Strategic Research and Innovation Agenda

Expert Advisory Group of the European Technology platform Networld 2020

Pervasive Mobile Virtual Services

July 2016

Table of Contents

Executive Summary

In 5G and beyond networks, the first fundamental change is the availability of processing and storage as core services provided by the network itself. We can compare this next generation network with a distributed computer, where a dedicated operating system controls the distributed virtual infrastructure, offering a common interface to applications. However, the definition of a reference architecture for next generation networks is challenging due to the required Key Performance Indicators (KPIs), which go from the transport of massive volume of data, to the capacity to connect trillions of devices. Current research projects are tackling several challenges towards a 5G network standard, but there are plenty of research activities still pending.

We envision a next generation network beyond 5G, capable of providing the universal availability of apparently **unlimited information services**. This next generation network will take advantage of: (i) new virtualisation techniques, especially focused on network functions; (ii) radio and signal processing advances to use new frequency bands; (iii) optical network improvements to reduce the end-to-end delay and, at the same time, increasing the available bandwidth. The diversity and number of users of the network (including machine-to-machine communications) will explode with this approach, bringing new opportunities to new sectors like the automotive, Industry 4.0, Entertainment, Energy and E-health, the so called *verticals*. We call this new paradigm "**Pervasive Mobile Virtual Services**", and this Strategy Research and Innovation Agenda presents several challenges and recommendations to address such new paradigm. This document is the result of contributions received from more than 140 researchers from industry and academia, to promote the necessary research in Europe to leader these initiatives.

1 Introduction

1.1 Our Vision: Pervasive Virtual Mobile Services

In the previous advances of mobile technology, the change between successive generations has been mainly (but not only) the increase of the capacity in the user radio access. It is true that other technical improvements like improved spectral efficiency, better coverage or packetized information transport were introduced, but these were mainly network internals, and were not fundamentally changing the service model. 5G, though, will imply a radical change in the way we currently understand mobile communications.

A first fundamental change is that the network will incorporate compute and storage services as part of its core services delivered to end-users. The network is no longer a transparent medium that merely transports packets end-to-end. The network will now incorporate, throughout its infrastructure, compute and storage capabilities, managed in an integrated way together with its transport capabilities. The network is therefore transformed into a distributed computer, where processes and applications are dynamically created, moved and deleted, depending on the information flows and customer needs. This is a much more complex service model than just transporting packets end to end, and will have a strong influence on the way telcos understand themselves. Technologies such as cloud computing, fog computing and mobile edge computing (MEC) will have a deep impact, and will be deeply interweaved with network function virtualization (NFV) and software defined networks (SDN).

A second fundamental change is the explosion of internet of things (IoT) involving machine to machine (M2M) communications. Current mobile networks are mainly oriented to human-held terminals and human-oriented applications. This has deep technical implications and has driven the network design to the model we have today. However, IoT will imply a much larger number of terminals with a more diverse range of requirements than human-held terminals will ever require. In the number of terminals, we expect a deployment of seven trillion terminals, three orders of magnitude above the number of the human-held terminals. This will pose huge technical challenges to many areas of the network. We need new network control solutions that can effectively handle the authentication, naming, addressing, routing and related functions for such a vast number of terminals. Radio systems will have to provide solutions that accommodate very low-traffic and low energy terminals as well as high-capacity terminals, plus additional requirements like low delay and localizations, all with high indoor penetration. Terminal technologies will have to deliver solutions with a variable combination of requirements at extremely low cost, very low energy, and very small size.

The combination of network computing services and internet of things drives a third fundamental change in the service model, consisting of the advent of the vertical sectors. Current mobile communication services have individuals as their main customers, and of course individual user will remain in 5G. However, we will observe a huge increase of corporate customers, the so called “verticals” that will subscribe combined transport and computing services in the shape of virtual distributed computers under complex SLAs. This virtual distributed computing services will be used by the verticals to offer a new range of services both to individual customers and to other corporate customers. Automotive, Industry 4.0, Entertainment, Energy and E-health have been identified as large potential vertical sectors, but many others will also exist, like financial, logistics, aerospace, security or smart-cities.

In addition to the three above fundamental qualitative changes in the service model, 5G will also imply deep technical modifications in the network internals, that may not be so much perceived by the end customers, but that also have profound implications in the technology and the business.

One extremely relevant change in the network internals will be the transformation of the control plane driven by NFV and SDN. The virtual computing capacities described above as a fundamental change in the telco service model will also be applied to the control and management of the network itself. The networks will be based on a set of very powerful, very cheap, specialized hardware elements, all integrated and controlled by flexible control modules. The network will be composed of commodity processing, storage, transmission and switching elements, with normalized data and control interfaces, all controlled by logically centralized and physically distributed software controllers. This implies a vast reduction of the current distributed control model based on protocol entities running on each network node which interact to jointly achieve the desired network behaviour. This distributed control model will be largely replaced by a centralized control model in which a reduced number of orchestrated controllers receive status information from, and issue commands to, each and every of the commodity hardware elements that compose the network.

Another extremely relevant change in the network internals will be the development in radio and optical technologies. 5G communications imply radio requirements not available today that have to be developed based on new technologies. The capability to use more spectrum that will be dispersed over many frequency bands, the huge pressure to increase spectrum efficiency, and related requirements like high capacity, low energy, low delay, location services and excellent coverage will lead to radically new radio developments. The 5G network will for sure be based on a multi-RAT approach, including new developments like millimetre wave, massive MIMO, beam steering and possibly non-orthogonal carriers. To coordinate all these access technologies, a Single Radio Controller (SRC) should be introduced. With this approach, it could be feasible to optimize user connections among all available access technologies, depending on the required resources at any time. By extending the SRC, other types of communications could be established, like network-driven device-to-device (D2D) connections, where user equipment directly exchange information under the control and coordination of the network operator.

From a purely technical point of view, we can provide a set of quantitative Key Performance Indicators (KPI) for 5G that constitute a measurable objective and specific directive for all projects involved in ongoing a future open calls. The 5G-PPP[1] has proposed the following list of KPIs that have been widely accepted by industry and academia:

- 1000 times higher mobile data volume per geographical area.
- 10 to 100 times more connected devices.
- 10 to 100 times higher typical user data rate.
- 10 times lower energy consumption.
- End-to-End latency of < 5ms.
- Location precision < 3m
- Mobile speed up to 500 km/h
- Ubiquitous 5G access including low density areas.

We may therefore see that 5G will imply a revolution in the communications business sector, and will also imply a profound change on all the other business sectors. The development of 5G is already in the works, with huge investments throughout the world. In Europe, the Commission has devoted 700 million Euros, leveraging an investment of an additional 7,000 million Euros by private industry and member states. Other world regions are also devoting huge investments to position themselves in this global race. 5G technology will start to be available in 2018, based on pre-standard technologies: some regions have already allocated specific spectrum to this end. Widespread deployment is expected to start in 2020, to achieve progressive penetration and improvements over the next years.

In spite of the radical scientific and technical innovation that will make 5G a reality, new discoveries in academic and industrial research are already paving the way for additional revolutionary advances. In this document, we will present a series of research directions, challenges and recommendations that, in combination, will enable our vision of the **"Pervasive Mobile Virtual Services"** (PMVS) concept. This concept is based on the universal availability of apparently unlimited information services. This implies that the information services will be totally pervasive, provided by a distributed virtual infrastructure combining proximity resources with huge infocentres, and making them transparent to the end user. A user will not have to adapt himself to the terminal and the way the services are delivered but rather the infoservices will have to adapt to the user, so that he has a totally natural experience. New terminal types embedded in cars, doors, mirrors, appliances, and new interfaces based on gestures, facial expressions, sound and haptics will be the basis of the interaction between humans and the infosystems.

This document presents the view of the Expert Advisory Group of the European Technology platform Networld 2020 on the research lines that will make Pervasive Virtual Mobile Services a reality in the coming years. The document covers four research areas:

- Virtualised Networks and Services

---

[1] https://5g-ppp.eu/kpis/

- Radio Networks and Signal Processing
- Optical Networks
- Experimentation with Verticals

In addition to this four broad areas, the document contains an additional chapter addressing other relevant research directions that can provide interesting results for future communications developments.

The document is structured as follows. Section 2 presents a brief summary about current research and innovation initiatives related with 5G. The aim of this summary is to provide the state of the art for the future research on Pervasive Mobile Virtual Services. Section 3 is devoted to detail the main detected challenges towards PMVS, while Section 4 contains the detailed description of the research and innovation actions required to achieve these ambitious goals. In Section 5 we present the proposed roadmap, which together with the recommendations analysed in Section 6, mark the strategy research and innovation agenda for the upcoming years in the information and communication technology field in Europe.

2 Research and Innovation in 5G

One of the most important research topic in 5G is network softwarisation, where some network functionalities are shifted from specific hardware to commercial off-the-shelf devices. There are mainly two new actors in scene: Network Function Virtualisation (NFV), where network functions are virtualised, and Software-Defined Networking (SDN), where the forwarding plane is decoupled from the control plane. Although there is a huge of effort put in these two paradigms, they are not mature enough to be massively deployed, and some improvements are necessary.

The architecture proposed by the ETSI NFV working group has mainly three parts: (1) the virtualized view of the physical infrastructure, (2) the virtual network functions themselves and (3) the management and orchestration block, controlling the previous parts. Regarding the first part, research is following several paths: introducing hardware-based virtualisation, where the hardware can help to share its resources; reducing the load (in terms of CPU, memory, storage, networking, etc.) imposed by virtualization with other approaches like Containers, unikernels, etc. For the second part, research is focusing on the split of network functionalities to effectively distribute them among the physical infrastructure, which is actually hided by the virtual view of the resources. Finally, the management and orchestration (MANO) is in charge of controlling the previous two blocks. On the one hand, the MANO must be aware of the physical infrastructure available to virtualise resources and, on the other hand, it must be able to start, stop, move, connect and destroy virtual network functions on top of this virtual infrastructure. There are different alternatives that implement MANO, but all of them are at a very early stage and more work is necessary in this direction. With the architecture described before, there are many open areas where active projects in 5G are currently tackling: multi-domain networks, where independent infrastructure providers should present a unified virtualised view to virtual network functions;

heterogeneous networks, where different technologies are used in the infrastructure; and multi-tenancy, where the business model is such that one clients of the virtual infrastructure provider may offer the same service to other clients. All these challenges present several issues that are being studied, but others are left for future work, mainly due to their complexity.

In SDN, the control plane is decoupled of the data plane, using the so called southbound interface (SBI) to exchange information. The control plane is usually composed of a controller and one or many applications, using the northbound interface. The most used protocol for the SBI interface is OpenFlow, which is still evolving nowadays. On the other hand, there is not a reference NBI to connect applications with controllers, and there is an ongoing work in this direction. Furthermore, there are several implementations available to use as a controller, but all of them have pros and cons.

Based on SDN and/or NFV, there are other initiatives to push forward in this direction of network softwarisation.

- For example, some projects are working on a software architecture to execute hardware-agnostic applications (network functions indeed). This is similar to virtualisation but reducing the load of the latter.
- Because the control plane and data plane are totally decoupled, it is possible to define a cognitive control plane, able to detect changes in the network in advance. Machine learning and/or Big Data can be used for this purpose. With this functionality, the network automatically adapts to future loads.
- As it is possible to move network functions from one point to another, it could be feasible to move the intelligence, storage, computation, etc. to the edge of the network, there where it is necessary.
- One of the most powerful services that can be provided by network softwarisation is the possibility to recover after failures. When a certain network functionality goes down, the MANO can detect such failure, instantiating a copy of the affected network functionality, probably with the same state as before the service disruption. Of course, the previous connections have to be restored too.

On the other hand, under the common drivers of improved spectral efficiency, increased system capacity, and network costs reduction, new advanced features are still required for optical control and data plane, able to support an evolutionary approach towards the scenarios of future 5G networks and services. In order to face the challenges of the rapid growth of traffic and services, the adoption of SDN-enabled architectures is particularly attractive for the operation of next-generation optical networks. In fact, these architectures enable to support the on-demand dynamic configuration of programmable network functions, such as rate, bandwidth, path adaptation, switching and slice-ability to handle a wide range of granularities (from the sub-wavelength to the super-wavelength) and highly scalable systems. In addition to the frequency domain flexibility, recent research interests move towards compressing more parallel channels with space division multiplexing (SDM) technologies based on multi-core fibers or multi-mode fibers. The SDM transport links promise an increase of dozen times more optical capacity as that of SMF-based links. However, it is still a big challenge to scale up the key network functions to the multi-dimensional optical network. In access networks, the challenge to be met is not limited simply to the realization of photonic transport systems supporting higher capacities at low cost; it also involves the introduction of technologies able to reach pervasively the users from the metro to the access segments of the 5G networks. In particular, the photonic transport network underlying 5G systems must support heterogeneous classes of traffic (as different as C-RAN fronthaul, high speed internet video and IoT), differing in modulation format, latency requirements and capacity. Indeed, future generations of fibre access (e.g. passive optical network) standards may focus more on meeting the demands of mobile networks than those of fibre-to-the-home subscribers, which has been the main focus up to now. Stipulated by the demand to reduce deployment and operational costs, a converged infrastructure both at structural and functional level is essential, instead of having multiple infrastructures delivering the same or similar services.

Within the domain of network control, some preliminary projects addressing the main challenges in this area, are scoped by an evolution and slow migration towards centralized approaches exemplified by SDN principles and the adoption of NFV. A tighter integration (e.g. where the cloud management framework better integrates the different network segments and technologies, including, where applicable, the radio/fixed access, aggregation, metro and core segments) is still missing. New architectural solutions are needed, building on top of current initiatives, and acknowledging that the operation of cloud infrastructures and of data networks can no longer be decoupled to a large degree. Such solutions need to be carrier-class, build on top of extensible but robust and well proven high-level frameworks, constructs and languages, and offer to meet the requirements of high-availability, robustness, or transactional semantics.

Finally, on the data center side, the estimated increase of traffic to be handled by multi-tenant data centers in the next few years puts great pressure on re-designing legacy DC networks (DCNs) to scale, for example, in terms of latency, bandwidth, and power consumption. Bandwidth of server connectivity in data centers continues to grow and is projected to reach 100 Gb/s in 2019 and 400 Gb/s in 2023. Meanwhile capacity and port count of switching ASICs are not scheduled to follow this trend leading to an inevitable disparity between demand and offer of ports and bandwidth at all levels of the data center network (DCN) fabric. Optical switching techniques and devices offer practically unlimited bandwidth pipe-lines between nodes on the network as switching is performed on the physical layer and is therefore bit-rate agnostic. Optical switching also offers “speed of light” latency as traffic between connected hosts remains in the optical domain. The use of multi-dimensional optical switching techniques in intra-datacenters networks is a challenge to be investigated.

3 Challenges towards Pervasive Mobile Virtual Services

The knowledge-based economy is redefining the foundations upon which the existing social and economic activity is built; starting from the way we produce things it will, progressively, shake-up all aspects of our lives. For example, it will allow overcoming technology-related restrictions that confine humans to implement their social functions at certain locations only, something that is raising the possibility for a new *nomadic* state of existence. Moreover, it will create the necessary conditions to overcome the current over-fragmentation in human skills. An important step to materialize this future is the creation of a *Digital Environment* where networks, compute and storage facilities, intelligence, control and management mechanisms, as well as any kind and type of "machines" and "things", are amalgamated into a *digital continuum*. The latter can be obtained through a combined research effort that will integrate competences in the field of virtualized services, optical networking, radio access and experimentation of verticals. This integration represents the highest challenge for the development of networks beyond the 5th Generation.

3.1 Virtualisation of Network Services

The virtualisation process of network services provided by the Digital Environment encompassing compute, storage and communication tackle several aspects.

First of all, it is important to review the traditional concept of a *domain,* largely identified as an administrative concept in communication networks. Software networks will impose a high degree of flexibility where multiple domains may exist at the same time and with different dimensions. On the other hand, such software networks allow the deployment of hierarchical multitenancy, where network providers may offer their (virtual) resources to other network providers, which can follow the same approach towards other network providers. To this end, it is necessary to extend the ETSI NFV MANO architecture to enable recursive deployments of functional components for multitenancy.

Nowadays network virtualisation stands on the shoulders of high-end commodity servers. Although it is true that the cost of these devices is low compared with dedicated hardware, infrastructure providers can reduce this cost by using other types of *hardware with constrained resources*. This way, it could be feasible to deploy distributed infrastructures moving network functions, storage and computation beyond the edge of the network. Anyway, virtualisation on resource-constrained devices is a challenging task, so it is also important to reduce the load imposed by traditional virtualisation environments with other approaches like containers and unikernels, for example.

With such complex infrastructure, it is important to provide mechanisms to ease the configuration of the network. Intent-based networking promises a reduction in the complexity of configuring networks, as orders will be expressed with intents and not with rules and/or filters. Anyway, some kind of *self-configuration* is desired. For example, the network may reconfigure itself if a user flow is affected by congestion. On the other hand, network controllers could provide information to applications to adapt their services to the state of the network. For example, when users are broadcasting a live event, the quality of the transmitted video can be adapted based on the information provided by the own network. This adaptation can be done based on some other information, like the quality of the signal, access technology, etc. The control plane could inform applications about user mobility, for example. In ultra-dense deployments, mobility management is challenging due to the frequent handovers when the user equipment moves between different small cells. Interference coordination for overlapping small cells has to be considered too. In such ultra-dense scenarios, users could decide the best provider to connect to, based on parameters like cost. Nowadays renting access to network providers on demand is far from feasible, and several parts of the network should be adapted to allow such service.

On the other hand, virtualisation imposes some negative effects regarding *energy consumption*, compared with dedicated devices. It is true that with virtualisation it is easy to switch on and off virtual functions when required, but even so, the energy consumption during operation times is higher than with dedicated devices. It is important to improve the performance of virtualisation for such environments to reduce the gap.

Finally, *security* is usually left as an optional functionality, but in the area of virtualised network services this must be considered from the very beginning. It is important to isolate the different workloads running on the same infrastructure, in order to avoid security issues of tenants accessing the information of other tenants.

3.2 Radio Networks and Signal Processing

Each day the world embraces more devices to connect everything, everywhere, and everyone. Driven by global digitalization and the emerging *Internet of Things* (IoT) or everything (IoE) paradigms, the number of connected devices is increasing at an exponential rate from the current ca. 10 billion to an estimated 20-fold within the next five years. The abundant exploitation of wireless sensors, gadgets, multimedia services, autonomous robots or tactile Internet, augmented reality, and other similar applications, will require unparalleled access rates with high reliability and low latency.

