Radio Access and Spectrum
FP7 - Future Networks Cluster
http://www.ict-ras.eu/

5G radio network architecture

This white paper is an outcome of the Radio Access and Spectrum cluster projects funded by the European Commission under the Seventh Framework Programme in the area of Future Networks. It presents a view from the European research community on the architecture aspects of 5G mobile and wireless communication systems.

1. Introduction

The use of mobile communication networks has increased significantly in the past decades, in terms of complexity of applications, their required capacities, and heterogeneity of device types. So far, this trend has always been met by significant technological advancements and will continue to increase. By 2020, Europe has to pave the way for a new generation of converged wired and wireless communication networks, which has to be developed and deployed to move forward to a future networked society. In this white paper, we present our perspective on such a 5G radio access network and focus especially on the arising challenges and new technologies that enable us to meet these challenges.

Looking back at the development of 3G (UMTS, HSPA) and 4G (LTE, LTE-Advanced) it is clear that these generations of mobile networks focused on creating new physical radio transmission schemes in order to meet new capacity requirements. From our point of view, 5G networks should consider both wireless and wired parts targeting a fully integrated solution. Furthermore, in order to address the user-oriented challenges, we foresee a continued evolution of existing functions, e.g., network densification into ultra-dense networks and device-to-device communications, as well as development of new functions such as moving networks and massive machine communications. This requires auto-integration and self-management capabilities well beyond today's self-organising network features, which have to be reflected in the architectural layer to achieve their full potential. Additionally, ultra-reliable communications put very stringent latency and reliability requirements on the architecture.

The white paper is organised as follows. In Section 2 we describe the most important challenges for 5G networks, including the rise of number of devices, increased requirements for capacity, energy efficiency, infrastructure issues and varying service requirements and characteristics. In Section 3 we present new technologies that will enable us solve these challenges, with a focus on network function virtualisation, cooperative communications, automated network organisation, flexible backhauling, as well as advanced traffic management and offloading. We conclude with some final remarks in Section 4.

2. Challenges

In this section we give an overview of the main issues challenge for 5G systems, as driven by the fast changing mobile network evolution and the forecast expansion of use cases and applications.

2.1 Broad variation of requirements and service characteristics

The main challenges for 5G system are the continued evolution of **mobile broadband** and the addition of new services e.g., massive sensor communication and vehicular-to-anything communication, requiring shorter set-up times and delay, as well as reduced signalling overhead and energy consumption [Bal13]. Mobile broadband of the future will have significantly increased traffic volumes and data transmissions rates, but also many more use cases. They include not only traffic between humans and between human and the cloud, but also between humans, sensors, and actuators in their environment, as well as between sensors and actuators themselves. Some new key applications with disruptive characteristics follow [METIS project https://www.metis2020.com/].

Firstly, **massive machine communications** (MMC) is envisioned, whose main challenges are [Fal13, Oss13]: i) to support 10-100 times more devices than today; ii) to allow very long battery lifetimes (on the order of 5+ years) of the wireless device; iii) to incur minimum signalling overhead; iv) to enable low-cost wireless devices; v) to support efficient transmission of small payloads with fast setup and low latency. At the same, it is desirable to have 99.999% coverage, while energy consumption and cost for the infrastructure should not increase. Despite some initial works done, e.g., in 3GPP [3gpp22.368], the service requirements for these machine type communications to realise the Internet of Things (IoT) are still not fully understood. This is mostly due to the multidisciplinary nature of IoT applications and the current lack of truly “massive” and large scale deployments of smart objects in an economically sustainable manner.

Secondly, **safety-critical domains**, which traditionally had their private infrastructure, will increasingly use mobile broadband networks. Example applications are: Assisted driving via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, with automated breaking in case of accident or adverse traffic conditions ahead; public protection and disaster relief (PPDR) systems; public transportation automation and control, e.g., the International Union of Railways is considering LTE as a dual use technology to complement and extend the “railways flavour” of GSM, called GSM-R, currently standardised in Europe for signalling in high-speed/high-availability trains; automatic control of smart grid elements, e.g., substations and electric vehicle charging stations, to balance network load and mitigate instability caused by, e.g., the introduction of renewable sources of energy. Such safety-critical applications typically require extremely short setup times and low delays. Also, it is important to achieve wide-area coverage, which can be done only through a smooth and efficient integration with alternative technologies, such as satellite communications and aerial base stations [ABSOLUTE project http://www.absolute-project.eu/].

Finally, we can expect a further growth of **mobile cloud-based applications**, which have unique characteristics in terms of latency and bandwidth. In fact, the most complex applications (e.g., speech recognition, navigation) are often, it not always, offloaded to a cloud server, so as to reduce the processing and energy burden of mobile devices. While this effectively makes the smartphone or tablet leaner, it stresses the importance of a reliable, low latency, high bandwidth connection to the Internet.

A final note is made on **video streaming**, which is already the biggest contributor to worldwide traffic today, at least in the fixed part of the Internet, and is expected to shift to mobile broadband connection as soon as the current technologies and billing plans will allow this. Moreover, the future video encoding and playback advances, including 3D, very high quality encoding, 4K resolution, and multi-angle, will further increase the capacity requirements.

Some examples of very diverse requirements for some use cases of business and social interest are reported in the table below.

