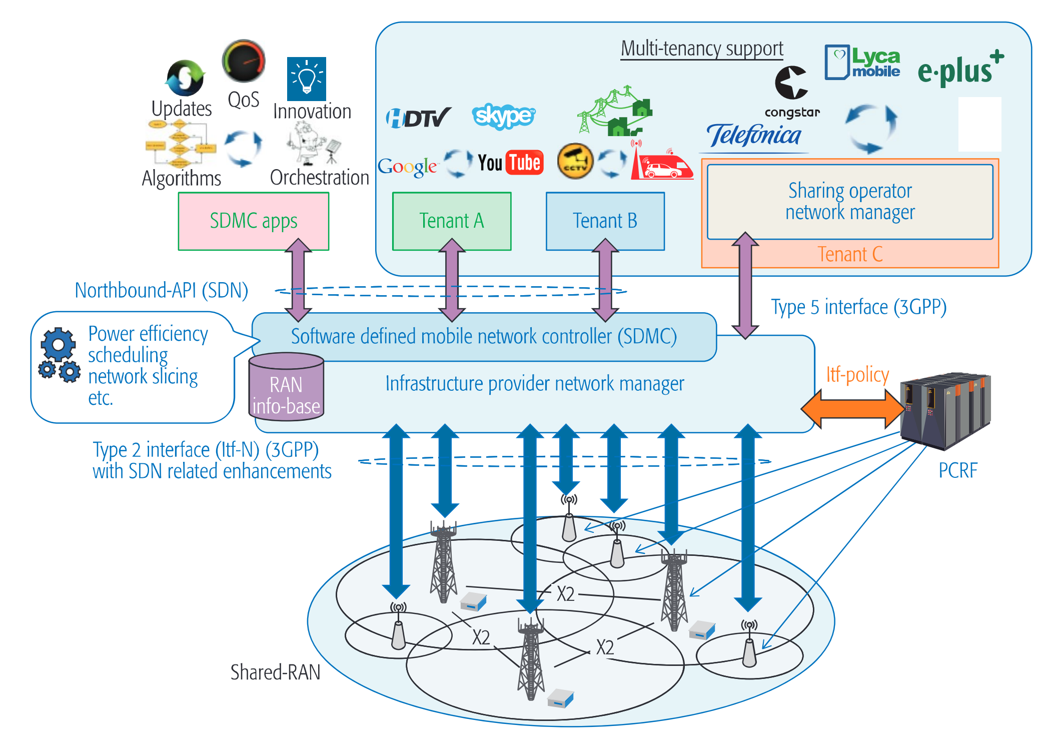

Figure 4. SDMC architecture and operations.

> By adapting the functions such as scheduling or channel selection to the specific needs of the applications or the scenario, significant performance gains can be achieved. For instance, the controller has a global view of the network, which allows for optimizing the resource allocation and scheduling across multiple BSs.

For instance, the controller has a global view of the network, which allows for optimizing the resource allocation and scheduling across multiple BSs.

Inter-Slice Resource Control: Following the network slice concept described above, infrastructure domain-hosted SDMC allows the infrastructure provider to assign unutilized resources to support third party services. Hence, the SDMC can allocate a network slice with a specified network capacity, a particular split of the control/data plane, and a selection of VNFs.

Standardization Roadmap

The International Telecommunication Union Radiocommunication Standardization Sector (ITU-R) is developing a longer-term vision of mobile networks and their evolution toward 2020 and beyond. It provides a framework and overall objectives of the future developments of 5G systems (referred to as IMT-2020) which involve several steps:

- In early 2012, ITU-R embarked on a program to develop "IMT for 2020 and beyond," setting the stage for fifth generation (5G) research activities, which are emerging around the world.
- In 2015, ITU-R finalized its vision of the 5G mobile broadband connected society, which will be instrumental in setting the agenda for the World Radiocommunication Conference 2019, where deliberations on additional spectrum will take place in support of the future growth of IMT.
- In the 2016–2017 timeframe, ITU-R will define in detail the performance requirements, evaluation criteria, and methodology for the assessment of a new IMT radio interface.
- It is anticipated that the timeframe for proposals will be focused on 2018.
- In 2018–2020 the evaluation by independent external evaluation groups and definition of the new radio interfaces to be included in IMT-2020 will take place.