Another challenge that has to be taken into account is the introduction of *end-user terminals* as elements of the network. For example, smartphones usually have several network interfaces, which may provide connectivity to other terminals. Including these devices under the control of the network would improve its overall performance. In particular, *vehicles* may become integral parts of the system, not just end-nodes. In essence, acting as mobile base stations and contributing to fill coverage holes, support local capacity needs that appear unpredictable in time and space and, ultimately, provide high quality of experience for passengers. In addition, vehicles will be able to sense environment to support a multitude of applications. For instance, vehicles could perform real-time traffic and environmental monitoring, which in turn will enable real-time traffic management to increase the efficiency of the transport system and reduce its environmental impact.

The combination of novel communication paradigms, including multi-hop D2D (device – to-device), delay tolerant networking where humans, animals and vehicles will be used as data carriers, should be fostered, in order to achieve the maximum degree of flexibility for the network, making it able to adapt to the ever-changing spatial-temporal characteristics of traffic demand.

The evolution of internetworking will imply multi-service networks carrying multiple traffic types over *shared physical resources*. Building such wireless networks to meet customers' diverse needs while ensuring efficient resource utilization will require cross-layer approaches that add support for multiple classes of services at the different levels of the network, as well as novel architectural concepts.

Satisfying the hugely growing need for spectrum to fulfil demand for much more capacity and increased data rates, will firstly rely on the use of *millimetre-wave and higher frequencies* where massive amounts of spectrum is available in addition to lower frequencies, and possibly also dynamic spectrum access to unused licensed or unlicensed bands or geographically-dependent unused bands.

Since at higher frequencies communication is restricted to relatively small ranges and mainly line-of-sight links, especially due to limitations in mobile terminals including the number of antennas that could be supported and battery power consumption, we will see further *densification* in the form of massive deployment of small-cells which will be connected to the network with efficient, high-throughput and low-latency optical and wireless backhaul links.

The need of un-interrupted services, resilience, flexibility, dynamic deployment will further push towards heterogeneous integrated network architectures which in addition to macro and small cells may also encompass *an aerial component*, including high altitude communications platforms, low and medium orbits satellite constellations, geostationary high throughput satellites, etc. For example, satellite terminals can be integrated in the communications architecture as network nodes, which can be used as complementary backhaul links.

Finally, the need to adapt the radio access network to the spatial and temporal variations of traffic patterns, will make the paradigm of *moving networks* essential; through low altitude UAVs (e.g. drones), and terrestrial vehicles, the network access points (base stations) will follow the traffic demand, ensuring maximised capacity and coverage, for both "things" and human-oriented devices.

3.3 Optical Networks

The main optical networking challenge to the creation of this Digital Environment is to effectively address *coordination* issues i.e. to provide a tangible framework for a seamless operation both locally and globally. This task is feasible only through the introduction of *automation* in the handling of the Digital environment, a process that requires: a) the creation of an *overarching Operating System* (OS) which will be an open-software, common, operational mechanism for all digital systems; b) a *Homogenization* in terms of the deployed technology i.e. to deploy the same set of standard and scalable technologies across all optical network sectors and across different operational domains. Through these, we envisage that processing, storage and optical networking systems will gradually converge towards an integrated platform exploiting commodity (or fully interoperable) hardware that is controlled and managed by a set of distributed autonomous controllers across different optical network domains and technologies. These conditions will allow the automatic response to dynamically changing services and user demands.

Another important research challenge is the one associated with *speed* and *volume*; specifically, with the creation of data flows at such volumes and at such speeds that would be hopeless to have them handled by centralized processing system architectures. What is sought is an effective way for coordinating architectures that optimize the balance between *decentralized*/distributed and centralized processing and storage infrastructures that should be interconnected by means of tuneable, i.e. flexible and easily adaptable, optical network infrastructures. In the same direction, we need to implement a two-fold strategy to avoid capacity crunch: first, to identify architectures that avoid transporting the entire flux of data created in the wireless/fixed access part back to the optical backbone. Second, we need to continue exploring research initiatives to ensure ultra-high capacity optical backbone networks will be there when we need them.

The final research challenge is associated with a paradigm shift from technology-based to *performance-based research objectives*. A technology-agnostic access network is the necessary framework for making the most from both wireless and optical technologies, without any prejudice, and without any primacy of one technology over the other. Performance, including CapEx and OpEx, should be the final criterion for selecting, or not, a particular combination of technologies and architectures, something that is creating fresh interest towards Access-Core integration concepts. In parallel, the process of homogenization is also setting performance as the ultimate criterion for selecting, or building, a particular infrastructure since these features, i.e. the deployment of common technology blocks in building networks and their joint operational framework and performance, will finally define the ability to slice resources in the quest for Infrastructure-as-a-Service (IaaS).

3.4 Experimentation and Verticals

Being able to address the above challenges through an integrated approach requires *interdisciplinary groups of researchers*. Teams including engineers (communications, electrical and software engineers), economists and specialists in digital economy, sociologists, lawyers, bio-engineers, user's associations and professional associations may work together and be able to understand better the whole set of implications before defining or deploying a new service. Projects should not only provide large-scale experiments to proof new advances but also have a task devoted to this interdisciplinary approach thinking in the long term.

The testing and validation of research results, implemented solutions, products and services, in large-scale real life experimental infrastructures, is essential for the innovation process of the future Internet. Current available experimentation facilities should be extended in order to provide enhanced experimentation infrastructures on top of which, *third party experimenters* e.g. SMEs or any digital asset owner over vertical sectors will have the opportunity to test their vertical applications and solutions in an integrated, cooperative and fully featured infrastructure fine-tuned to the characteristics of each vertical sector.

The support of *verticals in experimentation facilities* will attract much wider range of stakeholders compared to existing general purpose or vertical dedicated solutions, ranging from SMEs and industrial partners developing products and services over vertical sectors to public bodies (e.g. municipalities) and organisations (e.g. automotive safety organizations).

While many verticals have been identified in the context of 5G, there are some areas left uncovered. *Public Protection and Disaster Relief, mission critical applications*, are currently not sufficiently considered. They require specific types of testbeds. Moreover, the development of the *Industry 4.0* also calls for peculiar experimental facilities. In both cases the application requirements are very challenging, in terms of latency and 99.999% reliability. These challenges need to be tackled through specific approaches.

Finally, the inclusion of the *social aspects* into networks requires more attention form the experimentation viewpoint. The sociality of humans, and the sociality that their belongings inherit, can be exploited for the sake of better network resource utilization.

4 Research and Innovation for Pervasive Mobile Virtual Services

4.1 Virtualised Networks and Services

4.1.1 Converging the NBI: fostering a unified approach to Intent Based Networking

One of the current trends in networking is Intent Based Networking. By providing an abstracted view of the network, the different configuration and maintenance tasks than need to be performed by a network operator are significantly simplified. This principle can be applied to Software Defined Networking in a more ambitious scale. Current 5GPPP projects are exploring ways to construct 5G network architectures around SDN and specific SDN controller platforms they augment. In the framework of the 5GPPP, it would be beneficial if consensus on the definition of network intent and its expression could be reached.

As an example of an application that would benefit, we have machine learning controller SDN. The concept is that we have a machine learning engine that takes measurements from the underlying network, infers the network state and proposes a series of actions to improve a specific network KPI. With current state of the art, the actions are expressed directly in commands that are sent to the SDN controller of choice.

A more sensible action would be to look for the possibility of expressing these actions in a controller-independent way. We could thereby develop once and reuse in the different platforms that are emerging from Phase 1 5GPPP projects. This controller-independent way could be intent. However, we are still far from a definition of intent that is controller-independent. Moreover, we currently observe that there is a significant number of mutually exclusive initiatives to define intent, most of them linked to specific SDN controller platforms. As a consequence, there is an intrinsic risk when choosing a specific intent idiom at this point in time. We need standardisation.

Currently, this shortcoming is starting to be recognised in different fora and there are first initiatives to bring different currents of thought together. As with other initiatives (e.g. NFV) there is an opportunity for Europe to get involved in this discussions and take a leading role, letting all the stakeholders benefit in different ways: 1) new research opportunities to produce the tools that, once deployed would imply 2) lower operational costs for more reliable and flexible networks, that would result in 3) more flexible and powerful services for the end-users.

4.1.2 Pervasive network virtualization

Since the dawn of computer networks, communication devices are designed to be dedicated hardware and software components, glued together. The main reason for this is performance. Although standards came to help to achieve the interoperability between devices of different vendors, the configuration of such devices are specific for each product. There are other other problems that network operators have to face with this type of equipment, but the most important one is the cost, both in terms of CapEx and OpEx. To tackle this problem, a new paradigm called network function virtualization (NFV) is being defined in standardization bodies like ETSI. The ETSI NFV working group has published several documents, including a reference architecture. The main building blocks of such architecture are the management and orchestration, the virtualized network functions (VNF) and the NFV infrastructure. Regarding this last block, the hardware resources available are hidden to the layers above, by inserting a virtualization layer, presenting virtual resources to the VNFs. The goal of NFV is to use commercial off-the-shelf hardware to deploy the NFVI. But, in order to reduce the gap of performance compared with dedicated communication devices, it is recommended to use high-end, expensive servers with special hardware to help with the virtualization: new set of CPU instructions, a non-uniform memory access CPU architecture to accelerate access to memory, network virtualization at the network interface card, etc.

Reducing the cost of the infrastructure is the next step of network virtualization, but it is important to take into account the performance of the overall solution. We envision a distributed infrastructure composed by a minimal number of high-end servers together with a plethora of resource-constrained devices, deployed in strategic places. High-end servers are necessary to execute the more demanding functionalities, like core routers. On the other hand, other network functions can be distributed in low and medium-end servers, deployed over resource-constrained devices, which have a reduced cost, low power requirements, small size, etc. Although nowadays the gap in terms of performance between these resource-constrained devices and high-end servers is high, the forecast in the near future foresees an important reduction in such difference. Due to the small size, weight and power consumption, it could be possible to easily distribute this type of devices where and when necessary. For example, we can imagine drones transporting small base stations/computers from one place to another. Thanks to virtualization, it can be easy to transfer virtual machines from one board (because the drone it is attached to is going out of battery, for example) to another; it can be possible to suspend one or all virtual machines, etc. The benefits of using constrained-resource devices are clear, so it is important to address the current challenges that virtualization imposes.

In order to fulfil the challenges of virtualization over resource-constrained devices, several stakeholders have to increase their effort in this field. Equipment vendors should improve the architecture and increase the set of instructions for these small devices. Software companies should optimize hypervisors and virtualization software to run on these devices. The management and orchestration functional boxes have to take into account the mobility capability of the infrastructure, if drones are available. The impact of this new NFV infrastructure in the TIC sector will be very high, increasing the employment and benefits of all these companies in Europe.

4.1.3 SDN Enabled Terminals

Thanks to the Software-Defined Network (SDN) approach, a centralized controller has a global view of the network and the capability to manage its traffic. Nevertheless, user terminals (e.g. smart phones), are not under the control of such entity. Usually, they are not considered part of the control and forwarding planes of the network since, traditionally, they were leaf nodes with just a single network interface. There are, however, few exceptions, such as TCP congestion control and DHCP address provisioning.

Nevertheless, current terminal devices have multiple interfaces with heterogeneous technologies, and with the advent of 5G we expect terminals to include even more interfaces and communication capabilities. Different technologies provide different capabilities that are useful in different environments. For example, Wi-Fi is typically used in smartphones to improve indoor coverage and increase the bandwidth. Bluetooth is used when low latency and predictability are required for short range communications, 2G in areas with bad coverage, etc. Nevertheless, the choice of the technology, interface and provider is usually left to the user of the device, or, in some cases to an application.

The increasing complexity of choosing between the interfaces that will be available for terminal devices, and the need for continuity in communications across different technologies as they become available or unavailable, make more intelligent and powerful mechanisms to control terminals from the point of view of the network necessary. This will simplify user's decisions providing always the best possible communication choice, even in complex scenarios such as the one shown in Figure 1.

Several proposals on access network selection have been put forward given the limitations of the current simple mechanisms. The 3GPP, for example, has developed and implemented the Access Network Discovery and Selection Function (ANDSF). ANDSF assists 3GPP terminals in the detection of non-3GPP access networks, such as Wi-Fi and WiMAX. It stores information on alternative access networks on a server that is retrieved by terminals. However, ANDSF rules are static in the sense that they are applied for an entire session regardless of network dynamics (e.g. number of terminals connected to a specific access point). Other proposal, the Media Independent Handover (MIH) standard developed by the IEEE 802.21 group deals with vertical handover between IEEE 802 access technologies. MIH could be used to provide terminals with suggestions about which access network they should connect to.

These proposals must be improved and extended for terminal devices to fully participate in an SDN enabled network. Terminals are even more complex than routers (for example, because of mobility and the many scenarios and situations in which they are involved), but they can be modified to include a SDN API such as OpenFlow (or any other API that can be used by the controller), following a network-centric approach seeking a sustainable high-quality user experience.

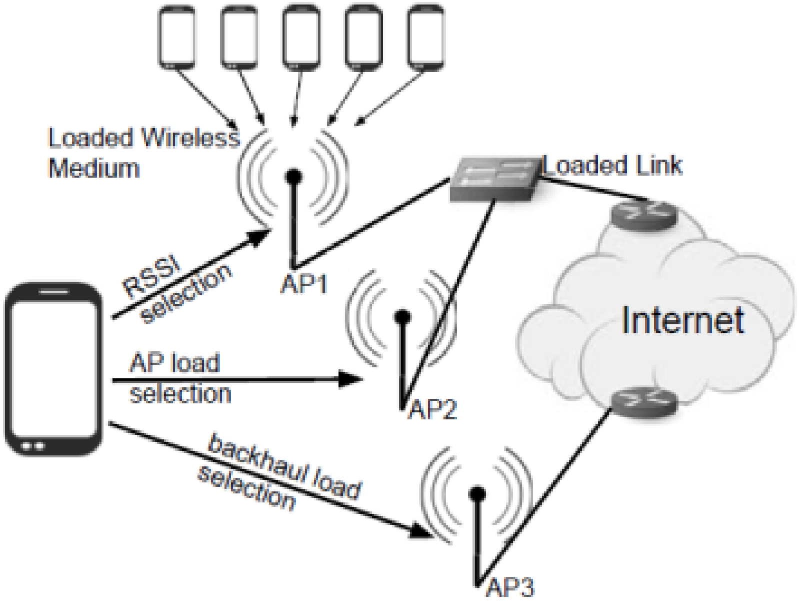

Figure 1 - Influence of parameters for network selection

Europe should continue leading the creation and standardization of global communication technology. By promoting the integration of SDN into terminals, Europe will continue contributing innovations in this field. Terminal manufacturers will have a specific feature that will differentiate their product, and operators will be able to provide a better service that will also benefit the users.

4.1.4 Platform security for cloud and NFV hosts

In a cloud or NFV environment, organisations or individuals run workloads on host systems which are managed and controlled by another organisation. The former are often referred to as "tenants", and the latter as "landlords". Tenants need to be assured that the hosts on which their workloads are executing are trustworthy, as otherwise any workloads which are executing may be subject to compromise, denial of service or other attacks. Compromise of a host system will generally not be visible to a tenant until a workload is negatively impacted, and some compromises may not be detectable all by the means available to the tenant.

In order to ensure that the workloads execute as expected, and to reduce the likelihood of security threats, the tenants must be assured that the host systems both meet the specifications expected, but also that they continue to perform as expected.

There are a number of areas of interest to the tenant and/or its workload. These include:

- that the expected hardware is present, and has not been changed
- that the BIOS and firmware is as expected

- that the underlying OS and virtualisation software (e.g. hypervisor) and associated management applications are as expected
- that other services such as libraries are as expected
- that storage and network services are as expected

Although mechanisms exist to perform such checks, they are typically performed at provision time and boot time: for the tenant, there are at several other times in the lifecycle of the host at are important:

- load time (when the workload itself is loaded or unsuspended);
- run time (when the workload is running);
- migrate time (if and when the workload is migrated to another host system);
- suspend time (if and when the workload is suspended);
- resume time (if and when the workload is resumed);
- completion time (when the workload is being decommissioned and removed from the host system).

At various of these times, the workload itself is not available to be reported to, and may not be the most appropriate entity for reporting anyway. In many cases, the details of the various aspects of the platform should not be communicated to the workload specifically, as some of the information about software versions, configurations, etc., might provide information which could allow tenants or other parties to exploit the host system or wider infrastructure.

Research topics include: hardware requirements; software requirements; appropriate levels of abstraction for information sharing; mechanisms for information sharing (directly to workload or to other tenant-controlled party); appropriate parties and channels for reporting; techniques for attestation at different points in the lifecycle; impact on workloads if various components (hardware and software) are compromised; mechanisms for mitigating impacts of compromise.

Robust work to allow tenants' workloads to trust the host systems provided by their landlords will allow for much greater opportunities to deploy sensitive and critical workloads across infrastructure in which trust is mainly established, currently, through business and contractual agreements. It will also allow landlords to offer more, and better differentiated, services to tenants, whilst also improving their own knowledge of their systems, vulnerabilities and compromises.

As more critical infrastructure - at the personal, organisational, national and international level - is virtualised, robust reporting and mitigating mechanisms are vital for business continuity.

4.1.5 Trusted Execution Environments (TEEs) for sensitive virtualised workloads

Although tenants must have some trust in the landlords in order to consider running their workloads on the landlords' host systems, there are some sensitive workloads which may not be appropriate for existing offerings, because the confidentiality or integrity requirements that the tenants have may be too high. Standard existing commercially available virtualisation technologies - hypervisors and Linux containers - currently allow users or processes which have super-user access to the host system to "peer inside" the workloads and to make changes to data within them. This means that data confidentiality and integrity cannot be assured.

The impact of this is that tenants cannot fully trust landlords to host sensitive workloads - and even if the landlord as an entity is trustworthy, individual employees may not be, and a tenant cannot be expected to manage attacks from individuals internal to the landlord. Hardware Security Modules (HSMs) provide one way to allow for sensitive data to be protected, but they are expensive, complex to manage and share between workloads, and difficult to program.

An alternative approach to HSMs is the use of Trusted Execution Environments (TEEs), which are based in processor hardware, and allow processes to execute in a protected environment where no other entity on the host system - including any super-user processes - can inspect their data or negatively impact its integrity. Applications are designed to run partly within a TEE, and all access to memory within the TEE is controlled at the hardware level, providing confidentiality and/or integrity protection: processes outside the TEE do not have access to unencrypted TEE data, though processes within the TEE may have access to data in other memory areas.