Requirement	Very strict	Intermediate	Relaxed
High bit-rate	Video equipments (3DTV, real-time streaming devices, remote conference)	Typical applications on smart phones and tablets, V2I	IoT, V2V
Fast mobility	Applications running on smart phones and tablets on the road, V2V/I	Everything else	Home and office appliances, IoT (most)
High reliability	PPDR, IoT (some), V2V/I	Everything else	--
Low latency	Game consoles, IoT (some), V2V/I, PPDR (some)	Web & mobile apps, cloud computing	IoT (some)
Low energy consumption	IoT devices (most)	Smart phones and tablets	Cabled devices

2.2 Energy efficiency

Classical designs for wireless communications, which tend to maximise rate, capacity and coverage, potentially lead to solutions where energy efficiency drops. Energy efficiency is understood from two viewpoints. On the one hand, the energy spent by the infrastructure may increase, implying high operational costs for the operator that will indirectly affect also the invoice of the finals users. On the other hand, some communication strategies require high computational burden at the terminal side having negative impact on battery lifetime. Hence, the intelligent use of energy becomes a major new target in addition to the classical design criteria.

Currently two approaches to reduce energy consumption on the **radio link** exist. First, small cells reduce the distance to the terminal. The main challenges of this approach are related to providing an economic backhaul solution and to minimise the additional deployment cost. The second approach is massive MIMO, where energy is more focused towards the user by means of more directive beams. In this way, less energy is wasted yielding interference for other users at the end. The challenges of massive MIMO include the diffusion of energy due to scattering in NLOS scenarios, limiting the achievable directivity, and the complexity of spatial multiplexing of users. Both in the terminal and at the base station, the goal of minimising the energy consumption per bit will require a paradigm shift in wireless system design to dramatically improve efficiency in terms of power and spectrum usage. Further research on implementation technologies is necessary, focused on low power hardware architectures and energy-efficient signal processing [COST IC1004 http://www.ic1004.org/]. Some approaches have been proposed on multi-hop cooperative networking, and wireless network coding [Car12].

There are further potential savings by operating the **network** with energy-efficiency in mind. Nowadays base stations consume a constant power, regardless of the traffic load. During off-peak traffic hours, small cells are switched off while coverage is maintained by macro-cells. For active base stations serving a single user, following Shannon's theorem, the most energy-efficient situation would be to use the full bandwidth and to reduce power so that the throughput target is met. However, an interference-limited multi-user scenario is more typical in mobile networks. Serving multiple users having different signal-to-interference ratios in a TDMA fashion -such as round-robin-, changing the power dynamically would result in unpredictable interference in adjacent cells. The same holds for OFDMA, implying inhomogeneous interference on different frequency sub-bands. Hence, current PHY and MAC layers design needs technology advances, including dynamic power control that is optimally coordinated among the users and with surrounding cells so that there is proportionality between the traffic and the energy consumption [5GNOW project http://www.5gnow.eu/]. There is a need for network architecture advances required to i) include small cells and larger antenna arrays efficiently into the network design, ii) switch on/off base stations depending on the traffic load; iii) achieve traffic proportionality at PHY layer.

Mobile devices with advanced capabilities such as smartphones or tablets may present important requirements in terms of energy, not only as far as transmission is concerned (which depends for example on the data flows, the type of application or the wireless network topology) but also regarding other components such as CPU, screen or audio devices at the user equipment. The offloading of applications today hosted by the mobile terminals towards the serving base stations or a (micro) data centre may also contribute to energy efficiency [TROPIC project http://www.ict-tropic.eu/]. This way, the execution of resource-hungry applications is shifted to processing elements that have more efficient computational and caching capabilities.

There is also a need to reduce energy consumption in the **backhaul network**, both in RAN and core, in order to reduce network operational costs. Energy efficiency in the backhaul becomes increasingly critical as the access segment of the network consumes up to 90% of the total telecom network energy cost. Historically, this huge number is related to the use of copper; with the increasing use of optical fibre, the energy requirement is reduced. The access network has a distributed (tree) topology to aggregate the traffic. The enormous heterogeneity of fixed and wireless final-drop technologies (i.e. e.g. FTTH, PON, AON, WiFi, WiMAX, UWB etc.) makes economies of scale rather problematic. More unified and standardised fixed access solutions would allow much higher volumes, and thereby higher integration densities, much lower cost and reduced energy consumption. For instance, the use of an active remote node, originally put forward in Ethernet PONs [Chan10], was recently proposed as a common platform for fixed-wireless convergence [HARP project http://www.fp7-harp.eu/]. This node locates the network intelligence closer to the end-users and performs statistical multiplexing of traffic from fewer users, which allows to handle locally some traffic flows (such as the signalling between cooperative base stations), therefore reducing the backhaul load and enabling a more energy-efficient operation. Moreover, such lower-level aggregation requires less power-hungry circuitry which, in turn, also makes it possible to use renewable energy sources only.

2.3 Network infrastructure

Small access nodes, with low transmit power and no precise planning requirements, are conceived to be densely deployed, resulting in an Ultra-Dense Network (UDN). This approach will improve spectral efficiency by reducing the distance between transmitters and receivers, and to improve macro-cell service by offloading wireless traffic, thus freeing radio resources in the access. Network densification is a way to increase the capacity and data-rate towards 2020.

UDNs are a step further towards low cost, plug-and-play, self-configuring and self-optimising networks. 5G will need to deal with many more base stations, deployed dynamically and in a heterogeneous manner, combining different radio technologies that need to be flexibly integrated. Moreover, a massive deployment of small access nodes induces several challenges such as an adverse interference scenario or additional backhaul and mobility management requirements, which 5G needs to address [CROWD project http://www.ict-crowd.eu/]. 3GPP is currently working on small cells solutions to reduce the inter-site distance [3gpp36.932] but, at the time being, pilot contamination and interference still limit the possible densification. Different levels of coordination/cooperation among small cells are key to enhance the network capacity and keep interference at an adequate level, to manage mobility and spectrum, to ensure service availability and response to non-uniform traffic distribution between neighbouring access points.

