Received: 22 December 2016 | Revised: 17 May 2017 | Accepted: 18 May 2017
DOI: 10.1002/ett.3205

RESEARCH ARTICLE

WILEY

Toward the network of the future: From enabling technologies to 5G concepts

Dario Bega[1,2] | Marco Gramaglia[2] | Carlos Jesus Bernardos Cano[2] | Albert Banchs[1,2] | Xavier Costa-Perez[3]

[1]IMDEA Networks Institute, Madrid, Spain
[2]Telematic Engineering Department, Universidad Carlos III de Madrid, Madrid, Spain
[3]NEC Laboratories Europe, Heidelberg, Germany

Correspondence
Dario Bega, IMDEA Networks Institute, 28918 Madrid, Spain; or Telematic Engineering Department, Universidad Carlos III de Madrid, Madrid, Spain.
Email: dario.bega@imdea.org

Funding information
European Commission, Grant/Award Number: 671636

Abstract
There is a wide consensus by the research community and the industry that it will not be possible to satisfy future mobile traffic demand and application requirements by simply evolving the current fourth-generation architecture. Instead, there is a need for a considerable revision of the mobile network system: such an effort is commonly referred to as the future *fifth-generation (5G) architecture*, and large-scale initiatives all around the globe have been launched worldwide to address this challenge. While these initiatives have not yet defined the future 5G architecture, the research community has already invested a very substantial effort on the definition of new individual technologies. The fact that all new proposals are tagged as 5G has created a lot of confusion on what 5G really is. The aim of this article is to shed some light on the current status of the 5G architecture definition and the trends on the required technologies. Our key contributions are the following: (1) we review the requirements for 5G identified by the different worldwide initiatives, highlighting similarities and differences; (2) we discuss current trends in technologies, showing that there is a wide consensus on the key enablers for 5G; and (3) we make an effort to understand the new concepts that need to be devised, building on the enablers, to satisfy the desired requirements.

1 | INTRODUCTION

It is well known that mobile data consumption is exploding, driven by the increased penetration of smart devices, better screens, and compelling services, among other factors. At the same time, emerging communication services impose new requirements on the network: use cases such as tactile internet, vehicular communications, high-resolution video streaming, road safety, or real-time control place have stringent requirements on throughput, latency, reliability, and robustness. It is widely agreed that all these new requirements and demands cannot be provided by simply evolving the current fourth-generation (4G) architecture. Therefore, novel architectural patterns and solutions must be introduced. The core network will be especially impacted by this re-engineering, but also the access will incorporate new technologies. The new architecture that will result from this redesigning effort is commonly referred to as the *fifth-generation (5G) architecture*.

Driven by the above trends, there is currently huge ongoing worldwide effort toward the definition of the new 5G architecture, with initiatives such as the (1) 5G Infrastructure Public Private Partnership (5G-PPP)[1] in Europe, (2) 5G Americas[2] in America, (3) IMT-2020 (5G) PG[3] in China, (4) 5G Forum[4] in Korea, and the (5) Fifth Generation Mobile Communication Promotion Forum (5GMF)[5] in Japan.

TABLE 1 Fifth-generation (5G) key performance indicators (KPI) according to the different initiatives[a]

KPI	5G-PPP	5G Americas	IMT-2020 (5G) PG	5G Forum	5GMF	METIS-II
Data rate	10 Gbps	100x	10 Gbps	50 Gbps	10 Gbps	10 Gbps
Latency	5 ms (E2E)	5x to 100x	1 ms (E2E)	1 ms (E2E)	1 ms (E2E)	ms (E2E)
Connected devices	1 M/km^2	10x to 100x	1 M/km^2	10x to 1000x	10 000 x cell	10x to 100x
Capacity	10 Tbps/km^2	1000x to 5000x	10 Tbps/km^2	1000x	1000x	1000x
Energy consumption	10x		100x	1000x		10x
Reliability	five nines	*high*	five nines	*hyper*		*ultra*
Mobility	≈ 500 km/h	> 350 km/h	> 500 km/h	> 350 km/h	≈ 500 km/h	
Cost	*ultralow*	*< 4G*	100x	*hyper low*		*as today*
Service creation time	90 min					

Abbreviations: E2E, end to end; 5GMF, Fifth Generation Mobile Communication Promotion Forum; 5G-PPP, 5G Infrastructure Public Private Partnership.
[a]10x means 10 times better.

Standardization activities such as 3GPPP SA2,[6] SA5,[7] and technical specification group radio access network (TSG-RAN)[8] are the other side of the coin. These activities that we will detail further range from the definition of the technologies that improve the underlying wireless interfaces to other that, by leveraging on ongoing cloudification trends, improve transport technologies and adopt the softwarization paradigm. The fact that many of these proposals are tagged by the 5G label has produced some confusion on what 5G is and which are the technologies that will actually conform the future 5G network.

The aim of this paper is to review the major ongoing activities in this area and put some order on the current flood of supposedly 5G building blocks. While the 5G architecture has not yet been defined, and hence any attempt to define its technological components is a mere speculation, we provide a thorough review of the current trends identifying the key performance indicators (KPIs), new concepts, and their enabling technologies considered necessary for the future 5G network. The remainder of this paper is organized as follows. Section 2 presents a view of the consolidated performance that 5G systems should satisfy, highlighting the similarities and discrepancies between the requirements provided by the different initiatives deriving the KPIs proposed. Section 3 describes the technologies upon which the new 5G concepts rely to enable the expected requirements. In Section 4, the 5G concepts, ie, the approaches the 5G architecture will based on, have been identified, while in Section 5, we detail the ongoing worldwide activities. We conclude our work in Section 6.

2 | THE NEED FOR 5G

As in the design of any system, the objectives in terms of KPIs are key for the deployment of future 5G systems. To this end, the main driving bodies behind 5G have recently dedicated very substantial effort toward not only identifying but also quantifying the objectives of the 5G technology in terms of KPIs. Table 1 depicts the key KPIs that have been proposed by the main driving actors of the 5G technology, including Europe (ie, 5G-PPP and METIS-II[9]), America (ie, 5G Americas), China (ie, IMT-2020 (5G) PG), Korea (ie, 5G Forum), and Japan (ie, 5GMF).