While TEEs are becoming widely used in client and mobile environments, their use in virtualised server environments is fairly new, and a variety of topics are appropriate. One of these is what types of workloads will most benefit from the use of a TEE. These are expected to include:

- secure monitoring
- cryptographic key management
- firewalls and other secure routing services
- services managing customer data
- financial services
- legal services
- government services
- pharmaceutical and medical services
- remote services (e.g. base stations) that are not physically secure
- big data

Research topics include: techniques for using Trusted Execution Environments (TEEs); secure provisioning of TEEs; remote management of TEEs; appropriateness of particular workloads to TEEs; migration of workloads using TEEs; design patterns for use of TEEs for various workloads; protection of workloads within TEEs from denial of service or resource starving attacks; availability of reliable time sources to workloads within a TEE.

The increasing move to virtualised services requires that tenants place more and more trust in landlords, but it is impossible for landlords to guarantee confidentiality and integrity for all workloads in all cases. Use of TEEs will allow tenants to roll out a wider range of services to landlords, leveraging economies of scale and providing for simpler management. It will also allow tenants such as financial services and government departments much greater assurance that their data and processes are appropriately protected. Landlords will also benefit by being able to make stronger guarantees to tenants and offer different types of service to tenants with various security requirements.

4.1.6 Trust domains for virtualisation and NFV

There are multiple components involved in the running, management and orchestration of the tenant and landlord domains. The different components may well have different security policies, and therefore sit in different trust domains, even when they sit in the same administration domain. A better understanding on how trust domains should be defined, provisioned, managed and de-provisioned is required to allow for more secure cloud and NFV deployments.

Separating trust domains is complex, as there are many different operations or fields of influence where different components may have impact on each other. To give one example, a security monitoring system run by a tenant may not trust:

- other tenants who happen to be hosted by the same landlord, who may try to disrupt the proper operation of the landlord or other tenants;
- the landlord to perform actions as requested, and not to interfere with any of the data being collected and delivered;
- other components owned and administered by the same tenant (e.g. build systems, web servers or database systems).

This lack of trust may span a variety of capabilities. For instance, the security policies that the monitoring system maintains, any information it has collected, and cryptographic keys are all information that are contained within the component itself, and need protecting. But it may also require that only trusted parties should be able to provision, start, stop, suspend, migrate or de-provision it. An understanding of exactly which components, in which trust domain, should be able to perform these actions is also key to the correct operation of the security monitor in this example.

Operations that may sit within different trust domains include, but are not limited to:

- monitoring (operational and security)
- tenant infrastructure management
- host infrastructure management
- SDN infrastructure management
- storage infrastructure management

Research topics include: categorisation of sensitive operations requiring trust domain definition; setting up of trust domains; real -time and forensic identification of possible compromises or breaches of trust; mitigations to trust breaches; transfer of trust between trust domains for components through their life-cycle; software and hardware tools for trust establishment and transfer; transitive trust between different actors and components; the impact of geographic, national and jurisdictional boundaries on trust domains.

In order to allow growth of the use of virtualised services and NFV by organisations across Europe and the world (including individuals, corporations, NGOs and government), clear trust relationships must be created, maintained and monitored to allow the appropriate actors to operate safely and with agreed responsibilities. Service providers and their users must be able to ensure that their property - physical, logical and informational - is under the control of the appropriate actors, and be able to define who has authority and responsibility for data, processes and physical assets.

4.1.7 Lightweight Virtualization of Mobile Network Operators

Mobile network operators work in a very competitive market where new features should be deployed quickly, with low implementation cost and providing high quality services. New mechanisms are required to satisfy these requirements and to enable new business models where new services will be provided following an “agile” paradigm.

Nowadays, mobile network operators have to install new equipment in order to deploy new services. Nevertheless, the Network Function Virtualization (NFV) paradigm will make it possible to provide new services using generic purpose hardware. By creating the technology to virtualize all the functions required by a mobile network operator, it will be possible to implement new features or even deploy a new operator in hours or minutes, and reduce costs by sharing the equipment and optimizing its usage.

Software Defined Networking (SDN) and Network Function Virtualization (NFV) will be key enablers for 5G networks. On the one hand, SDN decouples the control plane from the data plane allowing the deployment of differentiated services in a flexible way. On the other hand, NFV presents numerous advantages over traditional hardware-centric networking approaches in terms of cost, versatility and time-to-market reduction. Another complementary technology that is experiencing rapid adoption is virtualization using containers.

By combining such technologies, it will be possible to adapt the network to the number of users or to their demands. Mobile network operators will be able to create new instances of their systems by calling different NFVs according to the number of requests, or it will be even possible to create operators on demand that will provide a temporary service. It will be possible to test new disruptive services with small deployments, so that, in case of success, they will be made available to all the customers in a short time, without disturbing the operation of the existing functions.

In that regard, although less evolved than core virtualization, the Radio Access Network Virtualization paradigm will make possible the complete virtualization of the operator, i.e. the possibility for a company to provide mobile 5G communication services with fully dedicated resources, but without the need of owning any hardware infrastructure. For example, Cloud RAN (C-RAN) separates Baseband Processing Units (BBUs), which handle signal processing and resource management, and Remote Radio Heads (RRHs) that handle RF translation. With this approach, it is possible to share the same infrastructure, performing the baseband processing for different operators at the same BBU and using the RRHs according to the demand. A similar approach can be implemented with small cells, by assigning the small cells to the different operators when they need them. In other words, it will be possible to reduce the time to deploy a new operator from days to only hours or even minutes.

Furthermore, this will facilitate the adoption of new architectures. For example, it will be possible to move computing power to the edge of the network (Mobile Edge Computing, MEC), for applications that require ultra-low latency and high bandwidth.

There are multiple challenges that should be addressed to make Lightweight Virtualization of Mobile Network Operators a reality. Firstly, an abstraction layer should allocate radio resources in a completely transparent way from the point of view of operators, guaranteeing the agreed service levels. Secondly, an orchestration layer should provide the mechanisms to manage the lifecycle of the different virtualized functions, including their instantiation, deployment, monitoring, scaling, sequencing, and termination, without interference between the different operators. Finally, new business models should be explored and validated.

The technology for the complete virtualization of the mobile network operator will completely change the market. It will make it possible to share the infrastructure among several operators, but being always in control of their virtualized infrastructure. Operators will be able to deploy new services in hours or minutes, or dynamically adapt their infrastructure to meet the demand in almost real time.

4.1.8 Video transmission in uplink direction in 5G networks

The growing efficiency of uplink direction in future mobile 5G networks is extremely likely to bring new opportunities for video delivery. Uplink share of mobile bandwidth has been underutilized ever since but the growth of IoT devices, continuous decrease in video footage footprint (i.e. H265), the increasing role of intelligent CCTV and multiple facets of mHealth applications (e.g. teleconsultation) will change the share of uplink transmissions in future internet. It is also connected with emerging crowdsourced incentives to share live video feeds (e.g. Periscope mobile application for leisure activities, security and crisis management – e.g. FP7 iSAR+/SOTERIA), but also the increasing market of video games and augmented reality supported by mobile edge cloud.

The unprecedented growth of importance and the obvious niche in the domain of the target recognition video (TRV) has already been capitalized by most recent amendments to the ITU-T recommendation P.912 [1] which addresses need of systematic approach to measure QoE in the uplink (which is usually challenging due to scenario dependent quality). Directly linked with it is the rapidly growing market of utilizing drones for aerial imaginery and monitoring (of people, forests, lakes, mass events and so on). Crucial role of drones (and swarms thereof) in connection with video streaming for crisis management have been highlighted in recent report by EENA [2]. The latter identifies the increasing need for smoother and more stable delivery of HD streams from drones used in search and rescue operations. Intelligent CCTV systems like the ones proposed recently by FP7 INDECT [3] or FP7 ARENA [4] projects would highly benefit from possibility of effectively offloading calculations of video recognition targets to the cloud, still preserving the real-time requirement for identification of threats. Still the important enabler here is the efficient feedback loop which adjusts video and audio quality and synchronization with external sensors. Such feedback loop is at the moment non existing for uplink – whereas at the same time it is almost completely implemented in downlink (MPEG-DASH) where chunks of adjusted content are delivered based on instantaneous measurement reports from end-terminal. Besides for sake of greater control of quality and flexibility it is crucial that such feedback is not realized as an out-of-bound, proprietary, external signalling subsystem from the operator perspective. Instead it should be tightly integrated with the (SDN) architecture of the 5G operators (i.e. OpenEPC, SDN controller, etc.) to provide maximum of capabilities.

To cope with above mentioned challenges further R&D approaches have to be realized. Network softwareisation seems to be most promising concept towards development of video delivery techniques, focused on improving the overall efficiency. Thus the research in in next phase of 5G PPP shall focus on following aspects:

- The role of lightweight EPC solutions available for end terminals (smartphone, drone) shall be key enabler for introducing **improved control** over video traffic sent by the terminal. Such solutions would assure a truly holistic approach to the control of video delivery chain. This way more “standardized” approaches to **manage transcoding** processes deployed as a service inside an end-terminal or even in a cloud between a video streaming terminal and the ultimate recipients are necessary.
- The uplink direction will be playing an increasing role and thus will be changing the currently strongly asymmetric internet traffic share. Given that uplink video is mainly **event-driven** and **user generated** (i.e. eHealth, CCTV, Periscope kind of apps), its quality assessment should always be “scenario driven” and not simply subject to QoE assessment based on a known set of metrics. This invalidates some more typical QoE approaches [5] in favour of recently amended ones like e.g. [1] which highlights the need for novel methods to assess the majority of uplink video streaming (including novel approaches with crowdsourcing).
- Increasing requirements towards **synchronization of video and audio with parallel data from sensors** tend to prioritize premium QoS assurance inside the network as well as appropriate mechanisms to prevent synchronization loss and allow immediate reaction to network disruptions. For example, in an *eHealth scenario* a user is observed by a camera at home, while doctor remotely tasks patient to perform physical breathing exercises and controls the local charts showing the readings from chest-belt sensors. Premium quality is absolutely essential in assuring appropriate diagnosis and proper treatment of the patients undergoing cancer-treatment.
- Moreover, emerging 5G solutions should analyse and enable inclusion of real-time feedback from end-users, namely **crowdsourced/sensed quality indicators**, as an inevitable ingredient of “target recognition video” assessment. System wide introduction of new performance indicators and vertical-oriented 5G interfaces towards greater support of video delivery shall enable greater standardization in the area of adaptive video delivery.
- The above should be also aligned with activities towards greater energy consumption as well as security assurance for the overall e2e infrastructure.

The impact of delivering the above mentioned improvements would have several dimensions:

- The impact for improved video delivery is expected across all 5G PPP verticals that will be dealing with video delivery. Thanks to postulated solutions, the domain of uplink video delivery will be "regulated" to great extent and thus a multitude of TRV applications will directly benefit. Close alignment with end-to-end architecture of the operator access networks will bring more value to the video chain.
- Tight integration of an end-to-end architecture, with context driven approach, standardized realization of feedback loop and cloud for real-time processing of the feedback loop is a game changer in today's situation where video feeds are controlled from "outside" of the e2e delivery chain. SMEs will be able to focus on adjusting adaptation logic and be less constrained with the need to develop own "black boxes".
- It is already recognized that after EU-wide legislation of drones and accompanied harmonization of radio frequencies and management procedures, pressure to assure delivery of HD video streams from the field (or patient houses, car accident monitoring, etc.) will become urgent.
- Improved control over eHealth video delivery would drastically reduce the costs of traveling, medical diagnosis on-site etc. – providing highest level of video quality of remote consultation with synchronized video, audio and sensor readings in real-time.

4.1.9 Ultra-dense deployment as a Service

SDN, NFV and cloudification of the networks offer huge opportunities for the introduction of new mobile services and for the extension of current cloud infrastructures into new usage domains. The idea of ultra-dense deployments is considered to be one of the key enabling technologies for 5G. It promises low end-to-end delays, ultra-high data rates, support of huge number of simultaneous connections (mainly for indoors but not restricted to) for M2M communication and, through efficient frequency reuse, it aims at bringing the increment in the system capacity required from 5G networks. It is about a requirement of 100-times more system capacity that the expansion to more and higher frequency bands and the envisaged modulation techniques alone are not foreseen to tackle efficiently.

Ultra-densification of the mobile networks however hides a lot of challenges spread in multiple domains. Mobility management is an important challenge due to the high frequency of the handovers as well as the complexity of the underlying dense network infrastructure. Interference coordination is another issue that needs to be tackled due to the extremely high proximity of the small cells in the vicinity and the fact that smalls cells operate as a second tier over the macro cell layer. In addition, the concept of Central Office Re-architecting as a Data centre (CORD) is an appealing approach for operators since it decouples the service provision from the need for infrastructure ownership. It should be also considered that the Fiber-to-the-Home is becoming mature enough and the optical transport is the ideal candidate for front/back-hauling the deployments. To sum up, ultra-dense deployments will certainly face challenges that are related with their nature (density) as well as with the trends in the networking domain (multi-tenancy, cross-layer control also for underlying infrastructure) and at the same time achieve the ambitious goals of 5G.

SDN, NFV and cloudification are expected to have a key role in addressing all these objectives. In terms of networking performance, the logically centralized (but still resilient through appropriate clustering architectures) control of the network will secure a complete view of the network and thus an optimized network management. In terms of visualization, multi-tenancy, elasticity and cost-efficiency, the introduction of the necessary abstractions will enable mobile operators, carriers and cloud providers to compose, provision and exploit efficiently their resources.

Ultra-dense deployment as a Service, may be offered as a unique feature of CORD for mobile services. It should allow the provision of small cells and ultra-dense deployments in an automated way. After the antenna installation that could be performed by the subscribers themselves, the mobile operator should be able to manage the small cell and provide connectivity over any transport connection used by the subscriber for this purpose. All possible transport connection types should be supported and exploited efficiently through appropriate interfacing with broadband carriers. Mobile network functions may be visualized and be located in the data centre. Innovative mobility management and interference coordination techniques may be easily introduced as VNFs and be offered over non-expensive infrastructure, e.g., commodity servers and white-box switches. Softwarization of the ultra-dense deployments and decoupling from hardware should enable experimentation and thus innovative solutions tackling the challenges that ultra-dense deployments face will be easily introduced and verified within extremely short time to market time. New business models and pricing policies should be also introduced to define the business relationships in this new ecosystem.

4.1.10 Relations among network “softwarisation”, network management and energy efficiency

Among the challenges posed by the Future Internet in general, and particularly by the strong wireless/wired integration of the 5G environment, three broad topics, among others, can be seen as interacting and mutually influencing: i) flexibility, programmability and virtualization of network functions and services, ii) performance requirements (in terms of users’ Quality of Experience – QoE – and its mapping onto Quality of Service – QoS – in the network), and iii) energy efficiency. The first item stems from the evolution of the network towards a multi-purpose service-aware infrastructure, to deal with diverse and integrating paradigms as 5G, the Internet of Services, the Internet of Things (IoT), network-integrated cloud services, just to quote a few examples. Performance issues will have to be related with requirements implied by Terabit transport, highly broadband wireless and wired access, zero-loss and low-latency services. Energy-awareness cannot be neglected in view of sustainability, environmental concerns, and operational costs. In order to support all these objectives, network architectures, devices, and base technologies are undergoing a major redesign, including actions like hardware (HW) programmability inside networks and devices, HW offloading for higher performance, extreme virtualization paradigms to make different services sharing network resources, consolidation of services to reduce power consumption. All these actions entail a growing relevance of network management and control, in order to achieve desired trade-offs between utilisation of resources and network performance. Indeed, energy efficiency and the complexity of network operations and management have been already identified as two of the major sources of sustainability and scalability issues for state-of-the-art mobile and wire-line network technologies that, if not properly addressed, might definitely hinder the upcoming revolution in networking paradigms brought forth by Software Defined Networking (SDN) and Network Functions Virtualisation (NFV). In particular, Network Operators and Service Providers should face the ensuing potential increase in Operational Expenditures (OPEX). In this respect, the only significant OPEX sources that can be reduced by technological advancement appear to be the ones related to energy consumption and network management, which account for a figure equivalent to the entire infrastructure Capital Expenditures (CAPEX). With the upcoming SDN/NFV technological revolution, while CAPEX is envisaged to decrease, owing to the adoption of commodity hardware cheaper than today's devices, the OPEX spent in energy and in network operations is expected to significantly increase, unless specific solutions are included in future 5G technologies.

For these reasons, the inter-dependence among Key Performance Indicators (KPIs) concerning networks' and datacentres' QoE/QoS and energy consumption, along with the management and control strategies aimed at enhancing the former and thwarting the latter, should hold a relevant position in future 5G research plans. As regards energy efficiency, the massive introduction of general-purpose HW enabled by NFV would tend to increase power requests with respect to specialised HW solutions. As a matter of fact, for a given silicon technology, energy consumption largely depends on the number of gates in the network device/chip HW. This number is generally directly proportional to the flexibility and programmability levels of HW engines: with general-purpose CPUs, maximum flexibility can be obtained, at the price of reduced performance/power ratios; with very specialised ASICs, flexibility comes to a minimum, but the performance/power ratio can be greatly enhanced. The challenge here is in a careful exploitation of the flexibility offered by new technologies to revert the issue into making network programmability an enabling factor to achieve energy efficiency through management and control actions, without violating performance KPIs (and without causing the OPEX gain to be frustrated by increased management costs).

To this aim, three basic enablers can be envisaged at the chip/system level: i) dynamically programmable resources able to perform multi-purpose services; ii) specialized HW and in-network programmability for network/datacentre offloading to speed-up basic functionalities; iii) standby capabilities to save energy if a resource is unused. In this context, the presence of general-purpose HW (carefully balanced with specialised HW solutions and de-localisation of some services/functions to the access network) offers the possibility of dynamically moving services among the components of a node, or among nodes in a network: when the workload is low, many functions and services can be consolidated to aggressively share single general-purpose HW resources. Thus, even if a general-purpose/programmable resource consumes more energy than ASIC-based solutions, a smaller number of HW elements can be left active, in order to effectively handle the current workload; moreover, such decisions can be applied at multiple time scales (with fast dynamic adaptation or with long-lasting smart standby). Among the main questions to be considered here are those of which basic (sub-)functionalities need to be moved (and "frozen") to the offloaded specialized engines (best performance in terms of speed/energy), and of which ones have to remain in the programmability space (lower performance, but stronger sharing and more evolution opportunities). The solutions need to be identified by considering and effectively supporting the newest trends in Internet technology evolutions. Still going a bit further, the localisation of functionalities would play a significant role: bringing certain executions closer to the user (e.g., by exploiting the fog computing paradigm in the access network) can reduce latency, offload datacentres and increase flexibility in the allocation of resources. New Network Management frameworks and architectures will be needed, able to expose this underlying flexibility, by: i) extending SDN protocols (e.g., with energy-efficiency primitives); ii) virtualising services and network applications; iii) autonomically re-allocating network/ICT service tasks; iv) applying real-time and long-term analytics. In this way, what is currently perceived as a potential problem can be transformed into a strong opportunity toward increasing energy efficiency, completing the integration of networking and Information Technologies in the Internet, and providing maximum flexibility for network functions, protocols and services.