With the increasing density of networks, also the backhaul will become more heterogeneous and possibly also scenario-dependent (i.e., fibre, wireless backhaul or other non-ideal types of backhaul might be used depending on their availability). In addition, the connectivity among the network nodes may change in order to allow for fast direct exchange of data between them (which will be challenging in ultra-dense deployments). The heterogeneous backhaul structure will also influence the operation of the radio access networks, e.g. latency differences on backhaul links will impact inter-cell coordination and cooperation algorithms. Therefore, both radio access network and backhaul network need to be aware of limitations and capabilities of each other [TUCAN3G project http://www.ict-tucan3g.eu/]. This may for instance imply an extended SON applied to radio access networks which also uses information provided about the backhaul network.

The required flexibility of the network itself will require new concepts on network management in the backhaul such as the application of Software Defined Networking (SDN) principles in order to achieve fast re-routing and congestion control, mainly in the access part [Kre10, Ahm13]. SDN concepts enable us to adapt the operation of the backhaul network to the needs of the radio access network. For example, the selection of IP break-out anchor points may depend on the current backhaul traffic situation and QoS requirements in the radio access networks. Furthermore, the smaller the cells in the radio access network, the higher the temporal and spatial traffic fluctuations. This implies that also the backhaul network may experience a higher variance of traffic. Besides, current trends suggest that Infrastructure as a Service (IaaS) can be supported by small cells in order to offer innovative proximity services and to enable a series of advantages for end customers. With this approach, energy-scarce, capacity-limited mobile devices can offload highly demanding computational tasks into proximal fixed units or use them for storage. This entails that novel mechanisms are needed to efficiently allocate resources, understood in a wide sense (radio/computation/storage/energy), including contextual information metrics and clustering techniques for small cells.

Another important aspect in the network infrastructure is related to the exposure of end users to electromagnetic field (EMF).There is today a public concern concerning EMF induced by wireless networks. By reducing the distance between receivers and transmitters, small cells enable the minimisation of the power emitted by the mobiles phones and the total EMF exposure because, currently, the most important contribution is linked to the user equipment. 5G architecture -combining small cells, heterogeneous networks and offloading- should inherently enable minimising the human EMF exposure [LEXNET project http://www.lexnet-project.eu/].

3. New enabling technologies

In this section we introduce the most promising enabling technologies that are expected to be used extensively in 5G radio networks to tackle the challenges identified above.

3.1 Network Functions Virtualisation (NFV)

Network Functions Virtualisation (NFV) [Nfv12] refers to the implementation of network functions in software running on general purpose computing/storage platforms. This approach allows the deployment of network functions in data centres and to leverage from virtualisation techniques. By contrast, the state-of-the-art is to implement network functions on dedicated and application-specific hardware. Hence, the main motivation for NFV is to leverage from the economy of scale of high-volume hardware platforms, to reduce time-to-market and innovation cycles within telecommunication networks through software updates rather than hardware updates, and to exploit novel data centre technology. NFV has recently attracted significant interest from the industry, which has led to the creation of a dedicate industry study group at ETSI.

Implementing network functions in software on standard IT platforms will allow for new flexibilities in operating and managing mobile networks. In mobile networks, NFV is currently discussed in the context of virtualising the core network [You13] as well as logically centralising the base band processing within the RAN, so-called Cloud RAN (C-RAN) [Gua10]. C-RAN still requires specialised hardware in data centres in order to satisfy the hard real-time requirements in mobile networks. Furthermore, C-RAN does not allow for a functional decomposition which implies that the RAN functions are decomposed in individual modules which may then be managed and operated on different (virtual) machines and provided by different software vendors. While C-RAN enables both full centralisation and distribution of (digital) RAN functions, this needs not to be the case with a general NFV implementation where only a subset of all modules may be implemented centrally or the radio access points implement all functions based on general purpose hardware [iJOIN Project http://www.ict-ijoin.eu].

Another important topic in mobile networks which may be improved by implementing network functions in a data centre is resilience. This allows for re-assigning functions between either virtual or real machines. For example, rather than running functions in a data centre, they may be run in a Radio Access Point (RAP) at lower computational complexity [TROPIC project http://www.ict-tropic.eu/]. Furthermore, NFV and implementing mobile network functions in data centres allows more flexibility in terms of resource management, assignment, and scaling. This has also an impact on the energy efficiency of networks as only the required amount of resources may be used and over-provisioning of resources can be avoided. This resource orchestration could reuse management algorithms already developed in the IT world in order to exploit resources as efficiently as possible.

As mentioned, NFV is already applied on core networks and first trials are performed demonstrating that critical mobile network functions such as MME, HGW/PGW, or HSS can be implemented on standard IT platforms. A critical enabler of this development is, besides virtualisation technologies, the availability of high-speed IP networks and the possibility to manage them more flexibly through SDN. Interest on the latter is confirmed by the recent foundation of a working group on wireless and mobile within the Open Networking Foundation, which is the organisation that has standardised OpenFlow . In case of RANs, NFV may be more difficult to apply as it is either applied directly within network nodes such as RAPs or at more centralised locations which requires high-performance connections between RAPs and data centres. Those connections may not be available at all locations which imposes new challenges on implementations of NFV in RAN and managing networks composed of heterogeneous network nodes (macro-, metro-, and pico-cells), heterogeneous backhaul-connectivity (optical fibre, DSL, wireless), as well as heterogeneous location of RAN functions [Sab13, Ber13, Ros14].