We observe from the Table that, in addition to traditional KPIs for network performance, 5G also includes some additional indicators that are crucial for the upcoming network. Indeed, classical indicators for network design such as peak data rates, average and cell-edge user throughput, and overall cell throughput will continue to be important for the 5G network design. However, additional KPIs also need to be defined:

- Because of the massive uptake of machine-type traffic supporting new vertical user groups in industry, public administration, and business, KPIs such as network availability, coverage (both deep indoors and for sparse rural areas), robustness, and reliability play a very important role.
- The current trends toward Internet of Things, which is one of the fundamental use cases of 5G, point toward the support for dramatically increased numbers of almost zero-complexity devices with long standby times, all of them essential to support such a use case.
- Other very important use cases in 5G such as tactile internet and vehicular communications require extremely low latencies, which is one of the most stringent KPIs included in the Table.
- Another major challenge is energy efficiency, driven by the need to support growing mobile data volumes without increasing the energy consumption, which translates to greener operations and the corresponding cost savings.

TABLE 2 There is rough consensus among different 5G players on the enabling technologies

	Spectrum	mMIMO	SDN	NFV	C-RAN	Local Offloading	Small Cells
5G-PPP	✓	✓	✓	✓	✓		✓
5G Americas	✓	✓	✓	✓	✓	✓	✓
IMT-2020 (5G) PG	✓	✓	✓	✓	✓	✓	✓
5G Forum	✓	✓	✓			✓	✓
5GMF	✓	✓	✓	✓	✓		✓
METIS-II	✓	✓	✓	✓	✓	✓	✓
Vendor 1	✓	✓	✓	✓	✓	✓	✓
Vendor 2	✓	✓	✓	✓	✓	✓	
Vendor 3	✓	✓	✓	✓	✓		✓
Vendor 4	✓	✓	✓	✓		✓	✓

Abbreviations: C-RAN, centralized radio access network; mMIMO, massive multiple-input multiple-output; NFV, network function virtualization; SDN, software-defined networking; 5GMF, Fifth Generation Mobile Communication Promotion Forum; 5G-PPP, 5G Infrastructure Public Private Partnership.

- Finally, because of the broad adoption of flat rates, mobile operators will have to support the growth in mobile data volume resulting from the above KPI without increases in subscription fees; cost efficiency will thus remain a key challenge for future network developments.

When comparing the data provided in Table 1 for the different actors, the main observation is that they all largely agree on the target performance of 5G systems. While the parameters provided by some of these actors are more concrete than others and there may be a slight deviation in some of the parameters, the numbers provided by different fora fall within the same order of magnitude in almost all of the cases. Therefore, the main conclusion is that there is a wide consensus on the performance requirements of future 5G systems.

It is worth noting that there is one KPI in Table 1 that is only indicated by 5G-PPP and no other forum: the service creation time. Indeed, the flexibility of easily customizing the network infrastructure to new services may be one of the driving design criteria in 5G. Therefore, it is somehow surprising that such a KPI is ignored by the other actors.

At a more general level, the KPIs provided in this table refer mostly to the data plane performance, and little emphasis is placed on the flexibility of adapting the network behavior to the specific requirements of the different operators and the services they are providing. Given the current trends toward virtualization and softwarization of the network driven by the need for flexibility, it seems that future networks should place much more emphasis on this kind of KPI.

3 | ENABLING TECHNOLOGIES

The requirements that need to be addressed by the future 5G networks clearly demand new technologies and architectures, as simply evolving existing 4G deployments would not be enough. While these new technologies, which we refer to as *5G enablers* in this article, are essential pieces of the future 5G technology, they will not suffice by themselves to satisfy the requirements identified. The new concepts required, along with the mapping between 5G enablers, concepts, and requirements, are studied in Section 4.

We next identify and describe the main 5G enablers based on the existing components being considered by the most relevant players in the research and standardization communities. It is important to highlight that there is quite a rough consensus on the technologies that are considered fundamental enablers for 5G among these key players.[1-20] Table 2* graphically details which of the identified 5G enablers are considered by each of these players. We blinded and aggregated the selected vendors (NEC, Huawei, Ericsson, Samsung, Nokia) to emphasize this point: we are not claiming to interpret future strategies of network equipment vendors nor providing their comprehensive vision of 5G. Our goal is to present distilled information from their white papers to provide useful insight on the relevant technological trends in 5G.

*We remark that we solely used the information available in each player 5G vision white paper.[1-20] Vendors' participation in standardization activities or product development is not considered in our comparison.

3.1 | Spectrum and massive MIMO

Future networks will need to cope not only with higher data rates, but they will also need to provide extremely low latencies and support a substantially larger number of connected devices. In order to address this, a combination of new advanced spectrum-efficiency mechanisms (eg, carrier aggregation techniques) and use of new frequency bands (such as 60 GHz and millimeter wave) are required. Unlicensed spectrum, for instance, can be used in combination with licensed spectrum (for critical control signalling and mobility handling) to boost capacity. More spectrums can also be obtained with authorized shared access, in which the cellular system can access additional free spectrum otherwise apportioned for use by other (nontelecom) services. The use of high-frequency bands also allows for massive multiple-input–multiple-output (mMIMO) technique that, by exploiting antenna arrays with a few hundred antennas simultaneously, can serve many tens of terminals in the same time-frequency resource, increasing the capacity 10 times or more, and enables a significant reduction of latency on the air interface.[21] Those new access technologies in the actual Long-Term Evolution (LTE) network paradigm would be implemented separately since it does not allow multiple connection utilizing different technologies. This way, it is not possible to completely exploit their benefits.

3.2 | Software-defined networking

Currently, it is extremely complex to express operators' high-level network policies, since configuring each individual network device separately using low-level and often vendor-specific commands is needed. Besides, networks are vertically integrated. The control plane and the data plane are bundled inside the networking devices, reducing flexibility and hindering innovation and evolution of the networking infrastructure.