4.1.11 Software Networks (SN)

For many decades, advertising has been based on a model relying on billboards, ownership of real estate (walls, signs, land, …) and long-term relationship between advertiser and customers (i.e., placing boards in well-known and accepted locations). Driven by the internet-based penetration into people's homes, together with the content digitization and softwarisation of advertising content provisioning, advertising has become a business that, in its extreme yet often occurring phenomenon of web browsing, exists for a few seconds (at most) and is built in fractions of seconds, initiated by user clicks on web-based content. Advertising has become an instant business.

We envision that the usage of network resources, the 'eyeballs of connectivity', will develop along the same lines of ultra-instant provisioning anywhere at any time by anybody who provides best connectivity for the best price when it matters.

In order to drive forward this vision, the connectivity industry must achieve an unprecedented software network flexibility with currently unmatched maturity, resilience and security levels of control in order to (a) support on the fly, plug-and-play extensions and reductions of the network infrastructure, i.e. support node and link churn in the order of seconds and (b) allow to change network's purpose, services and behaviour through a set of independently developed control applications, which can be downloaded and installed in an application store session.

In this respect, we have identified the following challenges:

- Unify compute, storage and communication resources, currently represented through frameworks such as NFV and SDN, at the control level through *common means of resource discovery, manipulation and identification*.
- Provide *resilient bootstrapping* of control by implementing the control of a network infrastructure from within this infrastructure itself, i.e. without relying

on secondary/additional/separate resources, uncontrolled by the same control solution.

- Support *dynamic inclusion and removal* of both network nodes and links in a plug-and-play fashion without impacting the network control as a whole (local / global separation).
- Efficiently support *different and often widely varying control requirements* (in terms of QoS for different applications, of different infrastructure scales, etc.)
- Provide *resilience* of any such network control solution, specifically in the presence of control dependencies, so that application operations can never break control, control cannot fatally break itself, and churn cannot globally break control.
- Expose a *general yet crisp network-level API*, possibly with different access levels (e.g. system-level API for system apps, low level API exposure for trusted apps, standard API for typical apps), for the applications and application developers.
- Enable *security models* for multi-behaviour (i.e. multi-application) operations within one infrastructure, with applications developed in an uncoordinated fashion.
- Demonstrate the concepts in a coherent and scaled near-to-trial environment
- Drive the roadmap of developments into standard organisations, ensuring interoperable and open approaches to truly multi-tenant and multi-domain communication environments.

We believe that many verticals, such as automotive, health, transport, industry 4.0, and many others cannot exploit the current static infrastructure control model of 4G with its over-the-top-only service provision. The new dimension of service criticality characteristic of those verticals calls for unprecedented service guarantees, yet service-dedicated approaches from the past do not easily support the economy of scales and put the creativity in the hands of the infrastructure owner. The solution lies in a move towards a truly flexible and instant control infrastructure. The latter will truly accelerate the move of these verticals onto a common, yet inclusive (in terms of network and access technologies) 5G infrastructure, *therefore enabling truly innovative, integrated services at affordable price levels*. The bidding like nature of providing network resources to end users will also facilitate the utilization of resources on-demand to users previously not under exclusive contract with the network operator, therefore *opening new markets to existing and future operators*, while *reducing energy waste* by not exclusively reserving resources that are often not utilized at a given instant.

4.1.12 Hierarchical multi-tenancy in NFV environments

The latest techniques for network virtualization have provided the basis for advanced sharing of physical infrastructures among several tenants, through the dynamic delivery of isolated and customized virtual network infrastructures which scale up and down automatically to follow the evolution of customer's business. The adoption of SDN technologies, open APIs, intent-based network modelling and network programmability has allowed to evolve this basic model towards a more advanced and recursive virtualization of network infrastructures. Virtual operators are able to further virtualize their infrastructure obtaining coexisting but isolated slices, dedicated to serve specific customers or test new services before production, in a flexible manner. This approach allows for CAPEX and OPEX savings through the dynamicity and automated scalability of the virtual infrastructure, enabling faster and less expensive testing as well as a simplified roll-out of new services.

In parallel, the paradigm of Network Function Virtualization (NFV) decouples network functions for hardware boxes: traditional network functions become virtual entities which can be de-composed in elementary features and re-composed with additional virtual blocks in complex services which can be instantiated on-demand and delivered in several specialized flavours. The adoption of hierarchical virtualization and multi-tenancy at the NFV level would extend the concept of multi-level infrastructure sharing to virtual network functions and services. Going beyond the simple sharing of networking, storage and computing resources, the entire services would be further virtualized in several coexisting instances running on the same substrate of virtual infrastructure resources and across the underlying VNFs.

Hierarchical virtualization in NFV contexts extends the concepts of resource virtualization and multi-tenancy from virtual infrastructures embracing Virtual Network Functions and network services. Instances of VNFs and/or VNF chains are further shared among multiple-tenants to deliver several customized instances of the same services without requiring explicit management of the underlying virtual infrastructure. This approach enables more efficient, simplified and automated vertical scaling of the virtual services, as well as the creation of new business cooperation and more flexible value chains across multiple service providers.

However, the service-specific nature of VNFs and VNFs' chains brings new challenges in the hierarchical virtualization approach, with impact across the entire NFV architecture. In particular, new requirements arise through the three following aspects:

- the NFV Management and Orchestration (MANO) architecture, internal functional components and APIs;
- the NFV Infrastructure (NFVI) layer;
- the design of the VNF.

In particular, the **NFV Management and Orchestration (MANO) architecture** should be extended to enable a recursive deployment of the functional components related to NFV orchestration, properly integrated with hierarchical instances of the Virtual Infrastructure Manager (VIM) to support multi-domain and multi-tenancy also at the virtual infrastructure level. Additional functional components may be also required to handle multi-tenancy provisioning and management at the service layer. Suitable **information models** should be defined to provide unified description of VNFs and VNFs' chains at the service level, in terms of service characteristics, capabilities, interfaces and dependencies, going beyond the current infrastructure-based perspective of VNF requirements in terms of computing, network, storage and forwarding graph connectivity. These information models should support an abstraction of the virtualized service with different levels of granularity and they are the key to enable the vertical and horizontal interoperability of the virtual services with external entities. Moreover, extended multi-tenancy concepts should be supported at the **NFV MANO API** across all the different management levels. Going beyond the current scope of infrastructure, VNF and Network Service Tenants supported at the VIM, VNF Manager and NFVO respectively, a new tenant concept should be defined on top of single VNFs and VNFs' chains to enable the sharing of these virtual entities and the delivery of additional services to vertical customers. The support of multi-tenant virtual network service sharing has indirect impact also at the **NFV Infrastructure layer**, which should be able to support multiple levels of abstraction and virtualization, both for computing and network resources, providing suitable isolation mechanisms. Finally, **VNFs** should be designed to provide multi-tenancy at the service level, enabling the delivery of multiple coexisting services in isolated virtual contexts and implementing secure and open APIs for configuration and programmability of service instances.

4.1.13 Towards an organic NFV architecture

We have seen a huge amount of projects and SDN/NFV deployments already. All networking-related projects promise to solve all network architectural issues through NFV and SDN by putting a virtual network and a logically centralized control into scene. We are almost forgetting that there was a time in which physical networks existed, and not everything done over them must be forgotten. In particular, the Internet has become a huge distributed and reliable system that took years to deploy and that accumulates the expertise of the networking community, and also its historical trial-and-error assets. Despite the practical approach followed in the Internet design, or maybe precisely because of that, the Internet has become a huge success, as demonstrated by the commercial viability of the final product. We must not forget this impressive legacy.

Its evolution enabled an organic growth of the system in the sense that it adapted to a variety of environments and traffic profiles for which it was not initially conceived. This is what we should aim at when designing future software-centric architectures. In this respect, giving software the starring role of future networks has some relevant implications in terms of flexibility in deployment of network functions (e.g., its location), for instance. In turn, this has a remarkable consequence, that is, well-established borders in traditional networks now get blurred. This has technical and administrative implications, since the role of all stakeholders needs to be re-defined. Therefore, the traditional conception of what a domain is needs to be revisited.

Furthermore, it is also likely that the growing heterogeneity of technologies, particularly, when verticals come into play, may be better served by the more flexible paradigms inherent to software networks. However, it is also true that all this additional complexity must be appropriately handled if we want to fully harvest all its potential benefits. Given the novelty of ideas and concepts behind the SDN and NFV paradigms, most work up to now has been devoted to architectural design and functional evaluation. However, if our target is to offer commercial 5G service in 2020, there is a clear need to move one step beyond. There is the need to move from small-scale proof-of-concepts to larger-scale carrier-grade ones to evaluate if the hype around these novel architectural paradigms holds under stress. We need to now focus on the design, development, and evaluation of scalable, available, and reliable virtual networks and services. And networking history has given us some hints on how this could be done.

We must design an architecture that allows the network to organically grow and adapt to unplanned scenarios. Centralized control and management elements simplify network operation, but the envisioned complexity and heterogeneity will be hardly handled in this way. Therefore, there is a need to reduce the accelerated pace that network redesign has taken in recent years and think again about how to best keep the simplicity of management offered by centralized entities whilst maintaining the availability and reliability offered through redundancy and decentralized decision-making. Therefore, the traditional tension between centralization and decentralization has to be revisited again. And this time it will probably have more implications than ever before, given the size, heterogeneity, complexity, and flexibility that we are requesting to 5G networks. Depending on the network service, hierarchical and/or peering relations may be in order, hence resulting in hybrid systems. And this needs to be revisited for the various building blocks of the architecture (i.e., infrastructure, control, management and orchestration, and network management applications). The appropriate composition should be enabled by the architecture, bearing always in mind the resource efficiency vs. "carrier-gradedness" trade-off. Implicit to this discussion is also the oft-disregarded fact that control communication channels between decision points and devices under their control is not ideal, and so, distribution of decisions to avoid network operation disruption is also relevant in this context.

Making these decisions in a dynamic way in scenarios defined by a big number of parameters and uncontrolled dependencies will be hardly done at a theoretical level. In the same way self-organized networking (SON) use cases were identified in mobile network scenarios by 3GPP, extended SON use cases will have to be identified to consider end-to-end optimizations including RAN, transport, and core networks and the variety of services on top.

For the same reason, experimentation is increasingly needed for such complex systems, because it will be only in this way that some of the issues related with scalability and reliability will be identified for practical systems. In this way, we will be able to start accumulating the trial-and-error assets that will characterize this organic growth process.

Moving from a functional design and evaluation phase towards scalability, availability, and reliability, an evaluation phase is fundamental for achieving the objectives of starting the first 5G commercial deployments in 2020. The issues identified above go in this direction and will impact the way vendors design their products and operators deploy and manage their networks. Eventually, this will result in better services to the end-users, including the special needs of verticals. In summary, all stakeholders somehow involved in the network will be affected, given that such organic NFV architecture spans and impacts the whole network.

The combination of the starring role of ETSI in the standardization of NFV and the European competitiveness on the mobile front position Europe at the forefront of the 5G race. And only by creating the appropriate synergies between both worlds Europe will be able to maintain such a privileged position. The architectural discussion proposed above goes in this direction.

4.2 Radio Networks and Signal Processing

4.2.1 Cross-Layer Design and Optimization of Next Generation Networks for Multi-Class Services Support

Significant gains in resource efficiency and/or QoS are expected from a careful design and optimization of next generation networks that extend over the different network layers, including the physical (PHY) and medium access (MAC) layers. At the same time, this next generation networks should allow connections with different QoS profiles to accommodate different traffic patterns, and which are destined to very different users such as handheld devices or IoT machines.

Research areas for Cross-Layer Design and Optimization of Next Generation Networks supporting multiple classes of services:

- Design of integrated radio access schemes for massive connections with diverse QoS profiles, traffic patterns and/or delay constraints
- Design of multiple access schemes, control messages, and reference signals supporting large numbers of simultaneous connections in each cell and dynamic resource assignment to devices requiring different classes of services.
- Design of flexible transmission waveforms satisfying multi-class services requirements while minimising inter-class interference.
- Design of multi-service aware MAC layer with, e.g., hybrid automatic repeat request schemes capable of adapting protocol parameters to data rate requirements and delay tolerance profiles of different classes of supported services.
- Design of dynamic and adaptive cross layer communication protocols optimized for QoS and power-consumption.
- Dynamic channel assignment schemes that provide differentiated quality to each service (depending on the requirements) and can cope with the changing network conditions
- Cross-layer performance analysis to assess the impact of PHY and MAC layer parameters on the end-to-end data delivery delay and throughput.

4.2.2 Radio Resource Management for enhanced Broadcast/Multicast services

Wireless transmission has the capability of sending the same content to multiple users with the same radio resources, therefore saving the very scarce radio spectrum. Broadcast/multicast techniques can be applied to the transmission of many different services, where video, that will account for more than 70% of the content exchanged over radio networks, is of course an important one.

The main problem associated with multicast today is the provision of different quality of service (QoS) to the users. When serving a multiplicity of users that experience different channel quality, multicast by its nature will select the transmission mode that guarantees the performance of the user with the worst channel conditions. In this way, most users will perceive a worse QoS than what they potentially could experience.

Research areas for Radio Resource Management for enhanced Broadcast/Multicast services:

- Development of new Radio Resource Management (RRM) techniques to fully leverage the potentialities of multicast while preserving a suitable QoS.
- New scheduling alternatives for Multicast considering different optimization strategies and user grouping users according to the channel quality and expected QoS.
- Design of coordinated transmission to users within multiple cells.
- Coexistence of broadcast/multicast services and Machine-Type Communications (MTC) and the Internet of Things (IoT).

4.2.3 Agile and Dynamic Radio Access and Integrated Moving Networks

The 5th generation (5G) wireless and mobile communications system is expected to enrich the ecosystem of telecommunications systems by enabling numerous use cases associated with a wide range of requirements. At the same time, the future system needs to be efficient from multiple perspectives, such as energy and cost, in order to be sustainable. On this basis, as the service demands can vary over time and space, the networks cannot be designed for the peak demand anymore.

To handle the inhomogeneous distribution of increasing traffic demand over time and space in an agile manner, the network needs to react quickly and dynamically to fulfill the varying service requirements. One conventional approach for providing coverage and/or capacity is to deploy fixed small cells, such as picocells and relay nodes, overlaid by macrocells. Small cells may be deployed by operators at certain locations with power supply facilities, and the locations can be determined via network planning. However, the full operation of such a dense fixed small cell deployment for the peak demand is not needed anytime and anywhere due to inhomogeneous distribution of traffic over time and space.

In this context, agile and dynamic radio access networks (RANs) will comprise moving networks (aka, moving cells) that aim at enabling flexible network deployment in terms of on-demand densification. The possible access nodes include wireless relays integrated into vehicles (aka nomadic nodes), e.g., within a car-sharing fleet and taxies to serve users outside of vehicles, moving relays to serve in-vehicle users by eliminating penetration loss, and D2D relaying to offload the traffic from macrocells.

Research areas for Agile and Dynamic Radio Access Networks:

- A flexible backhaul needs to be employed, where the capacity of the backhaul link plays a crucial role in the end-to-end user performance.
- The network architecture needs to support numerous access nodes on the move in a scalable way.
- The network management needs to take into account service demands and changing radio topology.
- The availability of moving access nodes cannot be guaranteed in the target service region; thus, this uncertainty needs to be factored in;
- Efficient mobility management schemes should support the operation of moving cells while confining the signalling overhead.
- Tracking of large sets of mobile channels at high speed to enable advanced spectrally efficient and robust closed loop (massive) MIMO schemes in the moving backhaul links.

- Design of closed-loop and cooperative interference coordination in ultra-dense heterogeneous networks including modern solutions based on interference classification and interference matching [6].
- Resource allocation and resource slicing for versatile QoS services to meet key performance targets on outage, throughput, latency and energy efficiency.
- Efficient mobility protocols in integrated moving networks.

4.2.4 Network-Assisted Self-Driving Objects

Self-driving cars, and self-driving/coordinated flocks of drones are attracting more and more attention, moving from curious TED-style technological toys, to tools of practical use with a potentially revolutionary impact on everyday life. Nevertheless, the current approach consists of putting both the sensing and the main intelligence of the control system into the vehicle itself, somehow forgetting the fact that the vehicle is connected to a wireless network with infinitely more powerful "cloud" computing capacity. We envisage cyber-physical cloud-based systems where a large number of coordinating and moving objects (cars, drones) are jointly controlled via wireless networks. The communication problem here consists of conveying the sensing information (typically the result of some analogue source, such as a video camera, inertial navigation parameters, GPS if available, radio localization measurements) to a cloud computer which executes the (joint) controller, and conveying back the control commands from the controller to the vehicles. The standard separated "layering" approach (source, compression, TCP/IP, MAC/PHY) yields intolerable delay and sensitivity to post-decoding errors (resulting in dropped packets, NACKs, retransmissions, and so on). For such system, we advocate a low latency and low complexity joint source-channel coding approach, where the sensing source data streams are directly mapped over the channel symbols (e.g., using unquantized QAM and OFDM) through a randomized rank-reducing linear compression map (e.g., exploiting the rich theory of compressed sensing).

Research Areas for Network-Assisted Self-Driving Objects:

- Joint-source channel coding & compressed sensing: leveraging the fact that linear maps are well conditioned with respect to noise (graceful degradation, no packet losses or retransmissions).
- Efficient design of plant state estimators/trackers from compressed sensor signals. In particular, in combination with computer vision and feature extraction from video.
- Investigation of the feasibility region of the joint control and communication problem: we need to study the intersection between the region of sampling rate, distortion and delay achievable by the network, and the "demand" set of the controller. If such intersection is not empty, the control and communication problem is feasible. Otherwise, the controller has to step back, possibly put some vehicle in "hold" mode, and ask for less.
- Extension of the same ideas to similar systems, e.g., very large number of home batteries, coordinated via wireless in order to guarantee frequency stability of the power grid in the presence of renewable sources, industrial robots in a wireless-controlled plant, and so on.

4.2.5 Highly-flexible channel coding for forward error-correction and incremental redundancy

Error correcting codes are at the heart of every communication system, allowing for reliable transmission over inherently unreliable channels. While current wireless communications are based on point-to-point transmission with reasonable reliability and latency, future 5G networks will include a wide variety of scenarios, requirements and services. Error correcting codes in such future networks need to be able to cope with point-to-point transmission, as well as broadcast and multiple-access schemes; they need to provide very high code rates for mobile broadband and very low code rates for ultra-reliable transmission; they need to support very long codes for highly efficient data transfer and very short codes for machine-type communication; and their decoders need to operate at low latencies and low power consumption.