3.2 Cooperative communications

Recently, multi-hop relay communication has been gaining global acceptance as one of the most promising technologies in next-generation wireless cellular networks [She09, Wij09, Loa10]. Present-day cellular systems have a single direct link between the base station and the terminal. In a multi-hop wireless network, the communication takes place over one or more links (hops) to form a multi-hop path between the transmitter and the receiver. *Multi-hop cooperative networks* have the capability to increase the capacity density and to reduce energy consumption by bringing the RAN closer to the end-user [ABSOLUTE project http://www.absolute-project.eu/]. Compared to the existing layered protocols, which include mechanisms such as retransmissions or multiple acknowledgements, multi-hop networks overcome such inefficiencies and prevent these mechanisms from scaling as required for high capacity density access networks. However, **multi-hop networks** often suffer a throughput penalty since the nodes operate in a half-duplex mode and therefore necessarily introduce inefficiency in spectrum usage, as multiple time slots are required to receive and then relay the information. Another problem is the latency due to multiple hops. On the other hand, **wireless network coding** has the potential of naturally adapting to problems related to dense, cloud-like, massively-interacting networks of nodes, since it is an example of the general concept of “network-aware physical layer”: functions like routing, conventionally performed at high layers of the protocol stack, are more efficiently carried out at the physical layer, which has the capability of processing signals directly and without loss of information. By looking at multiple communication flows jointly, instead of a single flow at the time, wireless network coding can overcome the efficiency and latency issues mentioned before for general multi-hop networks.

Furthermore, storing the data at the edge of the network, i.e., **caching**, will be a promising way of reaching high capacity in 5G systems [MOTO project http://www.fp7-moto.eu/]. In fact, in spite of increasing the wireless network capacity by employing advanced PHY techniques, high data rates might still not be achievable due to the limited backhaul. Since base stations have to serve users by bringing their requested content from the Internet through the backhaul, the capacity of this backhaul should also be in the same order of the wireless network capacity, in order to avoid rate bottlenecks, which are especially evident in densely deployed small-cell scenarios where low-rate backhaul links are preferred instead of fibre-optic connections due to deployment and operational costs. Equipping (small-cell) base stations in such deployments with storage units and proactively caching the content definitely helps to mitigate this bottleneck [Bas13] and benefits can be brought further by enabling direct access of UE caches and leveraging social networks via device-to-device (D2D) communications. Moreover, when using non-ideal backhaul, e.g. xDSL, the throughput is very asymmetrical, strongly limiting the user traffic and affecting the latency of the inter-base station communication needed for most of the applicable coordinated multipoint (CoMP) techniques. Theoretically, this limitation can be overcome by considering **over-the-air mesh communication** for signalling between base stations [COST IC1004 http://www.ic1004.org/]. However, this type of communication inherently requires research on many-to-many network architectures and protocols, which are fundamentally different from the existing one-to-many approach.

The large user data traffic demand in conventional wireless communication systems tends to increase the number of required access points or base stations per area in a network, inducing an adverse scenario where communications are severely affected by interference. One way of improving the spectral efficiency of the system is the use of **advanced coordination / cooperative schemes** among transmitters in order to combat the generated interference. In LTE-Advanced and its evolutions these schemes are known as CoMP. Coordination of transmitters by exchanging control-plane messages and interference alignment-based transmission solutions are under investigation [5GNOW project http://www.5gnow.eu/]. Another way of improving the spectral efficiency is to enhance the spatial reuse of radio resources when D2D communication is allowed for terminals in the same radio range. This solution generates additional interference, but the involved terminals employ much less power than the base stations, which means that a lower level of interference can be expected.

Also concerning the limitations related to the enhanced inter-cell interference coordination (eICIC) mechanisms in the case of small cell deployments, current research suggests a more flexible interference coordination approach in time-frequency and power domains which, when used in conjunction with a simple ICIC-based Radio Resource Management (RRM), has the merit of significantly increased user throughput. For instance, each base station can create its own collaborating micro-cluster, composed by the (mutually) interfering base stations in downlink: the information on resource and power allocation is distributed by each cell to its **micro-cluster peers** to optimise access to common resources. This could be further extended by sharing within each micro-cluster, e.g., the user location or other relevant information.

We conclude by noting that the mobile network infrastructures are currently evolving to reduce the range, hence the size and complexity, of base stations, while increasing the number and bandwidth of the physical connections between smaller cell sites. The wide deployment of optical communications networks, with fibre connections closer to the end users, make sense also for wideband connections between small cells, changing the current basic concept of traffic-scaled cellular deployment to a modern view of opportunistic spectrum-access based cooperative networking [Car12]. In conjunction to this cooperative small-cells scenario, the terminal will be acting as a local access enabler, managing radio communications not only from the user but also from surrounding smart objects. Radio network architectures can then consider the roaming user device (on the bus, in the street, inside the car, at home, etc.) as an **IoT relay node** able to provide coverage extension and to act as a gateway to the Internet for the IP-enabled smart objects.

3.3 Automated Network Organisation

Current trends in the definition of 5G wireless systems rely on evolving heterogeneous networks where macro-cells are overlaid with small cells to deliver improved spectral efficiency and coverage within an area. Such coexistence imposes difficulties to the traditional network planning, where new site locations are set based on expensive and limited tests often based on propagation models that may be inaccurate. Furthermore, achieving and maintaining optimal performance in future cellular systems will become virtually impossible with manual configuration, optimisation, and maintenance due to their incremental densification, which involves a rise of the number of parameters involved, as well as latency and accuracy limitations. In the past, automated network organisation has been addressed via, e.g., SON proposals, which have emerged as a possible solution for the issues mentioned above. SON has shown itself as a paradigm that can reduce OPEX and CAPEX while yielding optimal performance in LTE [Ham12].