Similar to previous mobile network generations, 3GPP is expected to also be the leading standardization body for 5G, and the corresponding roadmap is shown in Fig. 5. 3GPP has started to work on 5G in both the SA and RAN working groups. The current 3GPP Release 13 and the coming 3GPP Release 14 will provide enhancements to LTE-Advanced under the name "LTE-Advanced Pro." This will become the baseline technology for the evolution from LTE-Advanced to 5G. In parallel, 5G scenarios and requirements will be studied, which likely demand a revolutionary new architecture providing greater flexibility, as stated in the previous section. This work is expected to be completed by mid-2017.

SA1 has been working on a "Study on New Services and Markets Technology Enablers" (SMARTER) since April 2015. As a result, four additional study items have been created that include three vertical industries and one horizontal group. The verticals are enhanced mobile broadband (eMBB), critical communications (CriC), and massive IoT (mIoT); the horizontal study is on network operation (NEO). The latter deals with, among other issues, network slicing, interworking, and migration, as well as fixed-mobile convergence (FMC). In March 2016, another study item for 5G vehicular-to-anything (V2X) communication was agreed. SA1 plans to finalize its studies in June 2016 and then start normative work in 3GPP Release 15.

SA2 targets to finish its "Study on Architec-

The mobile network architecture evolution as discussed in this article impacts many different network components. Hence, in addition to 3GPP other standards developing organizations will participate in the definition of the future mobile network architecture.

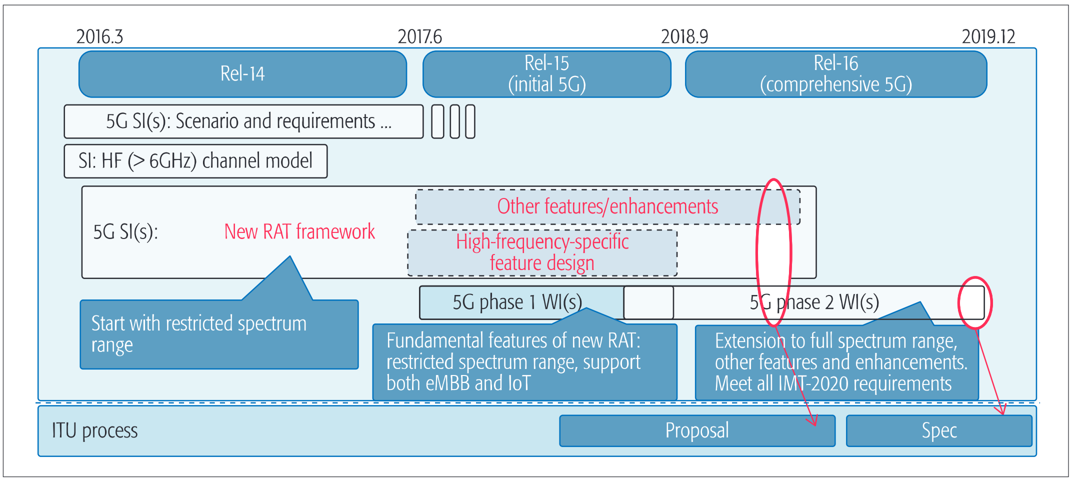

Figure 5. 3GPP LTE standardization roadmap toward 5G.

ture for Next Generation System" in September 2016. An important topic in this study will be the interface between the LTE-Advanced RAN and a future 5G core network (CN). SA2 has agreed to follow the Next Generation Mobile Network (NGMN) alliance, in particular Option 3 detailed in [12]. The new 5G CN will be able to support a new 5G RAT as well as an evolved LTE-Advanced and other RATs such as IEEE 802.11. This enables 5G network terminals to move between 5G and the evolved LTE-Advanced without any interworking between the 5G and 4G CNs, and thus provides a sound migration path from the LTE-based RAN to 5G.