The dominating channel code in current systems, like 3G and 4G, is the turbo code. Turbo codes were first published in 1993, and they have changed the world of channel coding, showing an error-rate performance that at that time was believed to be impossible at feasible decoding complexity. Besides their excellent forward error correction performance, turbo codes also provide robust incremental redundancy schemes, providing for efficient HARQ protocols. Highly optimised turbo decoder implementations are available and are used systems nowadays.

Despite all their advantages and mature status of development, turbo codes will not be able to fulfil all requirements of future 5G networks. Here are a few examples: (i) In the short-length regime, polar codes under list decoding achieve lower error rates than turbo codes. (ii) In the large-length regime, the decoding of low-density parity-check can be parallelised to a higher degree than turbo decoding. (iii) In the low-rate regime, turbo codes are known not to perform very well, and polar codes are a well-performing alternative, (iv) LDPC codes show 10 times lower power consumption than Turbo codes, (v) LDPC codes require 5 times less chip area for decoder implementation.

Research Areas for Highly-flexible channel coding for forward error-correction and incremental redundancy:

- Novel channel coding scheme with rate and length flexibility to support diverse services and communication scenarios
- Efficient hardware or software implementation with the required constraints on throughput, latency and energy efficiency.

4.2.6 Networked Signal Processing and Fronthaul/Backhaul

The evolution towards 5G radio networks will include a paradigm shift in signal processing, namely the use of distributed and dynamic processing of radio signals in different parts of the network. In the simplest scenario, processing will be shared between so-called remote radio-heads (RRH) and centralized data centres which perform joint processing of the information coming from many RRH. In future networks, we foresee that the processing split between RRH and data-centres will be dynamic depending on traffic load of the network, power consumption of the network, synchronization accuracy between RRH, processing capabilities of the RRH, and quality of the wireless fronthaul links. Moreover, we foresee the requirement for a very dense deployment of such processing-capable RRH with many different types of network interconnections with the data centres and mutual synchronization granularities. Note that this is in contrast with the notion of small-cells, where an entire base station MODEM and protocol stack is hard-coded in the processing elements of the small-cell device and little or no signal-level collaboration with other small-cell devices is feasible. This evolved RRH architecture is a truly distributed processing environment and will call for new computer science skill-sets for the radio system designers. When these transport architectures are coupled with the growing number of sensor-based applications, the processing elements could also be used dynamically for distributed data fusion at the application level.

Research areas for Signal Processing and Fronthaul/Backhaul:

- Massively distributed computing for radio signal processing. This is a complete paradigm shift and requires new skill-sets for the radio system designer.
- Potential for synergy/resource reuse with distributed application-layer signal processing (e.g. sensory data and multimedia)
- Network synchronization mechanisms and centralized timing/frequency resynchronization through signal processing.
- Design of efficient and reliable wireless fronthaul at mmWaves, i.e., the technology behind the so-called "wireless fibre"
- Optical technologies for Massive MIMO, Radio beam steering, Dynamic capacity allocation among pico cells.
- New efficient fronthaul transport protocols

4.2.7 Mobile Edge Computing and Mobile Cloud for Low-latency Mobile Services

Two relevant trends have recently emerged: wireless networks are required to support real-time applications with strict latency constraints (e.g., mobile gaming, IoT-based applications), and user terminals can conveniently offload computational and storage tasks to the Cloud. Unfortunately, these are often conflicting trends, as offloading services and applications to the infrastructure clearly leads to increased latency.

Mobile Edge Computing (MEC) can represent a valid solution to this problem: computing and storage resources can be placed at the edge of the network [7] so that task offloading can be performed with little additional latency. The "edge" concept can be pushed further to include (i) other user terminals in the set of available resources: each user device could indeed offer its computing and storage capabilities and constitute, along with its neighbouring devices, a mobile cloud. The mobile cloud would thus allow the exploitation of consistent resources that lie largely unused today and (ii) clustering local available edges of the network to enforce edge cloud capabilities that can be provided to local service requests [8].

The mobile cloud can include both mobile and stationary devices, each equipped with a CPU, memory and communication capabilities. It is intended to support distributed execution of intensive computational tasks as well as distributed storage of large content (e.g., IoT-based data). D2D communication can be used to build the underlying network and ensure the transfer of task input/output data and content items between devices. It can be conceived to closely interact with the local infrastructure (MEC), or to operate independently when it might be more convenient for the network or in the case of emergency networks.

Research areas for Mobile Edge Computing and Mobile Cloud for Low-latency Mobile Services:

- Novel protocols and algorithms are needed to efficiently manage the mobile cloud, in spite of the nodes mobility.
- Advanced clusterization techniques in order to federate locally edges of the network to serve multi user requests.
- Computational and storage tasks have to be properly defined, in such a way that their offloading is feasible in terms of latency and size of the transferred data.
- Tasks have to be assigned to the mobile cloud members so that (i) the tasks demand matches the resources offered by mobile devices as well as the communication capabilities between devices, and (ii) the task allocation is fair in terms of energy and resource consumption.
- User mobility should be predicted and accounted for since it impacts the communication capabilities and on the possible migration of virtual machines and, hence, the system performance in terms of latency and reliability level.
- The opportunity to offload tasks to the mobile cloud or to the MEC should be evaluated and an optimal split should be found. This case includes the case where the MEC is at the residential access network owned by the user and the mobile devices are all owned by the user.
- Incentive schemes are needed in order to motivate user devices to be part of the mobile cloud, unless they are all owned by the same user.

4.2.8 Processing of side-information for improving network performance

The acquisition and evolution of channel state information (CSI) is important for network throughput and dependable communications. CSI estimation and prediction in transceivers is heavily based on linear estimators and largely ignores the exploitation of side-information that promises large benefits. Significant side-information is in fact expected to be available in the next generation network, enabled by the ubiquitous availability of geographical information services and through the fusion of sensor data, e.g. device location and orientation, device speed, environmental map, device cooperation, road infrastructure information, driving route information, positioning and social networks, etc. Since, the number of degrees of freedom for the channel state increases strongly with the advent of wideband, high-mobility, and cooperative transceiver technologies an exploitation of side-information is a key focus for next generation systems.

Research areas for the processing of side-information:

- Identification of side-information type to be used for high throughput.
- Identification of side-information type to be used for dependable communications.
- Side-information exploitation approaches, also from unreliable sources.
- Side-information gathering and sharing techniques.
- Privacy, security, authentication and owner protection of side-information.
- Multidimensional random sampling in Inferential networks for spatio-temporal signal reconstruction of stochastic processes supporting emerging applications (e.g., environmental monitoring, crowd tracking, dynamic objects control).

4.2.9 Socially-aware dynamic resource allocation and network optimization

Device-to-device Communications is a core component of 5G networks due to its key role in boosting capacity, traffic offloading as well as enhanced user experience. Developing policies, algorithms and protocols at the intersection of Device-to-Device communications, social-aware networking and content caching at the edge, is a promising research direction for future networks. Such techniques will prove useful in large venues with high user density, e.g., sports games, fairgrounds, etc., where the number of users exceeds the network capacity, giving rise to congestion. This, in turn, gives rise to the new research paradigm of socially-aware dynamic resource allocation and wireless network optimization.

Research areas for socially-aware dynamic resource allocation and network optimization:

- Exploitation techniques for mobile user's data, e.g., spatio-temporal characteristics of the mobile users' traffic.
- Content delivery load reduction through offloading to local device-to-device communications through proactive, cooperative and opportunistic content caching.

4.2.10 Wireless ultra-reliable low-latency device-to-device communication

Ultra reliable low-latency device-to-device (D2D) communication is strongly demanded in vertical markets such as industry 4.0 or intelligent transportation systems (ITS) with applications such as:

- Replacement of cable connections to mobile actuators or sensor in industrial cyber physical systems (CPS), and
- For the provisioning of redundant sensor information (radar, optical, etc.) for autonomous vehicles to achieve a level of safety that is significantly higher than that of today's road traffic.

In both cases harsh radio propagation environments lead to non-stationary fading processes that have a strong impact on the reliability and latency provided by the wireless communication link. Compared to other machine-type communications (MTC) links requirements, ultra-reliable communications are particularly sensitive to ultra short and quasi deterministic delivery delay (towards low latency and zero jitter).

Research areas for wireless ultra-reliable low-latency device-to-device communication:

- Measurement, modelling and emulation of multi-node communication channels in the targeted 5G frequency bands: due to the fast movement of vehicles and the low height of the antennas the resulting fading process has strong non-stationary properties. Hence its second order statistics such as the root mean square (RMS) delay spread, RMS Doppler spread and path loss vary strongly over time. For multi-node vehicular scenarios these properties must be measured and numerical models need to be developed. Moreover, the design of ultra-compact multi-antenna systems integrated inside devices should be investigated in order to improve the robustness of radio link.
- Low-latency physical layer modulation formats for low packet error rates at short packet lengths: for reliable communication links the utilization of all locally available diversity mechanisms (time, frequency, space) is needed.

- System-level test methodologies: the strong dependence of the propagation conditions on the coordinates of the transmitter and receiver requires system-level test methodologies that take this dependence into account enabling the test, e.g. of control algorithms for connected autonomous vehicles.
- SDR testbed implementations: a software controlled radio platform enables the flexible test of different physical layer modulation formats and multiple access control algorithms.
- Suitable protocols for low-latency and zero jitter operation in 5G networks: a distributed multiple-access control (MAC) mechanism is needed that enables scheduling within the context of a local bubble supervised by the 5G network.

4.2.11 Satellite networks as part of next generation networks

Next generation networks shall address the needs for mobile data and connectivity of those 3 billion people who live beyond the reach of mobile networks in rural, suburban and remote sites. Cost effective solutions for ubiquitous communications need to go beyond the 5G mainstream. New technologies are radically bringing down costs for satellites, including new launchers, assembly line satellites, on-board processing, and so on. These technological advances in satellite technology need to be matched by a corresponding dramatic development of signal processing and communication capabilities, as well as by a holistic and systematic understanding of hybrid terrestrial/satellite communication network design for seamless services over hybrid terrestrial satellite networks. Satellite communications are the only means to provide full coverage and mobility. Research and innovation has to be done in order to reduce/optimise the satellite service provision costs.

Research Areas for Satellite networks as part of next generation networks:

- Development of suitable channel and interference management models.
- New interference resilient codes, interference constructive modulations and joint scheduling and precoding or detection schemes.
- Analysis of the compatibility and of the required adaptations of the radio interface (waveform, framing structure, etc.), and related impairments countermeasure, developed for terrestrial scenarios to satellite systems.
- Resource allocation and management techniques, e.g., non-uniform cooperative resource allocation strategies considering various parameters specific to each cell, segment and access/backhauling technology. In order to lower deployment costs, satellite systems aim at full frequency reuse among their users, and between terrestrial links.
- Design of integrated satellite terrestrial network management, control techniques and hierarchical backhauling architecture.
- Study and demonstration of the feasibility of efficient integration of Satellite and Terrestrial networks.

4.2.12 Satellite signal processing and communications for new network architectures and constellations.

New network architectures and constellations, e.g., High Altitude Platforms and Low/Medium Earth Orbit constellations promise lower communication delays and better integration with aerial and terrestrial networks, but pose new problems into the required signal processing for communications.

Research Areas for Satellite networks as part of next generation networks

- Channel models for the user and the feeder links accounting for Doppler shift and Doppler rate much more important than in Geostationary Satellites.
- Fast synchronization algorithms for un-interrupted services.
- Advanced multicarrier waveforms robust to time-varying channel conditions.
- Radio Interfaces for new constellations and network architecture, e.g., mega constellations, including possible direct link to the user devices.
- Advanced Return Channel Interfaces for Remote or Distributed Access Stations.
- Gateway and inter-satellite link design to optimize cost and deployment efficiency.
- On-board signal processing. to further increase downlink data rates.
- Space qualified technologies supporting advanced signal processing for the entire satellite life.

4.2.13 Extremely high and optical frequency satellite communications

New satellite links can employ extremely high frequency (EHF) bands and particularly Q/V band (40/50 GHz) and W band (70/80 GHz) for the user and feeder links. For example, two types of scenarios can be addressed: i) Aeronautical systems operating at EHF bands (AERO) to provide broadband Internet services to the cabin of commercial airliners; ii) Systems accommodating feeder or mesh professional links at EHF bands (MESH) to provide ultra-high data rates involving professional users with medium to large ground stations. For the inter-satellite communications, free-space laser links shall be addressed to achieve higher directivity, data throughput efficiency (several Tbps), and power efficiency.

Research areas for extremely high and optical frequency satellite communications:

- Development of accurate channel models in all application scenarios.
- Design of pointing and tracking algorithms for optical free-space communications.
- Development of optical communications techniques for EMI, low mass, low volume and mechanical flexibility characteristics of fibres on board.
- Photonic systems such as Photonic Payloads (including electro- photonic ADCs and BFNs) and fully photonic communication and sensing.
- Digital processing, analogue processing and photonic processing of microwave signals.

4.2.14 Terahertz Communications

All widely deployed commercial wireless technologies used today operate on frequency bands below 6 GHz, but millimeter wave (mmWave) frequencies such as 28 GHz, 60 GHz have gained significant interest for 5G cellular access with the targeted peak data rate more than 10 Gb/s.

While this is certainly the way to go, the rather limited available consecutive bandwidth (up to 7GHz at most), poses a constraint on the maximum achievable individual and aggregate data-rates. For example, according to the Edholm’s law of bandwidth, wireless Terabit-per-second (Tbps) links will become a reality by year 2020. With 7 GHz of bandwidth, this would require the use of a modulation and coding strategy with a spectral efficiency above 100 bits/s/Hz. This is not realistic at all and has not been achieved at lower frequency bands, where the technology is much more mature. A complete paradigm shift in solid state and radio frequency (RF) technologies as well as in their operation in wireless networks is required to provide the high rate connectivity, the core lubricant of digital society and its renewal. In this context, Terahertz (THz) band communications are quickly gaining momentum. THz-band communications are envisioned as a key wireless technology to satisfy real time traffic demand for vehicular communication, by alleviating the spectrum scarcity and capacity limitations of current wireless systems. The THz band is the spectral band that spans the frequencies between 0.1THz and 10THz. While the frequency regions immediately below and above this band (the microwaves and the far infrared, respectively) have been extensively investigated, this is still one of the least explored frequency bands.

Research areas for TeraHertz Communications:

- Transceivers, antennas and solid state circuits supporting Tbps data rates at (sub-)THz bands with an order of magnitude efficiency improvement ready for practical development.
- Waveforms, equalization, interference management, and information theoretically optimal coding schemes for Tbps radio systems with massive antenna arrays and true TX-RX hardware (HW).
- Optimization theory based, truly energy-efficient radio networks (protocols, control algorithms, routing solutions with tolerable control and signalling load) capitalizing the Tbps TX-RX HW and waveforms to enable sustainable connectivity and the vertical applications of the digital world, e.g., vehicular communications.
- Precise channel modelling.
- Dedicated antennas and both related miniaturization and massive antenna design.

4.2.15 SDR and CR for M2M communications

Despite a common agreement about the lack of readily available radio frequency channels for new communication systems, much more needs to be done to promote the efficient use of the spectrum. Notwithstanding the intense research effort in the last few years on Cognitive Radio (CR), the spectrum use still follows very rigid rules, and dynamic allocation mechanisms are almost exclusively applied in free bands. It is precisely in these bands where some innovation ideas have arrived to stay, and some companies are exploiting new inspiring business models mainly focused to Machine-to-machine (M2M) communications. The different mechanisms provided by CR for multiplexing services on a given spectral band should be also exploited in licensed bands; research effort is needed to specify new robust low-power waveforms able to coexist with primary terrestrial and satellite services. A secondary service cannot interfere the primary use of the spectrum, and cannot expect any special treatment from the incumbent; thus, it is very relevant to deal with the interference from the primary service. In addition, device-to-device communications and full duplex operation are becoming a reality, and intra-system interference needs to be considered in addition to the primary interference. Lab work is essential for proof-of-concept tests, and the availability of flexible platforms can attract more researchers into the field.

Research areas for SDR and CR for M2M communications:

- Low power programmable systems in the form of fast prototyping platforms, SDR-like, to develop new communication solutions.
- Agile and flexible RF front-ends, including compact antennas and even with full-duplex capability and including frequency bands all the way up to hundreds of GHz, allowing for dynamic selection of transmission and reception frequencies, as close as desired.
- Criteria definition to switch between Full-Duplex and Half-Duplex and Self-Interference cancellation technique in MIMO Full-duplex context
- Interference context aware adaptive transmission (PHY/MAC) protocols able to optimize the operation of M2M communications in crowdy shared spectrum bands.
- New methodologies to test the proper operation of secondary users.

4.2.16 Open-source tools for radio network innovation

Open-source has made a very significant impact in the extremities of current wireless and fixed networks, namely in the terminals due to the Android ecosystem and in cloud infrastructure due, in part, to the OpenStack ecosystem (http://www.openstack.org/). The access network components, or more specifically the embedded systems comprising the radio modem signal processing and access-layer protocol implementations, can be seen as the final frontier for open-source implementations. This is mainly due to the revenues generated by intellectual property in this segment of cellular technology. Through innovation in open-source licensing, the OpenAirInterface (OAI) Software Alliance now provides a joint academic-industry ecosystem for the core (EPC) and access-network (EUTRAN) of 3GPP cellular systems with the possibility of interoperating with closed-source equipment in either portion of the network. In addition to the huge economic success of the open-source model, such an ecosystem can be a tremendous tool used by both industry and academia to jointly foster wireless innovation. More importantly it ensures a much-needed communication mechanism between the two in order to allow academia to have a direct impact on cellular technologies which are controlled today by major industrial players in the wireless industry. In the context of the evolutionary path towards 5G and beyond, there is clearly the need for open-source tools to ensure a common R&D and prototyping framework for rapid proof-of-concept designs. This will reduce the cost of radio network components since the great majority of the development will be collaborative and include university, research centres and a more general population of programmers as part of the workforce. This is also valuable training for engineers that will later join industries using similar or the same core implementation.

Research Areas for Open-source tools for radio network innovation:

- Developing of a software library of signal processing and upper layer protocol fundamental functions with controlled execution delay, such that they can be interconnected and run on general purpose machine (e.g., cloud computers) and provide predictable and reliable performance.