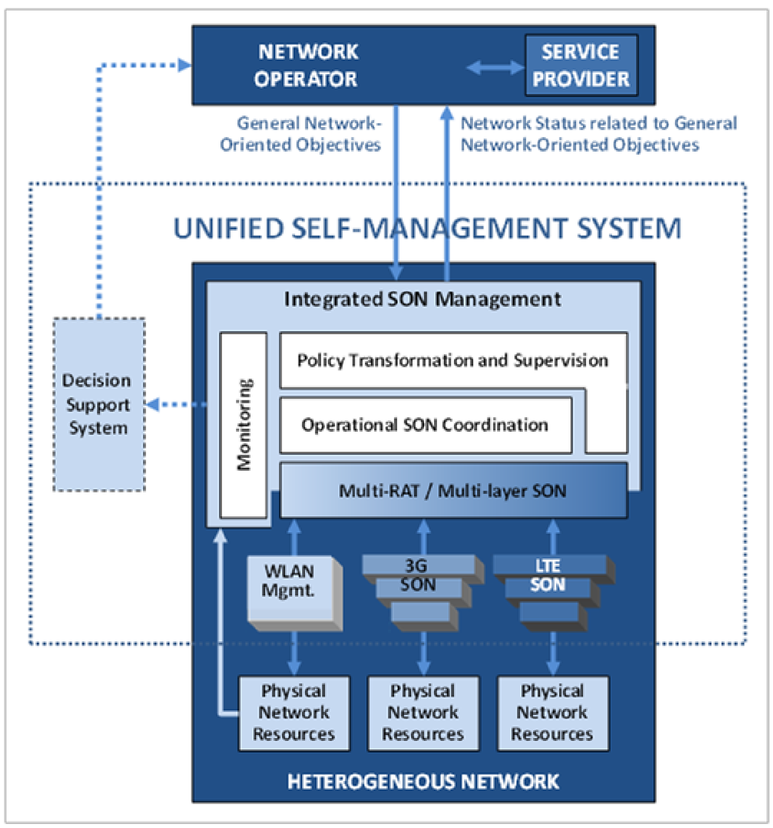
Self-configuration, automatic neighbour relation, self-organised carrier selection, and self-healing mechanisms are examples of automated network organisation techniques. Self-configuration, for instance, (i.e., automatic configuration of emission power, antenna tilt, etc.) allows newly added base stations to be self-configured in line with a "plug-and-play" paradigm, which is particularly important in the case of small cells. In addition, as far as costs for network planning and deployment are concerned, such techniques can also alleviate the burden of operators coming from manually managing neighbour relations. Indeed, automatic neighbour relation (ANR) was the first SON technique to be included in the LTE specifications. In contrast to legacy networks, where a significant amount of time and resources is needed to identify failing base stations and fix such a situation, 5G systems should have self-healing functions built in. This will allow them to detect failing base stations immediately and to take further measures while ensuring no significant degradation of service for the users. Finally, self-organising carrier selection and interference management will help to reduce cost and improve service reliability.

Automation seems the only sensible approach to cost-effective management of future operationally complex heterogeneous mobile access networks. The key enabling element is a unified self-management system, which controls the complex network environment as a single entity. This self-management system shall enable the network operator to specify network-oriented objectives regarding, e.g., desired service coverage, resource efficiency and quality of experience, and shall effectuate these objectives in the unified and automated optimisation of the underlying integrated access networks. The self-management shall then perform resource management and tune the radio parameters of mobile access networks in line with network-operator-defined targets.

Such a unified management system will provide considerable gains to the operators in terms of (i) enhanced resource efficiency, implying increased capacity and hence delayed investments in network expansions and/or equipment upgrades; (ii) improved manageability and hence lower operational costs; (iii) enhanced performance in terms of service availability, seamless session continuity and user-level quality of experience and (iv) enhanced configuration flexibility, therefore supporting different resource utilisation strategies and fast transitions in case of redefinition of network targets. Interestingly, automatised network controllers and SDN techniques offer a promising practical paradigm to implement a unified management system. Furthermore, an SDN-based approach goes beyond a mere unified management system, and indeed provides tools for jointly orchestrating radio and backhaul resources *on demand*, working at the time scale of IP flows [CROWD project http://www.ict-crowd.eu/]. However, the applicability and suitability of SDN for future dense wireless networks is currently under investigation, and industry-grade software and interfaces for SDN operation are still to be studied and designed, as mentioned in Section 3.1.

The following is a description of the two key elements for future unified self-management system [SEMAFOUR project http://www.fp7-semafour.eu/].

The first element is the **integrated SON Management**, as presented in the figure below. The top part depicts a service provider, which maintains a Service Level Agreement (SLA) with a network operator, contractually formalising their agreement regarding performance and tariffs. Integrating such performance obligations with its own business strategy, the operator formulates its network-oriented objectives and provides these as an input to the integrated SON management layer, which serves as its interface to the self-management system. The key purposes of this layer are (i) to transform these objectives into dedicated execution policies for specific SON functions; (ii) to supervise and coordinate these SON functions; and (iii) to monitor and analyse their performance according to the objectives, providing input to periodic operator reports, SON management, SON functions and decision support systems.



The second element is **Single / multi-RAT / layer SON functions.** These will reside at the functional layer below the integrated SON management (see figure above). These SON functions control the physical network resources in different RATs and layers and can be implemented in a distributed fashion in the network elements, or in a centralised fashion in the network management system. Numerous SON functions have been developed so far, including mobility robustness optimisation (MRO) and mobility load balancing (MLB) SON functions. They mostly focus on single-RAT/layer scenarios. Future SON functions will be needed to target multi-RAT/layer SON functions, addressing amongst others advanced traffic steering between WiFi and 3G/LTE cellular layers, dynamic allocation of spectrum over RATs and layers, and the automated (de)activation and tuning of site sectorisation.