The RAN working groups are targeting the first true 5G features to appear in 3GPP Release 15 (i.e., in the second half of 2018). This implies that 3GPP will complete its initial 5G specifications right before the Olympic Winter Games 2018, which will take place in Korea. The focus in this 5G "Phase 1" will mostly be on enabling new spectrum in high frequencies above 6 GHz. More features for implementing architectural enhancements will follow in 5G "Phase 2" with 3GPP Release 16 (i.e., by the end of 2019) in time for their submission to the IMT-2020 as well as the Olympic Summer Games 2020 in Japan.

Despite these planned architectural enhancements, further efforts are needed by 3GPP and other standardization bodies to accomplish the migration from 3GPP EPS toward a new 5G architecture. A completely new type of interface has to be designed and standardized when the "network of functions" is going to replace today's "network of entities," as pointed out earlier. Furthermore, the use of network slicing for multi-tenancy and multi-service described above requires a flexible execution environment that is capable of supporting the diversity of network functions in parallel. The application of SDN concepts promising this flexibility to mobile radio networks is, however, still in an experimental phase, although the C-RAN concept, RAN virtualization, and their expected centralization gains have been discussed for several years.

The mobile network architecture evolution as discussed in this article impacts many different network components. Hence, in addition to 3GPP, other standards development organizations (SDOs) will participate in the definition of the future mobile network architecture. Most notably, the following SDOs will be involved in addition to 3GPP:

- The European Telecommunications Standards Institute (ETSI) NFV industry specification group (ISG) has created a framework for virtualization of network functions. This framework has been applied successfully to VNFs, mostly in the CN. In the RAN, where hardware still plays an important role, implementation of NFV concepts is more difficult [14]; for example, the C-RAN concept with a fully centralized and virtualized RAN was among the first use cases, already discussed in 2012 in ETSI NFV. However, as of today, there are no large-scale commercial implementations. In order to gain more impact, the ETSI framework must be extended to be applicable not only to virtualized hardware but also to non-virtualized, bare metal hardware [14].
- The ETSI MEC ISG is looking at how to provide IT and cloud computing capabilities within the RAN in close proximity to mobile subscribers, allowing content, services, and applications to be accelerated, and increasing responsiveness from the edge.
- The Open Networking Foundation (ONF) is the leading force in the development of open standards for the adoption of the SDN concept. However, in order to provide the benefits described above, the SDN protocol functionalities developed by ONF (e.g., OpenFlow and OF-Config) need to be extended to cope with 5G requirements and toward 3GPP EPS.
- The Internet Engineering Task Force (IETF) is also considering the use of Internet protocols (e.g., IPv6 and IP Multicast) in 5G networks, although the work required does not have a clear scope yet. There are proposals for using IETF developed protocols such as locator/ID separation protocol (LISP), host identity protocol (HIP), and information-centric networking (ICN) to address shortcomings of the current 4G CN for the support of additional 5G functionalities (e.g., reducing network latency or supporting new mobility models). IETF is also working on the development of an architecture for service function chaining that includes the necessary protocols or protocol extensions for the nodes

that are involved in the implementation of service functions, as well as mechanisms for steering traffic through service functions.

Conclusions and Further Challenges

This article discusses the evolutionary 3GPP EPS mobile network architecture, and the need to provide a flexible architecture that integrates different technologies and enables diverse use cases. We introduce and explain various concepts such as the transition from a predefined set of functions grouped into network entities to a flexible network of functions, the network slicing concept, and software defined mobile network control, orchestration, and management. In addition, the relevance of different standards defining organizations has been outlined and their roadmap has been detailed.