4.3 Optical Networks

4.3.1 Optical Networks Enabling the Digital Environment

The Knowledge-based economy gives us the opportunity to operate with no geographical boundaries and, in response to this, global enterprises strive to become “virtualized” i.e. location-independent, themselves. This framework sets as the main challenge the creation of a unified ***Digital Environment*** where networks, processing and storage facilities, intelligence, control and management mechanisms as well as any kind and type of “machines” are amalgamated and then they are sliced and handled as a single *commodity* (service) afterwards. Actually an important challenge is the introduction of *automation* in the processes associated with creation, control, modification and removal of functions as well as the handling of the available resources across different global service agreements. The creation of this *Digital Environment* is a process that is implemented through a number of stages; although these are presented here sequentially, for conceptual reasons, it is almost certain that they will appear concurrently. The stages are:

Integration of Networks in the Access: Although it is crucial to ensure the European leadership in both r*adio* and f*ixed network* technologies, as stand-alone business sectors, it is equally important to consider their synergy, taking into account their relative merits. In fact, leadership can only be exercised through the exploitation of this synergy. The exploitation of a technology-agnostic network last-drop framework will allow making the most from the two technology platforms while alleviating the converged network from unnecessary duplications and inefficiencies that occur when building distinct networks for each one of these technologies. Thus, the challenge we are facing is to identify lower CapEx/OpEx and better managed platforms based on a combination of radio and fixed (optical) network technologies.

Integration of Networks into Clouds: Today we witness a convergence between network infrastructures and datacentres not only because datacentres are necessary in supporting Cloud services, the fastest growing segment of the overall network connectivity market, or because the Cloud postulates a dynamic and flexible network between remote locations. It is also because of the need to explore *Virtualization* to its full extend. That is, to utilize *virtualized* resources for datacentres and NFV alike regardless of location, while introducing advanced optical communication systems within the Datacentre Interconnect network (DCI). It is necessary to continue research in understanding the full benefits and challenges of virtualization, regardless of the geographic location of the physical resources. In this quest, both evolutionary approaches (where networks and datacentres are clearly identifiable entities) and clean-slate approaches (where networks and datacentres are integrated, in a single entity) are of interest.

Integration of Clouds and Machines; the "close" and the "far": The complete Digital Environment is finally formed with the integration of Clouds with Augmented or Artificial Intelligence (AI) entities, *Machines* and *Things* of any kind, equipped with communication and processing capabilities. Numerous business and welfare sectors would benefit from this convergence: from production systems, to energy, health and education etc. As an example:

- The convergence of compute and storage technologies with AI and networks is generating new production systems that are reshaping the manufacturing landscape, designating a return to local production paradigms, which will be managed and connected by a new global-local, cloud-integrated network of virtual and physical entities and workers.
- A new agrarian revolution is *ante portas* based on the convergence of precision farming and the autonomous vehicle control of robotic appliances.
- The embedding of “managed” intelligence in the main social interaction forms creates the condition where social activities will be implemented by means and through a layer of AI appliances.

The challenges to bring these innovations into life are for the short-range networking (the "close") as well as the long-range networking (the "far"). Regarding the former, in the access part the main challenge is a technology-agnostic service definition, creation and management and the integration of a plethora of “machines” and “things” in a seamless way. On the other hand, optical backbone networks do not only need to scale to efficiently and securely transport data but also to provide means to remotely control/monitor the numerous transport systems, local networks and a diverse range of families from Machines and Things.

This is a ground breaking development liberating human activity from any “localization” boundaries. However, the proclaimed "death of distance" is still illusive and it requires persistent multi-disciplinary research to be fully understood in a number of fields like network architectures, technology for truly low-cost end-user terminals, seamless orchestration/coordination mechanisms, etc.

An Overarching Operating System (OS): Networks, Datacentres and Machines/Things are today employing dissimilar OSs. To jointly orchestrate an unspecified, dynamically changing assembly of these resources requires the development of an overarching OS based on *open* software. This will be an important step towards automation in response to dynamically changing services and user demands. This OS will have a comprehensive end-to-end view of the network via a set of distributed autonomous controllers across different network domains and technologies. In the final phase in the formation of the Digital Environment, the SDN paradigm (either as middleware or as the main OS), needs to be extended to incorporate (apart from the Clouds) the OS of a constellation of diverse Machines and Things, to allow the joint provisioning, orchestration and resource slicing.

This converged Digital Environment will operate under gradually more **dynamic conditions** and it will create an unprecedented **volume of data** (Big-data) as elaborated in . These two trends are challenging the entire existing network architecture: network architectures with fewer segments (network collapse) and with fewer interfaces for data to cross in an end-to-end path, are necessary to maintain manageability and CapEx/OpEx at affordable levels.

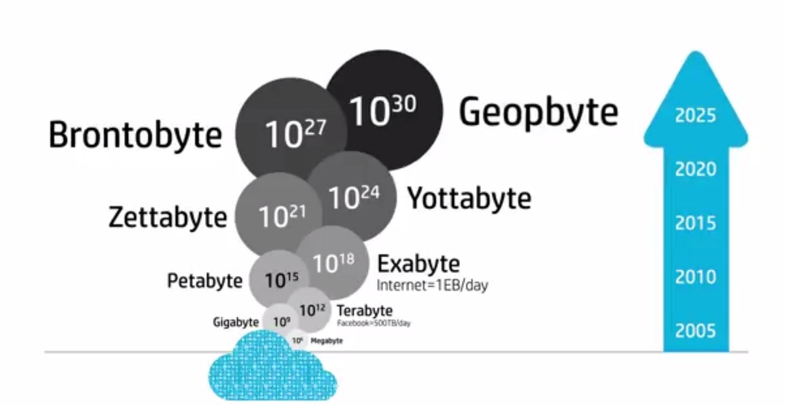

Figure 2 - The upcoming data explosion [source: Hewlett Packard, the machine]

In such a network architecture with fewer segments, the optical backbone/backhaul networks, which today are demonstrating a slowly varying traffic profile only because of the successive number of aggregation and grooming levels employed, would not only have to transport significantly higher traffic volumes but also would be more dynamic. To address the challenges of this new phase of networking, research should focus on the following priorities:

Focus on Manageability: The unified Digital Environment should exploit adaptable and flexible platforms implemented by means of automated and programmable infrastructures. Automation exploiting cognition and self-organization principles, programmability at the data-plane and the coupling with an open software-based control plane are important assets to even faster and more robust operations in the Digital Environment. Research effort should be expended to match the reconfiguration/response times in the converged Digital Environment, alleviating the current disparities as the one shown for example in . To this end, continuing research effort for dynamic and highly manageable optical data-planes is also in scope.

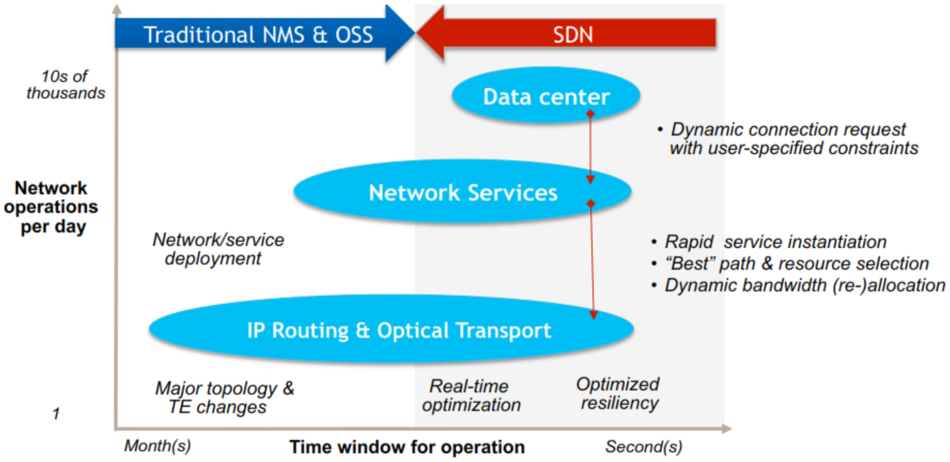

Figure 3 - Response times in the Cloud [source: Alcatel-Lucent]

Focus on Decentralization: The generation of a vast volume of data that is changing too fast and is setting too diverse requirements on the processing systems, creates the conditions for a “perfect storm” in centralized decision making systems. To cope with the emerging traffic volume and dynamicity, research should persistently explore decentralized and distributed information processing and traffic forwarding architectures and platforms. Moreover, the migration of Cloud services to the Edge, a trend that is moving computation and storage along with optical network connectivity closer and closer to the end-user, is an important tool to avoid the capacity crunch in optical core networks while providing superior QoS and QoE performance e.g. latency for mission critical applications.

Focus on Homogenization: The ambition of enterprises to operate both locally and globally, the opportunity to mix and match different suppliers and different types of connection technologies (fibre, microwave, satellite, etc.) and reliability levels, all these have the demand to use the same set of standard and scalable technologies as a common denominator. This requires a new brand of ubiquitous network building blocks like: standards-based compute systems, standard and self-adapted “optical modems” that autonomously adjust operational parameters, globally accepted encapsulation/framing technologies orchestrated by an open new automated control paradigm and operated by means of open software etc. These are the necessary ingredients for truly *programmable optical networks* capable to reconcile the “close” and the “far”.

Focus on Performance: Emphasis should be given to network performance-oriented, rather than technology-oriented, objectives. In particular, end-to-end, i.e. across the entire chain of interconnected systems, service-oriented performance indices are necessary to facilitate and encourage the benchmarking of competing and alternative technologies and platforms without prejudice. There should be no primacy of one technology over the other; instead, a sober analysis of the pros and cons of each technology will allow selecting the best alternative for the particular conditions.

4.3.2 Access Network Integration

A new landscape in Access: global enterprises, and in particular global service providers, in partnership with local providers, will leverage advances in the architectural framework, to be able to slice portions of their multi-tenant facilities, offering in this way diversified services to their customers that may span different countries or continents. To turn this into reality, two major advances are necessary in the access network part:

- The migration from the current overlay of independent transport networks employing a diversified portfolio of technologies to a *converged* network infrastructure.
- The virtualization of as many functions as possible of what a Central Office (CO) or a C-RAN are providing today, to allow the creation of a ubiquitous remote node hosting both wireless and fixed-line networks.

Currently, several architectural alternatives have been proposed regarding the optimal *point of network/service convergence* in the Access. Based on the various studies, the optimal point can be anywhere between the Metro and the end-user’s premises and the debates are heated. The discussions are spinning around the optimal scheme for the *x-hauling* a term which is a code name for the front/mid/backhauling (MFH/MMH/MBH).

Actually, the identification of the optimal point of convergence will significantly impact the functionality a particular architecture is offering, the performance it provides from the end-user's perspective, the technology an operator need to deploy to support it as well as the total number of interfaces involved in end-to-end data transportation and finally the overall system's CapEx/OpEx. Moreover, it has the potential to reshuffle the traditional segmentation of network layers since, based on what some alternatives are preaching, the rigid distinction between optical Access, Aggregation, Metro and Core could be replaced from a different architectural framework that is leading to *Access-Core integration*. The effectiveness of this integration becomes even more appealing when it is considered in parallel with *homogenization* and the introduction of network automation: it will lead to an optical network that will be built–to a larger degree-utilizing a general purpose infrastructure and it will consist of distributed clouds fused together into cohesive dynamic services platforms.

Today in Access, a multitude of "verticals" are relying on a large number of heterogeneous LANs to provide the necessary connectivity. These LANs are exploiting a diversified portfolio of technologies ranging from xDSL, G-fast, to point-to-point Ethernet, PONs and mobile networks, WiFi and WiMax, etc. The corresponding network transport infrastructures are running in parallel, unaware and independent of each other leading to functionality multiplication, CapEx waste and management complexity.

In the opposite direction, the scope of system/network convergence in Access is to identify the architectural framework for ensuring the integration of all these systems into a single, ubiquitous technology-agnostic optical transport platform. A key asset for such a system would be a technology and ownership-agnostic network control plane, capable of interacting with the application or service layers and capable to complete this processing automatically without intervention from the end user. We believe that this is a feasible objective that will be further supported from the programmability and accountability brought about by SDN and the advanced data storage and computation technologies. This new framework will make the most from any and all of these fixed and wireless technologies in providing user-transparent and end-to-end QoS-assured performance.

4.3.3 Access Capacity Increase

The demand for higher end-user capacity is progressively more and more driven by the necessity to provide what a digitized industry needs and, generally speaking, by the necessity to revitalize economic activity. It is widely recognized that optical fibre is the preferred high capacity transportation medium and a converged network will use this technology to transport traffic to and from the heterogeneous Access LANs.

Regarding fixed-line networks within the EU member states, various FTTx schemes are proliferating and optical fibre services are now widely deployed, whereas new standards, such as NGPON2, have been developed and trialled. Given the significant capital requirements for full FTTH deployment, alternative schemes offering fibre-like services based on copper, like G-fast, are considered. These schemes are sharing the optical feeder network of PONs since attenuation is limiting the distance such systems span to a few hundred meters or less. Equally, indoor wireless distribution systems and Wi-Fi installation may also take advantage of PON configurations to provide rates at the Gb/s range. To increase the capacity of PONs a higher number of WDM channels, each one operated at 10 Gb/s and up to 25 Gb/s, could be employed. The availability of the optical and electronic components and devices at an access-compatible cost will be a key enabler to realize the future converged networks.

Similarly, with the widespread availability of smart terminals, mobile services have increased exponentially, putting pressure on operators to introduce an even higher number and more diversified data services. As a result, operators and service providers are facing the difficult task of increasing the transport capacity of their mobile networks. In particular, the LTE-A deployment scenarios are using Common Public Radio Interface (CPRI) to set-up MFH links between a Remote Radio Head (RRH) and a Base Band Unit (BBU). Moreover, the wireless networks are exploring C-RAN with large scale centralization architectures where a vast number (thousands) of RRH are connected to a centralized BBU pool. The maximum distance can be 20 km for 4G system (LTE/LTE-A), while longer links (from 40 km to 80 km) for 3G and 2G systems are employed. This approach is causing many challenges since the current CPRI interfaces used in MFH induce huge bandwidth inefficiencies as they require multi-Gb/s data streams from Mbps radio channels. This is especially true in the case of multiple antennas employing MIMO techniques which results in a considerable increase of the MFH link's transportation capacity: the bandwidth of a single antenna with a single carrier is 1.229 Gb/s. For a RRH with 8 antennas and bandwidth of 40 MHz, the rate of the MFH interface could be up to 19.67 Gb/s, which is placing stringent requirements to optical transportation technology in terms of cost and power consumption. In the new generation of wireless systems employing array antennas and beam forming, the transport capacity requirements can be very high. So, mobile networks are also bound to use fibre for data transportation, considerably adding to capacity exhaustion.

Another parameter that should be taken into account is that dynamic capacity allocation between picocells, in response to the varying traffic demand per cell, will be extensively used since it will significantly improve the efficient use of a mobile network operator's resources. This dynamic allocation implies that capacity which is not needed at one picocell may be shifted to another cell where the demand exceeds the currently assigned capacity. This dynamicity of x-hauling is placing not only simply capacity requirements but also presents a pressing need to provide fresh architectural, monitoring and control strategy answers to a converged fixed-wireless network. In turn, these developments are setting a higher barrier to the requirements for effective monitoring and control strategies and fast routing in the backhauling parts of the network.

With the upcoming new round of innovations and services around the corner such as the widespread introduction of (semi)autonomous vehicles and Industrial "digitization" schemes such as distributed manufacturing, the expected dominance of IoT applications etc, will further exacerbate the aforementioned capacity requirements.

Conclusively, the research direction should be two-fold: a research objective should be to increase the capacity of existing fixed-line optical systems, in order to keep up with the growth of applications and services. However, even this would not be enough; in fact, it may have an adverse effect to the overall system stability if it is not accompanied by persistent research effort to come up with a converged network infrastructure in the Access that scales both fixed and wireless technologies proportionally to the capabilities these technologies have. A converged infrastructure will allow wireline and wireless operators and competitive global enterprises to concurrently use the same single, ubiquitous, infrastructure making the most of each technology something that will significantly reduce the overall rollout capital (CapEx) as well as OpEx.

4.3.4 Encapsulation Protocols and Latency Optimization

The convergence of a multitude of Access technologies will put additional strains on transport technologies in the Access network. The transport mechanism should be able to encapsulate dissimilar protocols and applications with different latency and packet loss requirements (QoS performance, in general) into a common format. Moreover, the essential packet-based services should offer statistical multiplexing functionality for efficiency and scalability. The challenges on transportation are particularly acute when considering a converged wireless and optical wireline network. Specifically,

- The centralization offered by C-RAN architectures, the BBU pooling, allows concentrating the baseband resources in a single location, so the wireless resources are dynamically allocated between different BBUs to support co-operative radio techniques such as Co-operative Multipoint (CoMP), self-adaptive load balancing, joint dispatching, etc. These C-RAN operations are facing scalability limitations today due to latency and other synchronization constraints between stations distributed in a large area; functions that previously resided in the same base station are now prone to variable delay due to geographic separation.
- In the MFH/MBH context, another challenge is to employ guaranteed-service transport techniques, which are taking full advantage of statistical multiplexing techniques without incurring additional delay variation. Architectures and transport mechanisms are needed to ensure low latency ensuing guaranteed, circuit-like performance into packet networks and providing timing transparency combined with statistical multiplexing gains for increased resource utilization efficiency.

4.3.5 Universal Remote Node Architecture

An architectural entity strongly linked to the concept of a converged access network is that of a Universal Remote Node (URN). A URN is a kind of a gateway node for all heterogeneous LAN technologies detailed in the introduction of this section. The URN could serve as an effective convergence point and its location could be anywhere in the Access or Aggregation segments. An important feature of the URN is that is it may collocate content distribution servers (CDNs), or even a mini Datacentre. Indeed, the URN is ideal for efficiently distributing cloud-based services to the users of fixed, mobile and Wi-Fi networks. These features will allow the URN to support a virtualised RAN (vRAN) functionality proliferating local cloud resources (“cloudlets”), which is an important step to buttress latency sensitive applications. Moreover, thanks to the virtualization features hosted in the URN, the node may implement a range of network functions like Broadband Remote Access Server (BRAS), Evolved Packet Core (EPC), Broadband Network Gateway (BNG) and security gateways. It may even include virtualized BBU latency and reliability-improving technologies. All these are important pieces in the broader scope of creating a technology-agnostic and programmable SDN-based access infrastructures based on commodity hardware which will further boost the *homogenization* process and the integration of all network segments.