3.4 Flexible backhauling

3G and 4G use different backhauling technologies (e.g. optical fibre, microwave links or even a satellite link) but in every case the backhaul is seen as providing "enough" QoS (quality of service) and as much capacity as the RAN may require. 5G RAN becomes more heterogeneous, thus requiring flexible topology and performance from the backhaul [e.g. iJOIN Project http://www.ict-ijoin.eu]. The backhaul for 5G needs higher flexibility also to unlock the potential of increased, more efficient and more flexible spectrum usage and to support new applications.

Carrier aggregation and the use of more frequency bands will lead to an increased mobile spectrum usage in 5G. **New radio technologies** will enable significantly higher spectrum efficiency by using inter-site coordinated multipoint, small cells and massive MIMO. Sharing of both, the spectrum and the mobile infrastructure will enable statistical multiplexing gains for the spectrum usage as well as increased density of base stations per operator without the need for new sites. Both approaches need a shared backhaul network used by mobile operators. Sharing will further increase the current heterogeneity of the network and will bring new specific requirements. Last but not least, new applications such as the "tactile mobile Internet" [Fet13] for the support of IoT will also have a big impact on the mobile backhaul evolution.

The mobile backhaul evolution for 5G is expected to follow five major trends: Open network architecture, end-to-end support for QoS and security, significantly higher data rates, reduced latency and network-assisted synchronisation. **Open networks** enable a shared infrastructure in which multiple operators contribute to a shared overall network. In a first step, operators integrate their proprietary infrastructure into the overall network. Virtualisation tools are then installed enabling the formation of coexisting virtual subnetworks. In a third step, the overall network resources are dynamically distributed among the operators. Distribution can be managed using a neutral broker trading the price of mutual resource utilisation by the sub-networks according to offer and demand.

The 5G RAN needs to verify actively and dynamically the supported QoS and the available capacity in the backhaul. Signalling between the backhaul and/or real-time QoS measurements performed by the RAN will be essential to guarantee QoS to end users. The native support for

MMC will require lower latency. Guaranteed bandwidth, latency and **end-to-end security** unaffected by other users' demands will be needed. For lower latency, for example, instantaneous handover is needed. But the data transfer over X2 or similar evolved interfaces to the new serving base includes private user data. This is also true for coordinated multipoint [Fri12]. Obviously, the current security architecture needs to be revised.

Data rates of the 5G air interface will be increased by a factor 1000, compared to LTE. The backhaul will follow this trend, obviously. Even more, inter-site coordinated multipoint enables a gain of factor 3 by exchanging user data in the clusters between 3 cells at distant sites, on average. Accordingly, the factor becomes 3000 in the 5G backhaul. Several 100 Gb/s per site will be needed [Jun13]. Although technologies exist or are already developed, cost is an issue. Low-cost and high-performance backhaul solutions will also be needed for small cells, both for LOS and NLOS deployment scenarios [SODALES project http://www.fp7-sodales.eu/].

Minimised delay is a driver for the backhaul evolution. Considering LTE, handover latency is due to framing delays and the X2 interface [Dim09], due to the centralised security architecture [Fri12]. A distributed security will be needed to protect private user data while reducing the latency. Further, the hop lengths between nodes can be minimised by distributing the intelligence in the network. Active switches will be placed in all aggregation nodes so that signals can be routed through the shortest path to other ports. Flexible distributed virtualisation for coordinated multipoint is described in [Kre10]. Extended with end-to-end encryption and guaranteed bandwidth, it is a good example what is needed also for the IoT.

Synchronisation is needed for higher spectral efficiency using coordinated multipoint and to minimise delays [SODALES project http://www.fp7-sodales.eu/]. GPS synchronisation is an example of a distributed approach, see [Irm11]. However network operators currently prefer the IEEE 1588 precision time protocol (PTP) over the backhaul because it is applicable also to indoor deployments. Note that the reference clock is passed over several aggregation nodes from a grand master to each base station. Native support for **network-wide synchronisation** is therefore needed in each aggregation node [Jun13].

3.5 Advanced traffic management & offloading

To handle the explosion of mobile wireless data offloading techniques have been proposed to improve the user experience for cellular services in overloaded areas. Offloading techniques towards the end-user either through WiFi (outdoor) infrastructures and femto-cells are currently being applied. New standards such as Selective IP Traffic Offloading (SIPTO), Local IP Access (LIPA) and IP flow mobility (IFO) are being proposed to optimise the data transfer from-to the mobile devices to the Internet.

However, the network densification envisaged in 5G is actually introducing a paradigm shift that the next generation traffic offloading techniques will have to take into account. Network uplink and downlink asymmetry will increase in 5G, hence they will need to be considered as two independent connections. In fact, many mobiles may find more efficient energy and throughput-wise to associate to two different Points of Access (PoAs) for uplink and downlink communications, respectively. More work is required to better understand how to assign traffic to each RAN under realistic network loading models and dealing with diverse types of traffic (e.g., balancing QoS for data, voice over IP and video streaming for instance). Today, data offloading modifies the service rate, which makes the net rate optimisation problem for all users under dynamic traffic offloading very complex. Further offloading is only possible if one uses the user equipment as a relay within a cell, or across several cells, building upon the recent developments in D2D communications in an LTE-A infrastructure. Strategies for D2D path establishment, or for managing opportunistic D2D communications need to be further investigated [MOTO project http://www.fp7-moto.eu/].