The software-defined networking (SDN) paradigm[22] separates the control and data-forwarding planes. Such separation allows for quicker provisioning and configuration of network connections. With SDN, network administrators can program the behavior of both the traffic and the network in a centralized way, without requiring independently accessing and configuring each of the network hardware devices. This approach decouples the system that makes decisions about where traffic is sent (ie, control plane) from the underlying system that forwards traffic to the selected destination (ie, data plane). Among other advantages, this simplifies networking and the deployment of new protocols and applications. In addition, by enabling programmability on the traffic and the devices, an SDN network might be much more flexible and efficient than a traditional one.

3.3 | Network function virtualization

In today's networks, every time a new service has to be deployed, operators have to buy proprietary devices, which often require a lot of time to be produced because of carrier grade quality requirements. In addition, this equipment needs physical space for its installation and energy to run. Last but not the least, a trained personnel is required to setup, configure, and operate it.

The new 5G requirements for more diverse and new (short-lived) services with high data rates have made operators even more reluctant to continue following the current networks' operation model. They are excited and hopeful with the advent of virtualization techniques in the field of networks, what is widely known as network function virtualization (NFV).[23]

The key concept of NFV is the decoupling of physical network equipment from the functions that run on them (decoupling the intelligence from the raw capacity). With this approach, network functions (eg, a load balancer) are now dispatched as software components, allowing for the consolidation of many network equipment types onto high, commercial-off-the-shelf–based, volume servers, switches, and storage, which could be located in data centers, distributed network nodes, and at end-user premises. The virtual network functions (VNFs) that provide network services can be flexibly reimplemented and relocated to different network locations as needed since they may run on general-purpose hardware, thus making it faster and cheaper to put new services into operation. Besides, combined with SDN, it enables multitenant and sliced networks in which multiple service providers share the physical resources, reducing the time and costs to deploy a new service.

3.4 | Centralized RAN

Centralized radio access network (C-RAN) is one possible way to efficiently centralize computational resources, by connecting multiple sites to a central data center where all the baseband processing is performed. Radio signals are exchanged over dedicated transmission lines (called fronthaul) between remote radio heads and the data center. With a pure C-RAN approach, only fiber links are today capable of supporting the required data rates (eg, about 10 Gb/s for time-division LTE with a 20-MHz bandwidth and 8 receiver antennas), being this need for a high-capacity fronthaul the main drawback of C-RAN.[24] The 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 radio access points has triggered the design of cloud RAN approaches. This allows flexible and adaptive software deployment, taking advantage of the enormous potential of cloud computing. In a flexible cloud RAN environment, different RAN functions can be optimally and dynamically allocated and moved between the radio access points and the data centers deployed within the network, even at the core.

Centralized RAN is therefore a key 5G enabler as it allows flexibly moving functions within the network, facilitating the achievement of lower latencies and the use of more efficient mobility mechanisms (eg, depending on the nature of the traffic, mobility anchors might be deployed closer to the end-user devices).

3.5 | Local offloading

5G networks are foreseen to share resources to cope with disparate traffic demands from heterogeneous users/applications (eg, Internet of Things and 4K high-definition video streaming). Additionally, some services may benefit from local processing capabilities at the edge of the network, whereas other services might demand a centralized processing because of privacy or legal concerns.

In this heterogeneous environment, local offloading strategies are needed to flexibly and opportunistically (1) allow for extremely high bit rates, low delays, and low power consumption exploiting the user equipment's proximity and (2) reduce network load and improve spectrum efficiency exploiting direct transmission among devices.[25] For mobile networks, the more promising technique is network-assisted device-to-device communications, where 2 nearby devices can communicate with each other with limited base station involvement. Besides all the advantages, for both service providers and users, device-to-device communication raises new challenges as security and interference management, requiring also new pricing models.[26] The European Telecommunications Standards Institute, recognizing the important role of local offloading strategies, has standardized a new technology called Mobile Edge Computing[27-29] with the aim of improving its efficiency. Mobile Edge Computing provides an information technology service environment and cloud-computing capabilities within the RAN, thus near mobile subscribers. This way, it is able to reduce latency, ensure highly efficient network operation and service delivery, and demonstrate computing agility in the computation offloading process.

Finally, in this heterogeneous environment, there is a need to flexibly and opportunistically allow locally breaking out selected traffic closer to the edge (ie, offloading the network core) and exploiting the use of different gateways for traffic with different connectivity and mobility requirements.

3.6 | Small cells

It is well known that increased spatial reuse (ie, denser networks and smaller cells) has been the dominant factor in the increase of the system throughput of cellular networks, as compared with new physical layer techniques. Therefore, the use of very dense, low-power, small-cell networks is a clear option to cope with future data rate demands. Ultradense deployments exploit 2 fundamental effects: (1) the distance between the radio access point and the user is reduced, leading to higher achievable data rates, and (2) the spectrum is more efficiently utilized because of the reuse of time-frequency resources across multiple cells.

Small cells do not replace but complement existing macro cellular deployments, which are still required to provide coverage for fast-moving users and in areas with low user density. The denser the network is, the higher the probability that an individual access point just carries a light load. Therefore, smart coordination and management mechanisms are required to achieve a more efficient use of spectral and energy resources.[30]

Since both higher individual per-flow data rates and aggregated offered loads are expected in the near future, small cells, together with new spectrum and MIMO, are key 5G enablers.