Another important advantage of URN schemes is that they will restrict the amount of traffic that is, unnecessarily, forwarded to Metro and Core networks since many functions will be locally implemented. Moreover, the URN concept will increase the utilisation of network resources, the synergies with CDN caching schemes and new business opportunities with over-the-top service providers. These different architectural approaches and the overall framework need to be thoroughly studied and benchmarked in terms of techno-economic and end-to-end functionality through a number of research initiatives.

4.3.6 Access-Core Integration

A wide variety of virtualized and commodity hardware, a technology and ownership agnostic network control plane as well as the software-defined functionality and the programmability it offers, are the key aspects of a URN which are changing the Access landscape. However, a parameter that is often ignored is the necessity to provision solutions that are equally applied to both urban and rural areas. This architectural step will complete the mosaic of chances allowing the seamless integration of Access and Core. The penetration of a converged optical access network directly to the Core network will lead to a considerable network architectural simplification. To achieve this there are many competing solutions to which the availability of a flexible Optical Distribution Network and its use for an extended reach is the common denominator. Considerable research effort is still needed, at both the architectural and system levels, to fully address the topic.

4.3.7 Networking Challenges in the Global Cloud Infrastructure

The development of Cloud infrastructures is perhaps the single most prominent technology driver of our time. The Cloud is the offspring of the convergence between networks and datacentres and as networks become programmable, dynamic cloud networking will become the fastest growing segment of the overall network connectivity market. The convergence between networks and datacentres is rooted in two processes:

- The need of telecommunication operators to reduce their operational cost as well as the response times of their networks. Another strong motive was the need to create global alliances, which require the homogenous framework a standard, scalable and manageable –through commonly adopted orchestration procedures-technology platform is offering.
- The increasingly higher role of datacentres in providing Cloud services in which the datacentre industry saw a new source of revenue.

It is worthy to elaborate on this process: as analysed in section 4.3.2, the pervasive end-user device connectivity in Access is implemented today by means of an overlay of independent network infrastructures. Under the existing paradigm, service convergence is implemented in Metro, something that is increasing the number of endpoints/terminations in that segment while it is changing the connectivity pattern within it. Moreover, to increase the quality of offered services to the end-user, it makes better sense to implement functions like content caching at these termination points than carrying traffic back and forth in longer, indirect, routes to a single host location (e.g. to a remote, centralized datacentre). As such, there is a necessity for distributing processing and storage facilities in the Metro segment, preferably in places collocating networking services. However, it would be difficult to augment the functionality of purpose-built equipment, something that was seen as an obstacle in using infrastructures as a commodity in global market exchanges (Infrastructure-as-a-Service, IaaS). Moreover, purpose-built equipment is also a roadblock to the rapid provisioning of new services and its upgrade is costly. To overcome all these, NFV i.e. the use of generic hardware with software-controlled functionality under a modularized system architecture, is allowing for functionality customization based on low-cost platforms.

So, at a time where datacentre facilities are distributed in Metro and progressively closer to the end-user, the Metro and/or Edge nodes of network operators are gradually deploying general-purpose equipment like datacentres do. This convergence is changing the way Metro and Aggregation networks are architected. Last, but not least, the intra-Datacentre optical DCIs are rapidly exploiting core network technologies (e.g. transceivers and switches) to implement the necessary connectivity, and ongoing research is aiming to demonstrate cost-efficient and low power consumption platforms for this purpose.

Evidently, we are witnessing a "fusion" where both Telecom and Datacentres vendors strive, in parallel, to develop common technology platforms, even at the chip level, and common software-upgradeable control and management systems facilitating dynamic resource slicing and management policies. The resources to be virtualised and managed in a dynamic and flexible way include but are not limited to, storage systems, servers, switches and network bandwidth in a process that will extend the benefits of virtualization across the entire network, i.e. regardless of geographic location.

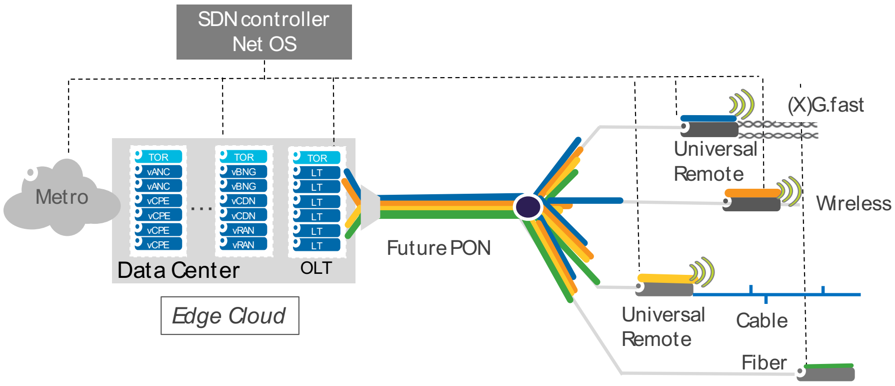

Figure 4 - A converged access network exploiting virtualization [source Nokia]

Conclusively, the convergence of Datacentre and Network infrastructures is evident in all network segments and it applies to all systems deployed therein:

- In the Access through the virtualization of CO and/or the MFH/MMH functions in the Edge cloud, facilitating a flexible, converged infrastructure as shown in Figure 4.
- In the Metro through the distribution of Datacentres in public networks and the evolution of Metro nodes to acquire an (intra-node) architecture that has common features with a Datacentre. We are progressing towards a point where Cloud service provider and Network provider Datacentres will become the natural exchange hubs of this market, as they represent the source and destination for much of the demand. The next step will be their complete integration.
- In the Core, where the interconnection of large Datacentre hubs, as well as of the centralized Web-scale data centres, will require orders of magnitude higher scaling of optical networks.

As of today, the convergence between Telecom networks and Datacentres is progressing particularly fast in the Aggregation segment: two important initiatives supporting this trend are CORD (Central-Office Reimagined as a Datacentre) from AT&T and FASA (Flexible Access System Architecture) from NTT that are both targeting the Edge aggregation network.

To implement the full integration of Datacentres and Network infrastructures, which would be a vital step for the proliferation of the Knowledge-based economy, research effort is necessary to progress the current state-of-the-art in a number of informatics and engineering fields. Priority should be given to the following topics:

- Ultra-high capacity optical core networks
- A dynamic adaptive, flexible optical metro network
- Novel node architectures exploiting virtualization by design
- End-to-end automated networks exploiting unified control/management-plane and monitoring platforms integrating cloud resources.

Achieving the seemingly “infinite” scale necessary for this new *digital everything* network, without incurring a CapEx/OpEx explosion, will require a new automated network fabric. This will require fusing (distributed) data centre infrastructures with massively scalable and flexible optical interconnection infrastructures forming up a single adaptive continuum under a new, open, network OS control. As a matter of fact, a longer term perspective of this convergence is that future networks may not necessarily require special purpose vendors: optical equipment that will be fully standardised and fully open (in the form of optical whiteboxes) may become the norm, so when empowered with a vendor-agnostic OS they would bring about *commoditisation* in optics.

4.3.8 Ultra-High Capacity Core

Core optical transport underpins the world's networks. Moving large amounts of data around countries, continents and globally is what optical transport has traditionally been relied upon to provide. Nothing else can provide transport over many kilometres for the enormous and exponentially growing data rates being seen in our networks. The enormous bandwidth demand is still the main driver for Core network evolution and we have seen in section 4.3.2 that MBH and the fixed-mobile convergence is scaling up this demand which should come with a drop in cost by an order of magnitude. Such a bold target will only be possible with innovation at all levels: from components up to systems and networks.

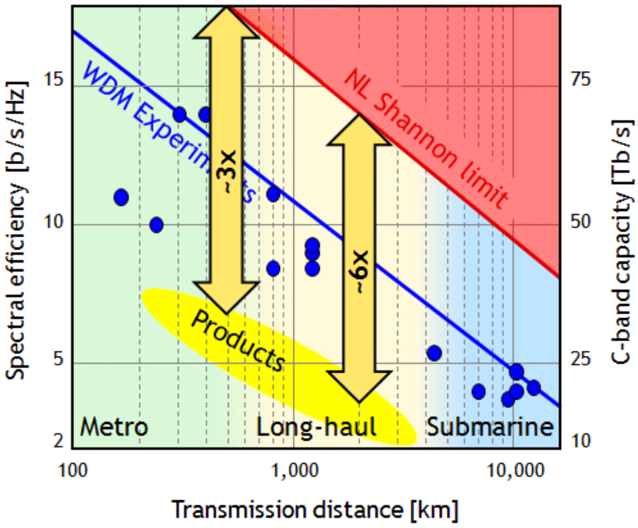

Figure 5 - Optical Systems Approaching the Non-Linear Shannon Limit [source: Nokia]

In current optical DWDM systems, the physical dimensions of time, polarization of light, amplitude and phase (quadrature) are already employed. Through these systems the maximum information capacity of fibre − the so-called 'Nonlinear Shannon Limit' – is being reached as shown in Figure 5.

Achieving further gains of 10x will therefore require leveraging the two remaining physical dimensions not yet exhausted - frequency and space. Cost effective expansion of the frequency (wavelength) space will be enabled by ultra-wideband optical amplifiers, transceivers and wavelength steering components, allowing utilization of DWDM operation in the more challenging longer and shorter wavelength "L" and "S" bands of the low-fibre-loss window, adjacent to the commonly used "C band" (Figure 6). The resulting system capacity scales directly with this increase in (optical) spectrum usage, resulting in a 3-5x capacity growth potential.

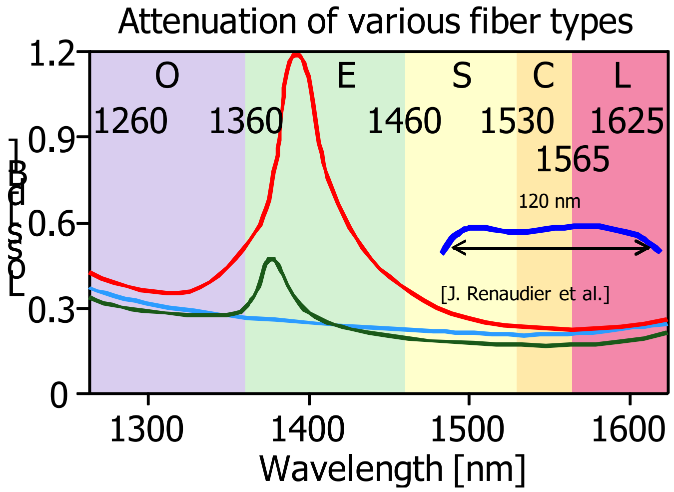

Figure 6 - Extending DWDM System Bandwidth from C to C+S+L Wavelength Bands [source: Nokia]

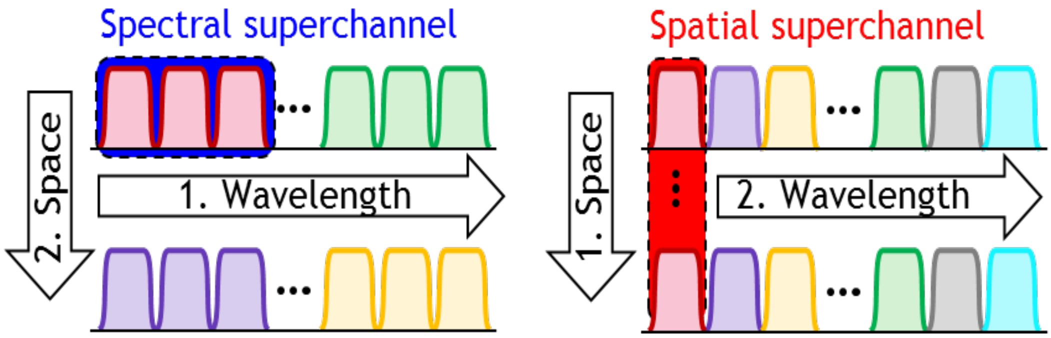

Figure 7 - Spectral and Spatial Superchannels [source: Nokia]

However, for each of the available wavelength channels, it is still important to maximize the transmission rate but we are almost at the limit of the digital signal processing speeds required for sophisticated higher-order modulation schemes. Therefore, alternative approaches will need to be employed, using lower-order modulation and more efficient usage of the available spectrum so, through WDM, to progress in the Tbps rates. Alternatively, the use of multiple spatial paths in parallel, as Space Division Multiplexing (SDM) technologies do, allow exploiting bundles of SSMF and, potentially in the future, multi-core or multi-mode fibres. Both approaches are forming up *Superchannels* and the two primary types of Superchannels are schematically shown in Figure 7. To continue the research effort in exploiting both schemes is really important for ensuring that demand will not surpass capacity supply in Clouds, so regional Datacentres can be seamlessly interconnected to regional ones for sustainable Cloud services.

4.3.9 Dynamic, Adaptive and Flexible Metro

The convergence is changing the way optical Metro networks are architected and operated. Metro networks, which are interconnecting edge clouds, access networks and enterprises, must be optimized to handle a completely different traffic profile with much larger variations in the volume and the connectivity requirements which now have to be implemented over different time scales. Indeed, *diverse* and *time dependent* 5G services need to be transported at different time scales – many of these services require enormous bandwidth – bandwidth too large for the IP layer to handle all of the dynamics: routers will not cost effectively scale to deal with, for example, future media networks carrying vast, multi camera 4k flows of up to 200 Gb/s.

The advances in the Access and, in particular, the different alternatives for service convergence (see section 4.3.2) will have a different impact on the way a Metro network is architected. Specifically, if service convergence is implemented in Metro then the number of termination points at Metro nodes will increase while the incoming traffic flows will have dissimilar transportation requirements in terms of capacity, granularity, QoS performance etc. On the other hand, architectures implementing service convergence in the Aggregation or Access will lead to a consolidation of the Metro nodes. In either case, an open issue is the technology to be used for protocol adaptation, encapsulation/mapping and (de)aggregation of traffic flows with dissimilar characteristics and attention should be paid to the identification of the interfaces. Regarding Metro transport, the connectivity pattern will become meshed between Metro nodes due to the centralization of IP edge functionalities in virtual server farms hosted in Datacentres that will be collocated with these nodes. Eventually, this collocation will lead to integrated hubs providing ubiquitous computation and storage facilities, abundant capacity and carrier-grade transportation performance.

Evidently, *diversity* is the word that will better define Metro. In response to this, new elastic networking technologies are necessary combining the flexibility of packet technologies with the scalability and the effectiveness of optical transportation. In fact, a combination of these two assets will allow supporting the on-demand dynamic reconfiguration of programmable network functions and seamless resource slicing. Advances in data-plane architectures will allow the handling of a wide range of granularities, from the sub-wavelength to the Superchannel, providing at the same time scalability and means to share the available resources through statistical multiplexing functions. Unless these challenges are unravelled and fresh answers are given, Metro may bottleneck the seamless Cloud evolution.

The following topics are important research priorities for the emerging Metro segment:

- The exploitation of novel encapsulation/framing and transportation schemes along the direction of an ‘optical-Ethernet’ which will allow inspecting, aggregating and forwarding flows efficiently based on simple forwarding rules. The mechanism should provide means to groom flows with dissimilar QoS and Class-of-Service (CoS) characteristics, to provide traffic engineering functions like statistical multiplexing and, yet, to be flexible and amenable to rapid reconfiguration.
- The exploitation of elastic and pluggable optical interfaces suitable for Metro and their integration into a dynamic transportation network platform, with an emphasis on dynamic data-plane solutions. As an example of such interfaces, we consider the case of intelligent and auto-tuned “*optical modems*”. These will autonomously adjust their operational parameters (capacity, carrier technology, modulation

format, optical bandwidth etc) in an optimal way to allow adaption to the specific path/channel conditions at each time. Such SDN-enabled adaption will allow operational optimisation as paths dynamically change or as the load is changing due to changes of SLAs or due to restoration procedures etc. Moreover, it will set the necessary framework for slicing transportation resources as an important part of their *commoditization*.

4.3.10 Novel Node Architectures Exploiting Virtualization by Design

The process of Datacentre and Network infrastructure convergence will have a deep impact in the way Core and Metro nodes are architected and operated while it will set new performance goals. It is evident that if networks are to play their role in a dynamic Cloud marketplace, the time window they operate in should shrink dramatically compared to what is the norm in contemporary networks. Moreover, the ability to transport data regardless of distance, i.e. across different domains, postulates that the nodes are implemented making use of common and interoperable technology platforms.

These two goals will require significant advances in both the data-plane and control-plane. Regarding the latter, the main building block of future Core and Metro nodes will be “white boxes” that are making the most out of advances in SDN and NFV. Regarding the former, a *fast tuning* of the entire interconnection fabric would be necessary and although the reconfiguration speed depends on a large number of parameters, data-plane technologies that may exploit a faster dynamic reconfiguration in a reliable manner are in the centre of attention. Profoundly, future Cloud nodes will consist of an “optical” and an “electronic” section and each are expected to face new challenges compared to the previous era. Specifically, during the first phase of convergence, Datacentres and Core/Metro nodes will be collocated so the traffic forwarding mechanisms between them should be standardized to allow a seamless traffic flows exchange. Progressively, as the two entities are integrated, the intra-node network may show architectural similarities and may face similar orchestration challenges with those witnessed in the inter-node (Cloud) network.

Regarding the early convergence stage, research on node architectures should address the following challenges regarding:

- *Flexibility,* with respect to bandwidth granularity (from Gbit/s, to Tbit/s and into the Pbit/s regime) and transportation parameters (e.g. elastic optical bandwidth, modulation format, etc). Moreover, flexibility mechanisms incorporating fine optical granularity bandwidth chunks are expected to provide optical grooming functionality to alleviate the scalability problems of their electronic counterparts. Specifically, fine optical granularity bandwidth chunks could be exploited in both *frequency* domain (with sub-GHz spectral resolution that can be transparently interfaced to 100 GHz optical channels) and *time* domain (for sub-wavelength TDMA sharing in the form of a slotted operation).
- *Scalability,* in terms of capacity (throughput into the Pbit/s regime) including multi-dimensional (time, space, frequency) capability. Switching complexity increases

dramatically with multi-dimensional operation and contemporary ROADM solutions cannot simply scale-up to support the requested switching functions.
- *Programmability,* in the sense of white boxes that would orchestrate the converged nodes of the multi-dimensional, multi-layer, multi-vendor and multi-technology interoperable transportation networks. These white boxes will be an integral part of node architectures that are vastly relying on commodity hardware for computation and storage and proprietary, but interoperable and SDN-enabled, hardware for optical signal generation and routing. As elaborated in the previous section, the SDN-enabled operation is the necessary framework for slicing transportation resources as an important part of their *commoditization.*

Regarding the final convergence stage, the process of transforming a Datacentre to a general purpose, ubiquitous Cloud node will set new challenges to the intra-Datacentre Interconnection (DCI) fabric. Here we are just highlighting some of them:
- *Architectural challenges* which include the identification of the functional relation between the different building blocks and the technologies that are necessary to process and forward both individual flows and aggregated chunks, scaling up to mulit-Tb/s, even Pb/s, capacity as analysed above.
- *DCI flow forwarding challenges*, which will take into account that optical technology and transparent forwarding will be implemented to reduce the level of electronic switching complexity. Concepts and technologies developed for a dynamic data-plane networking would find a new field of application here, including technologies for resource allocation algorithms for shared/collision-free access of the optical medium, scheduled switch operation, simplified labeling mechanisms, wavelength routing, etc.