Furthermore, 5G networks with the drastic increase in traffic load and number of devices connected will start to experience also a backhaul bottleneck beyond the current data capacity shortage experienced by customers in the wireless access segment. Offloading techniques will also need to increasingly combine and coordinate massive antenna configurations with strategies for decreasing the load on the backhaul, e.g., through femto-catching, out-of-band content loading, and increased D2D opportunistic communications. The densification of the infrastructure includes an additional challenge for effective management of offloading mobility, in terms of network association. To make real-time decisions regarding selective offloading, increased application, device and subscriber awareness are required to effectively manage the whole process. Consistent user experience and service continuity independent of the data offloading solutions implemented (femto, WiFi, opportunistic, IP flow mobility, IP layer management) demands transparent sign-on solutions across managed and heterogeneous network infrastructures. Hence, roaming agreement management across WiFi networks is an important issue. Seamless session handover additionally demands network readiness prior to device readiness.

4. Final remarks

A sudden change of the system requirements is expected in the next years, which needs an equally fast reaction to adapt the network architecture and protocols so as to efficiently support each use case. Unfortunately, the typical duration of the lifecycle from the analysis of requirements to the full-scale deployment of a technology is in the order of 7-10 years in the telecommunications industry, which is incompatible with the desired time horizon of some use case. For example, many IoT applications could bring significant economic and social impact already today since they are based on mature technology (except for communications).

5G solutions will have to enable service-aware optimal coverage, capacity, and reliability with lowest cost and energy consumption [METIS project https://www.metis2020.com/]. Different scenarios may require different grouping of functions to network elements. The trade-off between centralising network functions (whose main benefits are: resource pooling, easier deployment and management, and global optimisation) and decentralising them towards the network edges (achieving faster reaction, incremental commissioning, and potentially lower signalling overhead) must be investigated, e.g., which functions may be combined in common RAN elements to achieve optimal performance.

In addition, it is not only a matter of further optimisation of the radio connection itself. The new class of services and service providers requires a more open and service-aware network structure in order to customise the network resources and the management of the network (or parts thereof). It is also very likely that there is no single network architecture that can support all of the 5G scenarios in a cost-efficient manner: “one size fits all” is likely not viable!

The future network architecture will also be scenario and test case specific, i.e., it may be different in areas with low cell density compared to ultra-dense deployments, such as Mega-Cities. The future RAN architecture will include densely deployed heterogeneous radio access nodes provided by network operators, access nodes privately installed, and even moving access nodes. Many of the nodes will support multiple heterogeneous RATs and software-defined interfaces. Furthermore, data communications and management of the network will be based on a unified all-IP network.

References

[Bal13] Baldemair, R.; Dahlman, E.; Fodor, G.; Mildh, G.; Parkvall, S.; Selen, Y.; Tullberg, H.; Balachandran, K., "Evolving Wireless Communications: Addressing the Challenges and Expectations of the Future," *Vehicular Technology Magazine, IEEE* , vol.8, no.1, pp.24,30, March 2013.

[Fal13] M. Fallgren, and B. Timus (Eds.), “Scenarios, requirements and KPIs for 5G mobile and wireless system,” available online at http://goo.gl/64tKZt, May 2013.

[Oss13] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten H. Tullberg, M. A. Uusitalo, and M.Schellman, “The Foundation of the Mobile and Wireless Communication System for 202 and Beyond,” Proc. IEEE Vehic. Tech Conf, Dresden, Germany, June 2013.

[3gpp22.368] 3GPP TS 22.368, “Service requirements for Machine-Type Communications (MTC); Stage 1,” version 12.2.0, Mar. 2013.

[Car12] N. Cardona (Ed.), “COST IC1004 Position Paper on Horizon 2020,” available online at http://goo.gl/jiHzBv, May 2012.

[Cha10] C. A. Chan, M. Attygalle, A. Nirmalathas, "Remote Repeater-Based EPON With MAC Forwarding for Long-Reach and High-Split-Ratio Passive Optical Networks," *IEEE/OSA Journal of Optical Communications and Networking,* vol.2, no.1, pp. 28–37, January 2010.

[3gpp36.932] 3GPP TR36.932 “Scenarios and Requirements for small cells enhancements for E-UTRA and E-UTRAN,” version 12.1.0, Mar. 2013.

[Kre10] W. Kreher, J. Müller, V. Jungnickel, S. Jaeckel, F. Bauermeister, A. Forck, T.Haustein, “Coordinated Multi-point Transmission COMP Data and Signalling on X2-Interface using Additional VLAN Identifier,” European Patent EP20 100 178 477, 2010.

[Ahm13] H. Ali-Ahmad, C. Cicconetti, A. de la Oliva, V. Mancuso, M. R. Sama, P. Seite, and S. Shanmugalingam, “An SDN-based Network Architecture for Extremely Dense Wireless Networks,” *IEEE Software Defined Networks for Future Networks and Services*, November 2013.

[Nfv12] “Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges & Call for Action,” White Paper, October 2012, Darmstadt, Germany.

[You13] F.Z. Yousaf, J. Lessmann, P. Loureiro, and S. Schmid, “SoftEPC - Dynamic Instantiation of Mobile Core Network Entities for Efficient Resource Utilization,” *IEEE International Conference on Communications*, June 2013, Budapest, Hungary.

[Gua10] H. Guan, T. Kolding and P. Merz, "Discovery of Cloud-RAN," in *Cloud-RAN Workshop*, April 2010.

[Sab13] D. Sabella, P. Rost, Y. Sheng, E. Pateromichelakis, U. Salim, P. Guitton-Ouhamou, M. Di Girolamo, and G. Giuliani, “RAN as a Service: Challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network,” *2013 Future Networks Summit*, July 2013, Lisbon, Portugal.

[Ber13] C.J. Bernardos, A. De Domenico, J. Ortin, P. Rost, and D. Wübben, “Challenges of designing jointly the backhaul and radio access network in a cloud-based mobile network,” *2013 Future Networks Summit*, July 2013, Lisbon, Portugal.

[Ros14] P. Rost, C.J. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, “Cloud technologies for flexible 5G radio access networks,” *IEEE Communications Magazine*, May 2014.