4 | 5G CONCEPTS

Pushed by the rising of new technologies (what we refer to as 5G enablers in Section 3), new solutions need to be devised. Indeed, new algorithms and protocols are needed to exploit the above technologies toward achieving the goals identified in Section 2. Throughout this section, we review and classify the most important ones available in the literature. Figure 1 provides

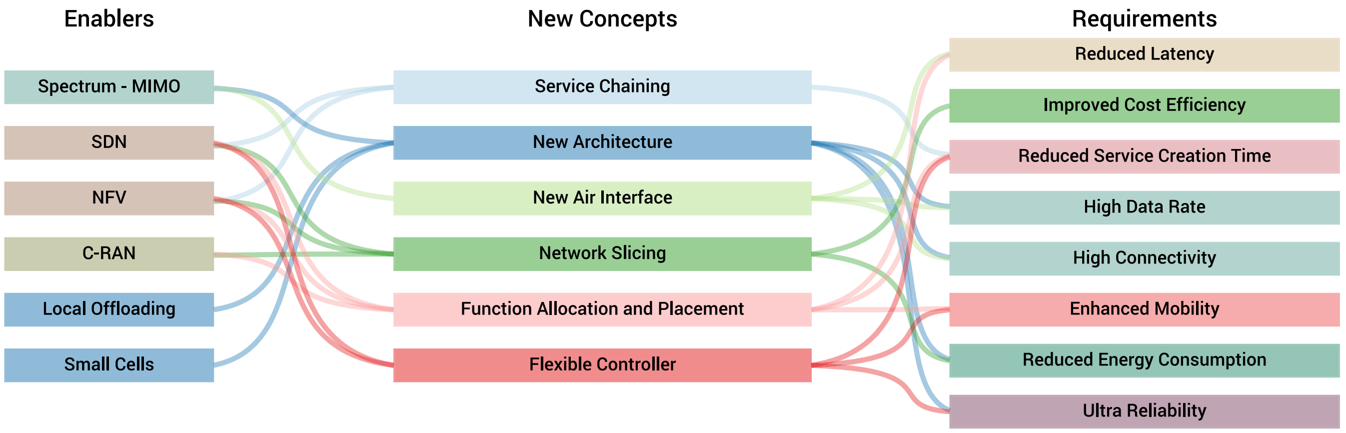

FIGURE 1 Fifth-generation (5G) new concepts, enablers, and requirements. The figure identifies the enablers upon which each concept relies and which requirements it contributes to satisfy. C-RAN, centralized radio access network; MIMO, multiple-input multiple-output; NFV, network function virtualization; SDN, software-defined networking

a graphical representation of the proposed taxonomy.† It shows the enablers that we have identified in the previous section, highlighting the new 5G concepts they mainly contribute to enable. Then, we disclose which requirements can be satisfied by means of the new concepts (we present only the principal connections among the actors to focus on their main role).

4.1 | Service chaining

A fundamental component toward achieving the flexibility needed in the future 5G networks is the self-adaptation capacity. Usually, network services are built on top of several, well-defined, functions (eg, firewalls and load balancers). In the legacy networking concepts, the placement of these functions was tightly coupled with the underlying network topology. The development of the SDN and NFV concepts has substantially changed the game. The possibility of running a network function almost anywhere in a data center (on general-purpose server hardware) decouples the sequence of network functions needed by a service from the physical topology. Network functions are hence not deployed according to their functionality (eg, placing load balancers close to the servers) but are defined and chained in an abstract fashion.[31]

The main advantage provided by this approach is *flexibility*. Chains can not only be instantiated in the network, but they can also be modified according to the quality of service (QoS) demands of the users. For example, a video optimizer or a content distribution network *middlebox* can be easily placed inside the chain on the fly if needed by the current network conditions. Therefore, service function chaining allows the rapid development of new services: new function chains can be deployed on demand, rather than forcing the modification of the network topology to insert a new function needed by the targeted service.

4.2 | Coalesced access architecture

Having to face the current trends in mobile data consumption, the requests of new services (eg, massive machine-to-machine communications), and the always-increasing number of connected devices, the current cellular-based network architecture clearly shows its shortcomings. Providing very heterogeneous services using the same infrastructure will not be feasible anymore in the near future, even with very efficient modulations and coding schemes. For this reason, future 5G networks will be based on a ductile access architecture, leveraging also on small cells and on smart flow offloading whenever it is possible. This fine-grained wireless access structure requires a very careful coordination among all the elements of the network: something unlikely achievable with the legacy architecture, but possible by using new 5G concepts as flexible mobile network controller (FMNC).

The optimized spectrum utilization, in exchange, will provide increased performance (in terms of available bandwidth and capacity) with increased efficiency from an energy point of view. Moreover, having small high-capacity cells will certainly improve the signal quality received by the user device, helping to reach the envisioned goals for reduced energy consumption

†Note that the colors are used to highlight how the enablers on the left-hand side relate to the concepts in the middle and how these cooperate to meet the 5G goals represented on the right-hand side.

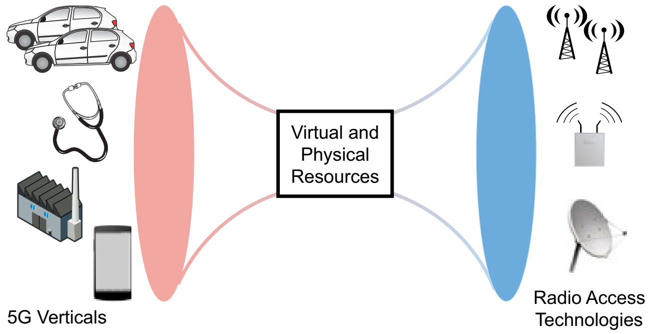

FIGURE 2 The future coalesced architecture: several fifth-generation (5G) verticals use heterogeneous access technologies. The 5G core provides a unified application programming interface for their management

and the overall reliability of the system. Figure 2 shows an example of how the future 5G networks provide and unified interface for the management and control of heterogeneous access technologies to various 5G services.

4.3 | New air interfaces

In 4G networking, the available radio access technologies were somehow limited to cellular ones: LTE Advanced (LTE-A) and Worldwide Interoperability for Microwave Access. In 5G networking, the intrinsic flexibility of the proposed architectures allows for the deployment of more heterogeneous radio access technologies. The rise of new communication techniques at the physical and medium access control layers fosters the research on new air interfaces. The availability of more and faster communication channels enables 3 of the envisioned goals of 5G: reduced latency, higher data rates, and reduced energy consumption.