It is in our opinion that it is highly important to consider the future evolution of 3GPP EPS not only as the introduction of a novel air interface but as the evolution of one mobile network architecture toward a "system of systems" where many different use cases, technologies, and deployments are integrated, and the operation of each system is tailored to its actual purpose.

Acknowledgment

This work has been performed in the framework of the H2020-ICT-2014-2 project 5G NORMA. The authors would like to acknowledge the contributions of their colleagues. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein.

References

[1] 3GPP, "TS 36.300; Overall Description; Stage 2," tech. spec., 2013.

[2] 3GPP, "TS 23.401; General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access," tech. spec., 2015.

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

[4] D. Wübben *et al.*, "Benefits and Impact of Cloud Computing on 5G Signal Processing," *IEEE Signal Processing Mag.,* Oct. 2014.

[5] A. de la Oliva *et al*., "Xhaul: Toward an Integrated Fronthaul/Backhaul Architecture in 5G Networks," *IEEE Wireless Commun*., vol. 22, no. 5, Oct. 2015.

[6] 3GPP, "TR 22.852, Study on Radio Access Network (RAN) Sharing Enhancements, Release 12," tech. rep., June 2013.

[7] 3GPP, "TS 23.251, Network Sharing; Architecture and Functional Description, Release 12," Dec. 2013.

[8] S. Paul and S. Seshan, "Technical Document on Wireless Virtualization," GDD-06-17, GENI, Sept. 2006.

[9] Y. Zaki *et al.*, "LTE Wireless Virtualization and Spectrum Management," *Proc. IEEE/IFIP Wireless and Mobile Networking Conf.*, Budapest, Hungary, Oct. 2010.

[10] T. Guo and R. Arnott, "Active RAN Sharing with partial Resource Reservation," *Proc. IEEE 78th VTC-Fall*, Las Vegas, NV, Sept. 2013.

[11] ITU-T Y.3011, "Framework of Network Virtualization for Future Networks," *Next Generation Networks-Future Networks*, Jan. 2012.

[12] NGMN Alliance, "NGMN 5G White Paper," tech. rep., Feb. 2015.

[13] C. J. Bernardos *et al*., "An Architecture for Software Defined Wireless Networking," *IEEE Wireless Commun*., vol. 21, no. 3, June 2014.

[14] Small Cell Forum, "Network Aspects of Virtualized Small Cells, Release 5.1, Document 161.05.1.01," tech. rep., June 2015.

Biographies

Peter Rost [SM] (peter.rost@nokia.com) received his Ph.D. degree from Technische Universität Dresden in 2009 and his M.Sc. degree from the University of Stuttgart in 2005. Since May 2015, he has been member of the Radio Systems research group at Nokia Germany, contributing to the European projects 5G-NORMA and METIS-II, and business unit projects on 5G architecture. He serves as a member of the IEEE ComSoc GITC, VDE ITG Expert Committee Information and System Theory, and as Executive Editor of *IEEE Transactions on Wireless Communications*.

Albert Banchs [SM] (banchs@it.uc3m.es) received his M.Sc. and Ph.D. degrees from UPC-BarcelonaTech in 1997 and 2002. He was at ICSI Berkeley in 1997, at Telefonica I+D in 1998, and at NEC Europe from 1998 to 2003. Currently, he is an associate professor with the University Carlos III of Madrid, and has a double affiliation as deputy director of the IMDEA Networks institute. His research interests include performance evaluation and algorithm design in wireless networks.

Ignacio Berberana (ignacio.berberana@telefonica.com) received his M.S. degree in mining engineering from Madrid Polytechnic University in 1987. In 1988 he joined Telefonica I+D, where he has worked mainly in wireless communications, including several European projects (CODIT, MONET, Artist4G, iJOIN). Currently, he is responsible for the Innovation unit in the Radio Access Networks direction of the Telefónica Global CTO office, dealing with long-term evolution of mobile access, including 5G systems.