[She09] G. Shen, J. Liu, D. Wang, J. Wang, and S. Jin, “Multi-hop relay for next-generation wireless networks”, *Bell Labs Technical Journal*, Vol. 13, No. 4, pp. 175-194, 2009.

[Wij09] C. Wijting, K. Doppler, K. KallioJarvi, T. Svensson, M. Sternad, G. Auer, N. Johansson, J. Nystrom, M. Olsson, A. Osseiran, M. Dottling, J. Luo, T. Lestable and S. Pfletschinger, “Key technologies for IMT-Advanced mobile communication systems”, *IEEE Wireless Communications*, vol. 16, No. 3, pp. 76 –85, June 2009.

[Loa10] K. Loa, C.-C. Wu, S.-T. Sheu, Y. Yuan, M. Chion, D. Huo and L. Xu, “IMT advanced relay standards [WiMAX/LTE update]”, *IEEE Communications Magazine*, vol. 48, no. 8, August 2010.

[Bas13] E. Baştuğ, M. Bennis and M. Debbah, “Living on the Edge: The role of Proactive Caching in 5G Wireless Networks”, *IEEE Communications Magazine* (to appear).

[Ham12] S. Hämäläinen, H. Sanneck, C. Sartori, “LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency,” John Wiley & Sons (Publisher), ISBN: 1119970679, 2012.

[Fet13] G. Fettweis, “5G - What Will it Be: The Tactile Internet,” Keynote talk at *ICC 2013*, available online at http://goo.gl/DJhg8E.

[Dim09] K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, Y.Timner, Handover within 3GPP LTE: Design Principles and Performance, *IEEE VTC-Fall*, 2009.

[Fri12] M. Fricke, A.Heckwolf, R. Herber, R. Nitsch, S. Schwarze, S. Voß, S. Wevering, “Requirements of 4G-Based Mobile Broadband on Future Transport Networks,” *Journal of Telecommunications and Information Technology*, vol. 2, pp. 21–28, 2012.

[Jun13] V. Jungnickel, K. Manolakis, S. Jaeckel, M. Lossow, P. Farkas, M. Schlosser, V. Braun, “Backhaul requirements for inter-site cooperation in heterogeneous LTE-Advanced networks,” in *Proc. OWITSN Workshop, IEEE ICC*, 2013, pp. 915-920.

[Irm11] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated Multipoint: Concepts, Performance and Field Trial Results,” *IEEE Comm. Magazine*,vol. 49, no. 2, pp. 102–111, 2011.

Acronyms

3GPP	Third Generation Partnership Project
ANR	Automatic Neighbour Relation
AON	Active Optical Network
CAPEX	Capital Expenditures
CoMP	Coordinated Multipoint
CPU	Central Processing Unit
C-RAN	Cloud Radio Access Network
D2D	Device to Device
DSL	Digital Subscriber Line
eICIC	enhanced Inter-cell Interference Coordination
EMF	Electromagnetic Field
FTTH	Fibre To The Home
GPS	Global Positioning System
HGW	Home Gateway
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
IaaS	Infrastructure as a Service
IEEE	Institute of Electrical and Electronics Engineers
IFO	IP Flow mobility
IP	Internet Protocol
IT	Information Technology
LIPA	Local IP Access
LOS	Line of Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MLB	Mobility Load Balancing
MMC	Massive Machine Communications
MME	Mobility Management Entity
MRO	Mobility Robustness Optimisation
NFV	Network Function Virtualisation
NLOS	Non Line of Sight
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditures
PGW	Packet data network Gateway
PON	Passive Optical Network
PPDR	Public Protection Disaster Relief
PTP	Precision Time Protocol
QoS	Quality of Service
RAN	Radio Access Network
RAP	Radio Access Point
RRM	Radio Resource Management
SDN	Software Defined Network
SIPTO	Selective IP Traffic Offloading
SON	Self Organising Network
TDMA	Time Division Multiple Access
UDN	Ultra Dense Network
UMTS	Universal Mobile Telecommunications System
UWB	Ultra Wideband
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access

Acknowledgments

RAS cluster chaired by Paulo Marques (CRS-i - http://www.ict-crsi.eu/)

White paper edited by Claudio Cicconetti (CROWD) with contributions from:

Gerhard Wunder (5GNOW - http://www.5gnow.eu/)
Isabelle Bucaille (ABSOLUTE - http://www.absolute-project.eu/)
Narcís Cardona (COST IC1004 - http://www.ic1004.org/)
Vincenzo Mancuso, Martin Dräxler and Antonio de la Oliva (CROWD - http://www.ict-crowd.eu/)
Tharm Ratnarajah and Lars Dittmann (HARP - http://www.fp7-harp.eu/)
Peter Rost and Albert Banchs (iJOIN - http://www.ict-ijoin.eu/)
Joe Wiart (LEXNET - http://www.lexnet-project.eu/)
Afif Osseiran and Hugo Tullberg (METIS - https://www.metis2020.com/)
Oscar Lazaro and Damien Lavaux (MOTO - http://www.fp7-moto.eu/)
Ejder Baştuğ and Mérouane Debbah (NEWCOM# - http://www.newcom-project.eu/)
Colin Willcock (SEMAFOUR - http://fp7-semafour.eu/)
Volker Jungnickel (SODALES - http://www.fp7-sodales.eu/)
Felicia Lobillo, Antonio Pascual-Iserte, Adrián Agustín, Olga Muñoz, and Mariana Goldhamer (TROPIC - http://www.ict-tropic.eu/)
Josep Vidal, David Chavez, and Francisco-Javier Simó-Reigadas (TUCAN3G - http://www.ict-tucan3g.eu/)