The current structure of LTE-A was designed to be an enhancement of third-generation networking. The targeted KPIs were related to the voice and data communications from mobile terminal use cases (ie, throughput, capacity, and blocking probability during calls). As time went by, the need for new services arose: some of them required very diverse characteristics that were just not targeted by the initially envisioned KPIs. Although the support for more enhanced service is currently being provided in LTE-A, a focused revision of the access network (that is, an evolution of second generation and third generation) is needed.

5G networks hence propose a complete paradigm switch: not only making more bandwidth available to the users but also achieving it through the seamless integration of new frequency bands in the range of 6 to 100 GHz (made available using mMIMO deployments), advanced spectrum-efficiency management methods (especially in the legacy sub-6-GHz band), and their integration.[32]

Among the considered innovations, there will be evolved waveforms, but also wireless network coding will play a major role during the definition of the new 5G air interfaces. This also tackles the medium access control layer, with the definition of an integrated frame structure capable of allowing very diverse traffic types.

The key point is not only to implement new access technologies but also to exploit them, allowing multiconnectivity, thus the possibility to connect the same user using different access technologies such as 5G, Wi-Fi, LTE, 6 GHz, millimeter wave, or visible light communications at the same time. The main innovation is not to develop new technologies but to utilize them together, improving, in this way, their efficiency.

The current consensus is that 5G, to be able to provide a very considerable high data rate and to reduce the latency, needs to combine the use of new frequency bands (higher frequencies), advanced spectrum efficiency enhancement methods in the legacy band, and seamless integration of licensed and unlicensed bands.

4.4 | Network slicing

Nowadays, very different applications share the same communication infrastructure, but communication networks were not designed with this in mind. With the trend of increased heterogeneity, 5G networks must be designed embracing this from the very beginning. Moreover, the final goal of 5G is not only to support very heterogeneous services but also to reduce the costs (operating expenses and capital expenditure).

Theoretically, this goal can be achieved by having several physical networks deployed, one for each service (or even one for each business). Isolated services can hence use their resources in an optimal way, avoiding difficult reconfiguration of hardware and network entities. Clearly, this approach cannot be applied to real networks and calls for a solution that allows both efficient resource sharing (ie, *multitenancy*) and utilization.

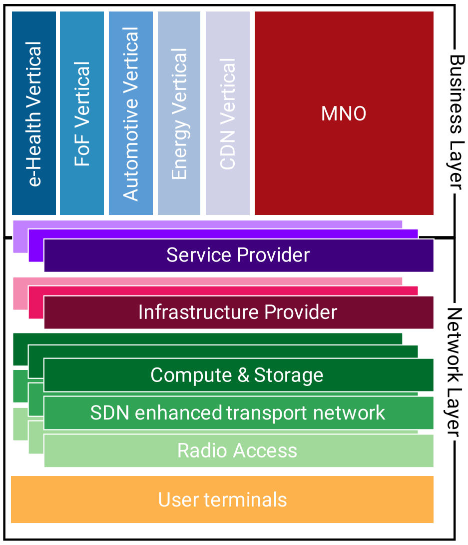

FIGURE 3 Multitenancy in a network-sliced representation. CDN, content distribution network; FoF, Factory of the Future; MNO, mobile network operator; SDN, software-defined networking

A mild approach to multitenancy, mostly passive, is already standardized and applied by many operators that currently share cell sites. However, the equipment still belongs to each operator, limiting, hence, the cost reduction. 5G networks will go one step further, pushing for the active sharing of resources among different tenants, allowing for the so-called *verticals*, where also nonoperators may need the use of the network.

This kind of approach can be reached because of the *programmability* feature of future 5G networks, which will be heavily based on the NFV and SDN paradigms. Hence, different tenants can share the same general-purpose hardware to provide all the needed functionality to the final users. A first approach in this sense was proposed by NGMN with the introduction of the network slice concept.[33] *Softwarization* techniques paved the way toward the virtualization of network resources; thus, fully decoupled networks can be built on top of virtual infrastructure laying on a shared physical infrastructure.

Therefore, a network slice can be defined as a subset of virtual network infrastructure resources dedicated to a specific tenant, who can use it to provide its envisioned service. The virtualization layer between the different slices and the physical infrastructure ensures the economy of scale that suggests the viability of the network slicing approach.

Network slices are created mostly with a business purpose: following also the 5G verticals spirit, an infrastructure provider will assign one or more network slices to each service of a service provider portfolio (eg, the vehicular network slice, the factory of the future slice, and the health net slice; see Figure 3). The required KPIs are provided only when needed and where needed, allowing, hence, better network utilization with the consequent running cost reduction.

Network slicing calls for a flexible architecture capable of orchestrating and configuring all the entities used by a network slice. This role is played by the FMNC described in Section 4.6.

4.5 | Function allocation and placement

If service function chaining defines the set of network functions (or middleboxes in the legacy jargon) that have to be traversed by the data traffic in a network slice and how to chain them (ie, how to ensure that the traffic traverses the different functions in the right order), their actual instantiation in the network is another part of the problem. Currently, with hardware middleboxes and their fixed location in the infrastructure network, flows are routed through the chain using static configurations. This approach clearly lacks flexibility, and it is certainly prone to configuration errors. The emerging NFV technology enables the paradigm switch from hardware to software packet processing, with the possibility of deploying a network function everywhere in the network. The flexibility provided by NFV (and SDN) comes at a price, while with legacy middleboxes, QoS problems were tackled by overprovisioning the network; using the NFV/SDN approach, the QoS management can be managed in a more efficient (but complex) way.

The increased flexibility can be used for many purposes, ranging from cost reduction because of better infrastructure utilization to more efficient and fine-grained network features. Enhanced mobility management schemes, for example, can be more effectively implemented by using this approach. Specific mobility-related functions may be located closer to the actual user locations and possibly relocated upon massive user mobility to always provide the best possible QoS.

Also radio functions may be allocated and moved flexibly across different network locations. Traditionally, service function chains in mobile networks only include elements that come downline the P-GW (eg, firewalls and TCP optimizers) because the digital signal processing hardware could not be detached from the physical base stations. The C-RAN concept extends the possibility of having function chains also for the baseband part.