Markus Breitbach (markus.breitbach@telekom.de) is working as a senior expert in the area of end-to-end network architecture. Before joining Deutsche Telekom in 2006, he developed concepts for UMTS base stations and their HSPA schedulers for a major infrastructure supplier. In the last years, he has been working on network virtualization. Holding both a Ph.D. in electrical engineering and an M.B.A., his ambition is to design innovative network concepts that fit well into the surrounding business picture.

Mark Doll (mark.doll@alcatel-lucent.com) received his Dipl.-Phys. degree in physics from Technische Universität Braunschweig in 2000 and his Dr.-Ing. in computer science from Karlsruhe Institute of Technology (KIT) in 2007. At KIT, he worked on mobility, multicast, and QoS support for the Internet. Upon joining Nokia Bell Labs, his work shifted to EPS CoMP and EPS air-to-ground communication for aircrafts and now focuses on post-cellular "user-centric" wireless access for 5G. He acts as 5G NORMA's technical manager.

Heinz Droste (Heinz.droste@telekom.de) works for Deutsche Telekom in Darmstadt on mobile communication related projects. Antennas and radio wave propagation belong to his knowledge field as well as system-level simulation and radio network planning. His current R&D activities at Telekom Innovation Laboratories focus on the optimization of EPS and EPS-A deployments where he is acting as senior expert and project manager. He is actively contributing to the EU funded R&D project 5G NORMA.

Christian Mannweiler (christian.mannweiler@nokia.com) received his M.Sc. (Dipl.-Ing.) and Ph.D. (Dr.-Ing.) degrees from Kaiserslautern University, Germany, in 2008 and 2014, respectively. Since 2015, he has been a member of the Network Management Automation research group at Nokia. He has co-authored numerous articles and papers on future mobile network technologies and architectures. He has worked in several nationally and EU-funded projects covering the development of cellular and industrial communication systems, among them H2020-5G-NORMA, FP7-C-Cast, FP7-METIS, BMBF-SolarMesh, BMBF-PROWILAN, and BMWi-CoCoS.

Miguel A. Puente (miguelangel.puente@atos.net) received his M.Sc. in telecommunications engineering from the Universidad Politécnica de Madrid (UPM) in 2012, including an information technology Master's degree from the University of Stuttgart (2010–2012). Since 2012 he is with Atos Research & Innovation in Spain, where he is involved in European research projects addressing 5G, EPS, Cloud Computing, Mobile Cloud/Edge Computing, QoE/QoS optimization and recursive Internet architectures. From 2014 he is a Ph.D. candidate at UPM.

Konstantinos Samdanis (samdanis@neclab.eu) is a senior researcher and backhaul standardization specialist with NEC Europe. He is involved in research for 5G networks and active in BBF on network virtualization, and published numerous papers/patents. He has served as a Feature Topic Editor of *IEEE Communications Magazine* and *IEEE MMTC E-Letters*, Co-Chair of IEEE ICC 2014 and EuCNC 2015, and edited *Green Communications* (Wiley). He received his Ph.D. and M.Sc. degrees from Kings College London.

Bessem Sayadi (bessem.sayadi@alcatel-lucent.com) received his M.Sc. and Ph.D. degrees from SUPELEC in 2000 and 2003. He is a senior researcher at Alcatel-Lucent Bell Labs. He has worked on several nationally and EU-funded projects. He has authored over 60 publications. He holds 20 patents and has more than 25 patent applications pending in the area of video coding and wireless communications.

It is in our opinion of high importance to consider the future evolution of 3GPP EPS not only as the introduction of a novel air interface but as the evolution of one mobile network architecture toward a "system of systems" where many different use cases, technologies, and deployments are integrated, and the operation of each system is tailored to its actual purpose.