The ecosystem of possible VNFs that have to be orchestrated within a network slice, each one with heterogeneous constraints to be fulfilled, calls for QoS-aware VNF orchestrators. A QoS-aware VNF orchestrator should place VNF into the right physical machines of a data center to minimize the used resources while guaranteeing the service level agreed upon for a given network slice.

4.6 | Flexible control

With the introduction of FMNCs, future 5G networks will bring the concept of *network programmability* beyond SDN. While SDN splits routing and forwarding capabilities in a switch using an SDN controller, the FMNC performs such split between *logic* and *agent* for any network function in the network. That is, the SDN principles are extended to all control, data plane, and management functions usually deployed in mobile wireless networks, which can ultimately be divided into 3 categories: (1) control plane functions, (2) data plane functions, and (3) wireless control functions.

The former points are a rather natural extension of the application of SDN principles, while the latter captures the key aspect of a FMNC: wireless control functions will not be implemented any more in specialized hardware (eg, LTE evolved NodeB), but rather be a piece of software. Therefore, many functions such as channel selection, scheduling, modulation and coding scheme selection, and power control will be provided using a software-defined approach. All these functions are performed by a (virtualized) programmable central control, which provides very important benefits for the operation of the mobile network.

The advantages are manifold. The first one concerns the increased flexibility of the network, one of the current problems that network operators are facing today in their wireless equipment (besides high associated cost). By leveraging the programmability of the FMNC approach, operators will be able to match their needs by simply reprogramming the controller, thus reducing costs. This approach also allows scaling up and down virtual functions, enhancing reliability as well. The flexibility is exposed not only to network operators but also to third parties that can acquire network resources fulfilling a predefined service-level agreement. Programmability also allows customizing the network, enhancing the quality of experience perceived by users.

The FMNC approach implies having a unique management point for the network: a logical centralized controller that homogenizes different network technologies (see Figure 4). By controlling a reduced number of FMNCs, network operators reduce the complexity of the network management. Dense wireless networks, as envisioned in 5G, are especially favored by the FMNC approach: the management of user mobility schemes and dynamic radio characteristics is in charge of the FMNC, which can use especially tailored algorithms according to the network slice they are deployed in. Moreover, if needed, VNFs can be deployed close to the users (ie, an automotive network slice), reducing their experienced latency.

New services can hence be enabled by just modifying the controller functions: services that were not initially included by an operator in its architecture design can now be introduced by implementing service-specific enhancements. The FMNC behavior can also be modified to meet specific needs of the application or to better adapt to a specific scenario. A good example is the management of base station schedulers: as the FMNC has a global view of the network slice, it can optimize the scheduling algorithms and the resource allocation across them. This concept can be extended to the resource control across network slices. The flexible mobile network controller allows the optimization of network utilization: a network infrastructure provider may allocate unused resources to demanding network slices, provided that the service-level agreement is satisfied for all the hosted network slices.

Another possible use of the FMNC is mobility management. As stated above, the FMNC is an extension of the SDN concept to any kind of network function in the mobile network. So, a straightforward amendment of the SDN dialect, capable of directly handling General Packet Radio Service Tunnel Protocol tunnels, may be used to directly control the Serving Gateway (S-GW) and Packet Data Network-Gateway (P-GW) entities of the network. However, the same idea can be used to directly control other low-level user flows, steering traffic through network functions implementing the C-RAN architecture (see Section 3.4). That is, one centralized flexible application *logic* can control heterogeneous network functions through specialized interfaces.

Therefore, the FMNC, following the SDN principles, has northbound and southbound interfaces. The northbound interface is used by FMNC *applications* to exchange high-level messages with the controller. The FMNC applies these high-level commands to the underlying SDN-/NFV-based networks through southbound ones that are used to actually configure them. With FMNC, service providers will be able to fit the equipment to their needs by simply reprogramming the controller using well-defined

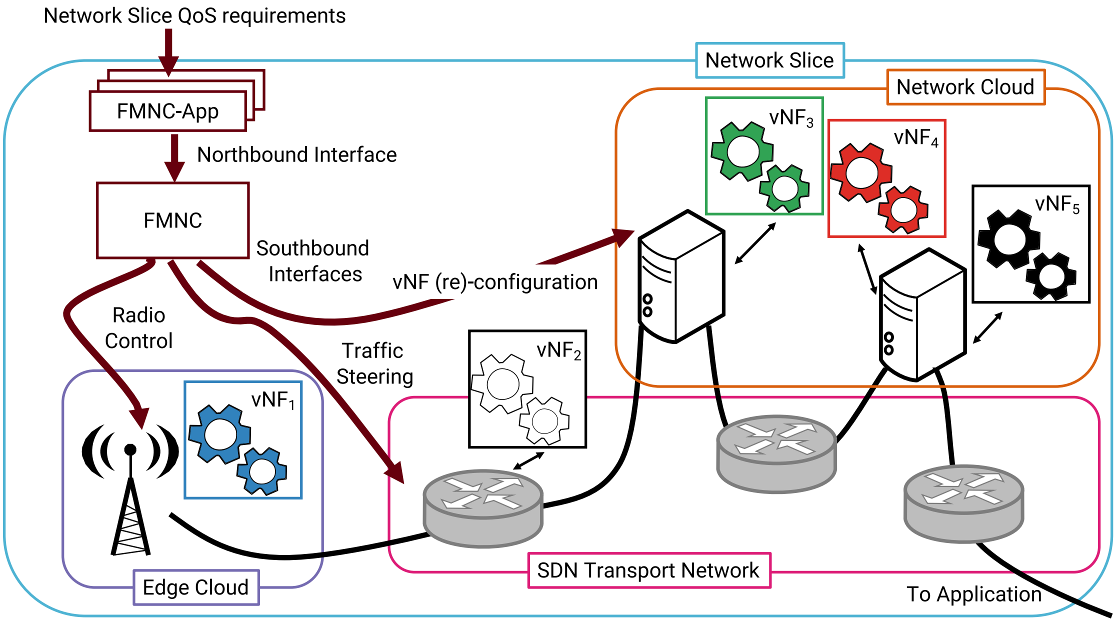

FIGURE 4 An example of the flexible mobile network controller (FMNC). QoS, quality of service; VNF, virtual network function

application programming interfaces, and thus enabling new service within a very reduced implementation, test, and deployment footprint.

5 | CURRENT ACTIVITIES

The attention and the effort around 5G networks have hugely increased in the last years with the emergence of worldwide initiatives with the aim of defining the new architecture, specific technologies, and solutions by 2020. The most relevant ongoing projects are as follows:

1. *5G-PPP* is a European joint initiative between the European information and communication technology industry and the European Commission with the aim of rethinking the infrastructure and creating the next generation of communication networks and services that will provide ubiquitous super-fast connectivity and seamless service delivery in all circumstances. It is composed of 19 different projects (Flex5Gware, 5G-XHaul, 5G-Ensure, METIS-II, Euro 5G, 5G NORMA, Charisma, Sesame, Selfnet, CogNet, Virtuwind, 5GEx, Fantastic 5G, Coherent, SONATA, Superfluidity, 5G Crosshaul, mmMagic, and Speed5G) where industries and academia's members collaborate. In July 2016, they released a white paper[34] that is focused, in particular, on the definition of the key points of the overall 5G architecture. After identifying the 5G requirements, it provides a first preliminary logical and functional architecture ranging from the physical to the management and orchestration layer. Wider attention is placed on softwarization (including NFV/SDN) in 5G, which is seen as an important enabler for the next communication network.
2. *5G Americas* is a wireless industry trade organization composed of leading telecommunications service providers and manufacturers and voice of 5G and LTE for the Americas. It focuses its efforts to advocate the advancement of LTE wireless technology and its evolution beyond 5G, throughout the ecosystem's networks, services, applications, and wirelessly connected devices in the Americas. In November 2016, 5G Americas released a white paper[35] that details network slicing implementation relative to 5G technologies, recognizing this new concept as one of the most important to be able to meet the different 5G use cases and requirements, including scalability and flexibility.
3. *IMT-2020 (5G) PG* in China is a program embarked by the International Telecommunication Union Radiocommunication Sector to develop the new International Mobile Telecommunications system and 5G. It is the major platform to promote the research of 5G in China, and its members include the leading operators, vendors, universities, and research institutes in the field of mobile communications. In September 2016, they released first-round results of the 5G Technology R&D Trial focused on the main key technologies for 5G, such as mMIMO, novel multiple access, new multicarrier, high-frequency communication, network slicing, mobile edge computing, control/user plane separation, and network function reconstruction. The results prove that the implementation of the above technologies leads to supporting the diverse 5G requirements, such as gigabit-per-second user experience data rate, millisecond-level end-to-end latency, and 1 million connections per

square kilometer. The next step will be focused on technical schemes of 5G air interface and network and system trial. The second-round results are expected by the end of 2017.[36]

4. *5G Forum* is an organization founded in 2013 in Korea. It is composed of mobile networks operators, global manufacturer, research institutes, universities, and governments. Its goal is to assist in the development of the standard and contribute to its globalization. By 2020, the South Korean government intends to commercially deploy 5G mobile telecommunication technology for the first time in the world, and they are planning to test 5 core 5G services during the Pyeongchang 2018 Winter Olympics, such as mobile 3-dimensional imaging, artificial intelligence, high-speed services, and ultra- and high-definition resolution capabilities.
5. *5GMF* is a Japanese entity founded in 2014 with the aim of contributing to the development of the use of telecommunications. In July 2016, they released an updated white paper[37] proposing 2 key concepts for 5G: satisfaction of end-to-end quality and extreme flexibility. The former means providing every user satisfactory access to any application, anytime, anywhere, and under any circumstance, while the latter is the feature of communications systems that will allow 5G to always achieve end-to-end quality. Furthermore, it identifies 2 key technologies to support the proposed concepts: advanced heterogeneous network and network softwarization and slicing.
6. *3GPP SA2, SA5, and TSG-RAN* have grown in parallel to the aforementioned large-scale initiatives working groups whose standardization activities are focusing on specific technologies and solutions that aim at addressing some specific requirements imposed on mobile networks by new services or scenarios. In particular, SA2 is in charge of identifying the main functions and entities of the networks, how these entities are linked to each other, and the information exchanged. SA5 will specify the requirements, architecture, and solutions for provisioning and management of the future 5G network (RAN, Core Network (CN), IP Multimedia Subsystem (IMS)) and its services. Their consistent integration with the radio architecture elements is defined by TSG-RAN.

We can easily understand that the worldwide attention and effort on defining and developing the new 5G network is enormous. It is worth noticing that the more relevant ongoing projects agree on the key requirements the new network will need to provide and on the key technologies and enablers, even if each of them targets a different goal.

6 | CONCLUSION

The architecture and the operation of future 5G networks have yet to be defined. However, there is already rough consensus on what the fundamental building technologies will be and where future 5G network should lead us. We have reviewed the most important enabling technologies, describing how they can be used to achieve the goals envisioned for 5G networking by the most prominent fora.

The goals of 5G networks are being addressed by applying novel concepts to the legacy wireless networks. We have listed many of them, underlining the interaction between the enabling technologies and the final goal. Despite that many current softwarization and virtualization technologies are considered to be overhyped, we remark their fundamental contribution toward the goals of 5G by placing them into the new 5G concepts landscape.

This landscape is, however, still blurred. In this paper, we made an effort to specify the current research works, shedding light on the future trends that will eventually build the 5G networking technology.

ACKNOWLEDGEMENTS

This work has been performed in the framework of the H2020-ICT-2014-2 project 5G Novel Radio Multiservice adaptive Network Architecture (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 herein. This work has also been performed in the framework of the H2020-ICT-2014 project 5GExchange (5GEx) (Grant Agreement No. 671636), which is partially funded by the European Commission. This information reflects the consortium's view, but neither the consortium nor the European Commission is liable for any use that may be done to the information contained herein.

REFERENCES

1. 5G-PPP. 5G Vision. The 5G Infrastructure Public Private Partnership: The Next Generation of Communication Networks and Services. http://bit.ly/1GZuqHW. Published February 2015.
2. 4G Americas. 4G Americas' Recommendations on 5G Requirements and Solutions. http://bit.ly/2pCiGvv. Published October 2014.
3. IMT. 5G Vision and Requirements. https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf. Published May 2015.

4. 5G Vision and Requirements of 5G forum. https://www.itu.int/dms_pub/itu-r/oth/0a/06/R0A0600005F0001PDFE.pdf. Published February 2014.
5. 5GMF. 5GMF Activities for 2020 and Beyond in Japan. https://5g-ppp.eu/wp-content/uploads/2015/10/2-20151020_5GMF_K.Satoh_.pdf. Published September 2015.
6. 3GPP SA2—Architecture. http://www.3gpp.org/specifications-groups/sa-plenary/sa2-architecture
7. 3GPP SA5—Telecom Management. http://www.3gpp.org/specifications-groups/sa-plenary/sa5-telecom-management
8. 3GPP TSG—Radio Access Network. http://www.3gpp.org/specifications-groups/ran-plenary
9. Monserrat JF, Mange G, Braun V, Tullberg H, Zimmermann G, Bulakci Ö. METIS research advances towards the 5G mobile and wireless system definition. *EURASIP J Wirel Commun Netw*. 2015;2015(1):53-68.
10. Osseiran A, Boccardi F, Braun V, et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. *IEEE Commun Mag*. 2014;52(5):26-35.
11. Nokia Networks. Network Architecture for the 5G Era. http://bit.ly/2qugPsC. Published March 2015.
12. Ziegler V, Theimer T, Sartori C, et al. Architecture vision for the 5G era. Paper presented at: 2016 IEEE International Conference on Communications Workshops (ICC); May 2016; Kuala Lumpur.
13. Ericsson. 5G Radio Access. http://bit.ly/2qurxzn. Published June 2014.
14. Samsung. 5G Vision. http://bit.ly/2quqvTX. Published February 2015.
15. Huawei. 5G Network Architecture—A High-Level Perspective. http://bit.ly/2q4PZ9b. Published July 2016.
16. Rost P, Banchs A, Berberana I, et al. Mobile network architecture evolution toward 5G. *IEEE Commun Mag*. 2016;54(5):84-91.
17. Marsch P, Da Silva I, Bulakci O, et al. 5G radio access network architecture: Design guidelines and key considerations. *IEEE Commun Mag*. 2016;54(11):24-32.
18. Monserrat JF, Droste H, Bulakci Ö, et al. Rethinking the mobile and wireless network architecture: The METIS research into 5G. Paper presented at: 2014 European Conference on Networks and Communications (EuCNC); June 2014; Bologna.
19. Agyapong PK, Iwamura M, Staehle D, Kiess W, Benjebbour A. Design considerations for a 5G network architecture. *IEEE Commun Mag*. 2014;52(11):65-75.
20. NEC. Network Evolution Toward 2020 and Beyond. http://bit.ly/1JCRP4n. Published March 2015.
21. Larsson EG, Edfors O, Tufvesson F, Marzetta TL. Massive MIMO for next generation wireless systems. *IEEE Commun Mag*. 2014;52(2):186-195.
22. Open Networking Foundation. Software-Defined Networking: The New Norm for Networks. https://www.opennetworking.org. Published April 2012.
23. NFV ISG. Network Functions Virtualisation (NFV)—Network Operator Perspectives on Industry Progress. http://bit.ly/1oAZ7yv. Published October 2013.
24. NGMN Alliance. Suggestion on Potential Solutions for C-RAN. http://bit.ly/2qJnpLN. Published January 2013.
25. Fodor G, Dahlman E, Mildh G, et al. Design aspects of network assisted device-to-device communications. *IEEE Commun Mag*. 2012;50(3):170-177.
26. Tehrani MN, Uysal M, Yanikomeroglu H. Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions. *IEEE Commun Mag*. 2014;52(5):86-92.
27. ETSI GS MEC 001. Mobile Edge Computing (MEC) Terminology. http://www.etsi.org/deliver/etsi_gs/MEC/001_099/001/01.01.01_60/gs_MEC001v010101p.pdf. Published March 2016.
28. ETSI GS MEC 002. Mobile Edge Computing (MEC) Technical Requirements. http://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/01.01.01_60/gs_MEC002v010101p.pdf. Published March 2016.
29. ETSI GS MEC 003. Mobile Edge Computing (MEC) Framework and Reference Architecture. http://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.pdf. Published March 2016.
30. NGMN Alliance. SmallCells: Recommendations for Small Cell Development and Deployment. Deliverable D5. http://bit.ly/2qjr95M. Published September 2015.
31. John W, Pentikousis K, Agapiou G, et al. Research directions in network service chaining. Paper presented at: 2013 IEEE SDN for Future Networks and Services (SDN4FNS); November 2013; Trento.
32. METIS. Deliverable D2.3. Components of a New Air Interface Building Blocks and Performance. https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D2.3_v1.pdf. Published April 2014.
33. NGMN Alliance. NGMG 5G White Paper. http://bit.ly/19gv3jU. Published February 2015.
34. 5G-PPP. View on 5G Architecture. http://bit.ly/2qJpvvr. Published July 2016.
35. 5G Americas. Network Slicing for 5G Networks and Services. http://bit.ly/2ri7twt. Published November 2016.
36. IMT-2020 (5G) Promotion Association. http://www.imt-2020.cn
37. 5GMF. 5G Mobile Communications Systems for 2020 and Beyond. http://bit.ly/2qJcB0f. Published July 2016.

How to cite this article: Bega D, Gramaglia M, Bernardos Cano CJ, Banchs A, Costa-Perez X. Toward the network of the future: From enabling technologies to 5G concepts. *Trans Emerging Tel Tech*. 2017;e3205. https://doi.org/10.1002/ett.3205