4.3.11 The End-to-End Cloud Automation

As has been elaborated in the previous sections, networks –as an integral part of Clouds- will become the central linking institution of the Knowledge-based economy. Global and regional enterprises are already making plans and they are creating the necessary alliances to allow them to operate both locally and globally (see section 4.3.1). However, such a seamless integration is a rather longer term goal since it requires the creation of certain preconditions before it becomes effective.

On a local scale i.e. at the node level, a leap in the overall efficiency fostering Cloud operations requires to jointly orchestrate networking, storage and computation resources. In this direction a deeper integration between the different physical infrastructures and the control and management layers will enable a more flexible, agile and programmable way to orchestrate the future nodes where Datacentres and Network facilities are either collocated or are integrated. The innovation brought by SDN has two main advantages: first, SDN-based control will help to harness the capabilities of optical technology and allow them to serve as an integral part of a flexible multi-layer network services infrastructure; second, exposing advanced northbound abstractions over the SDN layer will allow service provisioning while providing a platform to jointly orchestrate network resources and Datacentre resources. Through such virtualization, converged but isolated virtual infrastructures will emerge.

Therefore, the major operators have their own, fully functional but still isolated, network infrastructures across all network segments (Access-Metro-Core). Quite often, dissimilar technology and control paradigms are implemented in each segment and different evolutionary plans are made to each of them without considering the global implications. The unification of these management systems, to support seamless transparent end-to-end connectivity across regions/segments, is one of the advances SDN and NFV are promising to achieve. In this respect:

- Progress with respect to SDN is more tangible e.g. it is easier to comprehend that in a multi-technology network, controller federation is the winning approach for setting the regions/segments within a single operator’s domain under a single framework. Although this is a considerable step forward, it is not sufficient to build a ubiquitous ecosystem across different domains. So, what is currently pursued is *multi-tenancy* over network infrastructures which allow an operator to be the real infrastructure provider in one region/segment and to become a Virtual Network Operator (VNO) in another.
- Real multi-tenancy over physical infrastructures is made possible only though the extensive use of Virtualization/NFV. In fact, virtualization should be extended over the entire system of heterogeneous physical resources comprising the Cloud: storage nodes, computation servers, transmission subsystems, switching ports, aggregation levels and points, etc. Moreover, it should be extended, uninterruptibly, across all network segments: from the Core down to Metro, to the different Access networks and –at the final stage- to the appliances/devices themselves. A federation of VNOs is unthinkable without federating the physical infrastructures across all entities and domains. This federation will give to VNOs the flexibility to consider the virtualized infrastructure as a single entity end-to-end and it will allow them to optimally use network resources for providing the intended service.

A research challenge here is to identify and implement the processes and the mechanisms delegating virtualized resources to different stakeholders (tenants). This process is tag-named as resource partitioning and slicing. A multi-tenant, multi-technology, multi-domain management system should be able to create end-to-end partitions of resources. The isolation between the different resources, especially over the shared physical infrastructure, is an important issue to address since it must provide guarantees over capacity, admission, security, robustness and QoS performance. Therefore, multi-tenant management systems should be able to provide distinct virtual infrastructures and secure access mechanisms with configurable functionalities tailored to each tenant over a multi-domain network environment, constituted by owned and rented infrastructures. This process is further complicated by the various architectural choices made by operators. For example, for various reasons, different operators may implement different service convergence schemes for fixed and mobile Cloud traffic (see section 4.3.2). In this case, multi-tenancy for offering end-to-end services across multiple VNO infrastructures that implemented different aggregation stacks and/or technologies is a problem not yet addressed.

Another research challenge is to identify a clear framework for SDN controller federation and its interoperability with NFV in Core and Metro networks. Finally, cross-domain orchestration considering security and resource isolation issues has a long way to cover with early ideas but no concrete developments yet.

Finally, as the convergence will eventually include devices and any types of machines with AI, application-aware networking would be a feature delegating a higher level of control to the end-user. Application-aware networking postulates that control and management planes have evolved to the level where specific application parameters are interacting with the specifics of a multi-layer network model to make the most of the underlying infrastructure. This means to make decisions e.g. to either optically bypass nodes or to forward flows to grooming points with given level of processing, storage and transmission resources, based on service characteristics in terms of capacity, latency, jitter etc. Research objectives may include the creation of a common language for the applications to communicate with network abstractions, optimization algorithms etc.

Last, but not least, in the framework of Cloud orchestration and application convergence towards a fully flexible and programmable network, is that a form of cognition and/or self-organization should be introduced to improve automation. The dynamic network configuration at short response times requires introducing higher level of intelligence in Cloud orchestration tools and multi-layer monitoring information can be used as a feedback for the cognitive SDN control plane, and lead to better configurations of the entire network.

4.3.12 Augmented Intelligence and Monitoring for Network Services

Monitoring is essential for network operators, since it allows them to evaluate network performance. A fully flexible and programmable Cloud networking based on the aforementioned technologies requires complex tools for network analytics, network troubleshooting and optimization.

Future Core and Metro networks will consist of

- General-purpose data processing systems, that would be operated and upgraded exclusively based on software.
- Specific-purpose systems like for example, transmission modules.

In this framework, the monitoring techniques are aiming to address both network operation challenges as well as service integrity challenges. Regarding network operations, the monitoring function allows collecting data related to the quality of the transmission and triggering reconfiguration in case of some quality degradation. In a SDN-based control plane, the monitoring functions will be realized by programmable monitoring technologies enabling cross-layer troubleshooting, further improving network stability. Capitalizing on this by monitoring traffic samples, data analysis might be used to predict traffic matrices for the near future, a knowledge that can be afterwards applied to reconfigure and re-optimise network resources, detect anomalies, anticipate failures, etc. In other words, it allows to use statistics to ease the introduction of an automated decision making process. Moreover, real time monitoring information can be used to dynamically reconfigure data-plane parameters to cope with signal degradation at specific (sub)-wavelengths, activating the appropriate resiliency mechanisms.

In the next generation networks, data will be collected from a large number of network nodes and data analytics would be mainly based on packet-level monitoring without excluding the potential of collecting data directly from the optical layer or even to correlate data from both layers. Data analytics and traffic monitoring requires powerful architectures and algorithms, in analogy to the methods used in Big Data. The available data will be collected from a large number of geographically dispersed nodes but latency related restrictions will prohibit the transportation of these data to a centralized repository placed for analysis. As a result, a distributed intelligent platform would be needed to manage the geographically dispersed networking resources. Overall, a combination of distributed and hierarchical architectures need to be implemented for efficient network monitoring and real-time analytics provisioning which would play an important role to the creation of an automated, dynamically reconfigurable and adaptive Cloud infrastructure.

Regarding service integrity, in the framework of VNF, the telecom services would be decomposed into multiple software-based components running on general purpose hardware devices. However, given that these services would run on system dispersed in different locations, service parameters could differ considerably depending on the type of the Network Function, the virtualization technology, the applied hardware acceleration or network stack optimization.

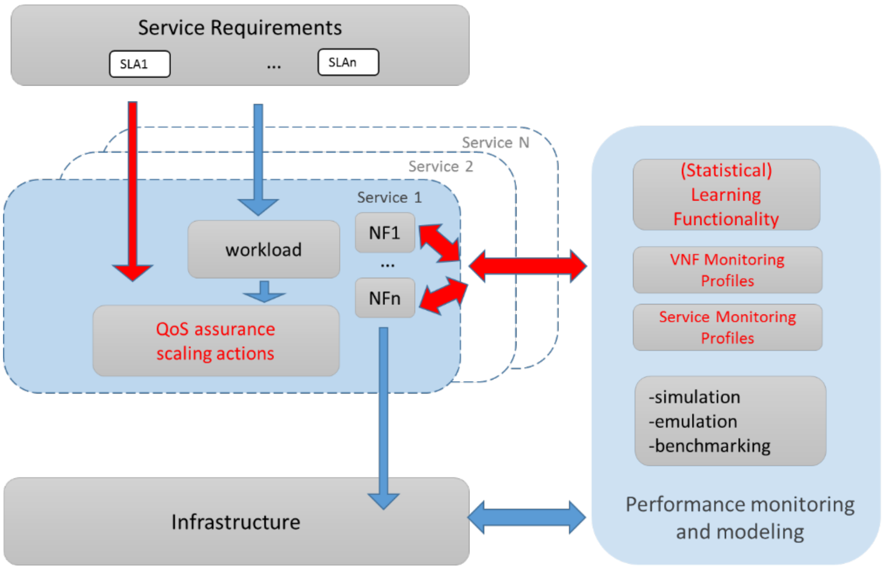

Figure 8 - Operations related to service monitoring [source: iminds]

To this end, the network operation, infrastructure-focused monitoring techniques and platforms discussed above should be extended to support the evaluation of the impact of different virtualization technologies and network processing acceleration options.

Parameters linked to SLAs, QoS or even KPIs might be used as elasticity monitoring indicators allowing the control mechanisms to continuously adapt hardware resources to the workload as shown in Figure 8. The research objectives would include the design of mechanisms to model and predict performance of combinations of NFs under a range of different deployment scenarios. Profiling VNF's can be done in a white-box approach to derive performance parameters by borrowing methods and techniques from statistical/machine learning. This will allow deriving new methods for service decomposition to a large amount of nodes, enhancing the elastic properties of any network service and, thus, to use the resources with higher efficiency. By investigating the trade-offs between these scaling actions, algorithms would be distilled, choosing the most efficient function of the required QoS parameters. Actually, through this process there will be an effective translation of SLA/QoS parameters (like end-to-end latency etc) to infrastructure resource parameters like processing, storage or network transportation units (CPU, memory, and bandwidth).

4.3.13 Hybrid Visible Light Communications and Radio Networks

In addition to millimeter-wave small-cells, the use of free-space optical channels has been proposed for high-bandwidth wireless data communications, especially in the form of visible-light communications (VLC). VLC approaches using conventional LED lighting equipment that can be bought in hardware stores used in conjunction with relatively low-cost and tiny photo-sensors have been shown to be feasible for transporting high-bandwidth spectrally-efficient OFDM waveforms, currently up to 1.6 Gb/s with a single-color LED. One of the interesting consequences of using VLC is that the deployment of communications infrastructure is coupled with lighting, and, moreover, good lighting coverage will ensure good communication links. This could even be true in the case of non-line-of-sight scenarios since modern indoor environments are conceived with materials that guarantee good lighting conditions even with reflections. Exploitation of this symbiosis will be extremely beneficial in locations such as airports, shopping centres, office buildings, super-markets, etc., which are all examples of locations where high-throughput wireless communications are expected with even todays smartphones. Another clear benefit of the symbiosis is energy consumption, in the sense that the communications component does not require additional power for transmission since it exploits the high-power transmission used for lighting.

Similarly to radio-based systems, multi-port transmission (i.e. MIMO and/or adaptive beamforming) is feasible with LED lamps, which are usually an aggregation of many low-power emitters. Innovative multiport processing for LED-based transmission is still in its infancy. In particular, the benefits of MIMO transmission, whether point-to-point or multiuser-MIMO, are yet to be determined. This is primarily due to insufficient understanding of the propagation characteristics of free-space optical links. Coordinated multipoint transmission (CoMP) may be simpler with VLC transmitters because of the number of lighting sources that are available in typical indoor environments. Finally, through CoMP, VLC transmission could provide very accurate positioning possibilities.

This is due to the fact that multiple optical channels with significant carrier-spacing (orders of GHz) can be used to allow terminals to position themselves with extremely high accuracy. Industry cooperates to promote the use of VLC through the so-called LiFi or light fidelity consortium. LiFi approaches the standardization of VLC waveforms and protocols through the now out-of-date IEEE 802.15.7 standard proposal. The pitfalls in this approach are firstly that the physical-layer waveform does not consider and is not particularly well-adapted to high-spectral efficiency techniques (OFDM, MIMO, beam-forming etc.) and secondly that the proposed protocols are not conceived for tight (low-layer) heterogeneity with radio-systems. These are both extremely unfortunate with respect to the massive future usage of VLC technology. In the context of low-power terminals, like smartphones, VLC can make maximal impact if used for high-throughput downlink communications, either in the form of multicast/broadcast radio bearers or unicast bearers whose uplink counterpart (including necessary low-layer signalling for channel state information and channel decoding integrity) is transported using legacy radio services. It could be argued that a tight interaction between radio and optical components should be considered at the level of baseband processing as well. Since OFDM transmission is feasible on a free-space optical link, it is definitely worth considering using the same basic waveform and protocol stack for radio and VLC components. This would allow for a common baseband processing platform in both the small-cell transmitters and terminal receivers. Moreover, current access-layer protocols are perfectly adapted to the use of Downlink-only component carriers. The 40 MHz of so-called Supplemental Downlink Channels (1452-1492 MHz) will be used widely in Europe in conjunction with bi-directional carriers in the coming years. Adding VLC-based downlink channels could prove to be a cheap solution for boosting downlink bandwidth.

Research areas for VLC:

- Propagation measurement campaigns for the VLC component for: co-located multi-port transmission achieving point-to-point and multiuser spatial-multiplexing with a single LED array; highly-selective adaptive beamforming techniques using feedback on the radio-link; distributed multi-port transmission using multiple spatially distributed LED arrays.
- Proper statistical modelling of the free-space optical channel for assessment of fundamental communication limits.
- MIMO techniques for point to point and multi-user scenarios, with a focus on LED arrays.
- Baseband signal processing units for VLC and radio transceivers.
- Analysis of the capacity of the free-space optical link (even in the single-user case).
- Investigation of innovative radio resource management techniques that exploits colour multiplexing.

4.4 Experimentation with Verticals

4.4.1 Introduction

a) Inter-disciplinary Research for the Future Internet

The Internet, considered as a network infrastructure, has facilitated an enormous amount of contributions improving citizen’s welfare with a fascinating list of innovations and boosting economic growth. Nowadays we cannot decouple the infrastructure from the services and business anymore. In the future, relationship among technologies, applications, services, social impact, privacy models, business models, pricing models, the concept of IPR, security models, management and deployment of services and regulation should develop in a coherent way.

The appearance of new agents and technologies has been, and it is being, made at the cost of a large amount of tussles that challenge the established rules, regulations and law enforcement. As example, we may think about network neutrality and data privacy. At the very beginning, the tussles were mainly about technical and economic interests of the different parties. Nowadays, these tussles affect the individual citizens and the society as a whole.

Policy makers need the support of many different experts who give their expert advice from their point of view. Policy makers and regulatory bodies need a “unified” advice that may consider all of these aspects altogether.

We need the integrated capacity of creating not only new technology for the Future Internet but new applications and services too. As an example, we may consider network neutrality and data privacy debates. Both require a broad knowledge and understanding of different aspects. Most of the debates are biased towards some technical aspects while other focus on regulatory aspects or economic models or particular interests.

Being able to address these debates globally requires interdisciplinary groups of researchers. Teams including engineers (communications, electrical and software engineers), economists and specialists in digital economy, sociologists, lawyers, bio-engineers, user’s associations and professional associations may work together and be able to understand better the whole set of implications before defining or deploying a new service. Training of postdoctoral fellows with this interdisciplinary collaborative work is necessary.

Projects should not only provide large-scale experiments to proof new advances but also have a task devoted to this interdisciplinary approach thinking in the long term. Training teams of future thought leaders on multi-disciplinary aspects of the Internet, including engineering and economic aspects but also social, legal and regulatory aspects is an investment in talent that will help filling a gap in human resources in Europe. These qualified researchers and innovators will be able to direct future Internet and Internet services development in industry, or wisely advise regulators and policy makers.

b) Experimentation for the Future Internet, and Verticals

The testing and validation of research results, implemented solutions, products and services, in large scale real life experimental infrastructures, is essential for the innovation process of the future Internet. This is because: a) it enables the efficient testing of products, application and services; b) it minimises the time-to-market; c) it provides to small and medium-size developers and innovators, who cannot afford testbeds or testing equipment, sophisticated testing tools and facilities.

Current available large scale experimentation facilities (under the umbrella of FIRE, FIRE+ and other initiatives) provide the means for the efficient testing of innovative solutions. However, the introduction of the 5G ecosystem with its native ability to support new innovative vertical solutions and applications in the area of e.g. automotive, media delivery and e-health, characterised by special requirements, impose new challenges, because the currently available experimentation facilities which provide general purpose infrastructures and tools lack to capture the idiosyncrasies of the aforementioned vertical systems and applications. On the other hand, in such sectors, the preparation and building of monolithic, specific-purpose laboratories is time-consuming, costly, and in many cases unsatisfactory to the requirements on scalability, physical distribution, mobility, and ubiquity. Therefore, the use of experimentation facilities with the hybrid target of supporting both general purpose research experiments and experiments fine-tuned to vertical sector needs, seems to be the most viable and cost efficient solution.

Current available experimentation facilities should be extended in order to provide enhanced experimentation infrastructures on top of which third party experimenters, e.g. SMEs, or any digital asset owner over vertical sectors such as automotive, media delivery and e-health will have the opportunity to test vertical applications and solutions in an integrated, cooperative and fully featured infrastructure fine-tuned to the characteristics of each vertical sector.

The support of verticals in experimentation facilities will attract much wider range of stakeholders compared to existing general purpose or dedicated vertical solutions, ranging from SMEs and industrial partners developing products and services over vertical sectors to public bodies (e.g. municipalities) and organisations (e.g. automotive safety organizations). The vertical enabled experimentation facilities will be beneficial especially for vertical SMEs which have not the means to develop dedicated testbeds, as they will have the opportunity to test the developed solutions in a real life 5G environment under different meaningful vertical-specific configurations, while they will be capable of evaluating their products to vertical related KPIs. Therefore, on this open and large-scale 5G vertical experimentation infrastructure, innovative service providers will reinforce their businesses, while start-ups and SMEs will also find a large variety of new opportunities

Phase 1 and 2 of the process set up by the European Commission for the development of R&D activities focused towards 5G provide emphasis to the verticals that require new air interfaces, novel network paradigms, and set challenging requirements that will be met through the 5th Generation of mobile radio networks. At the same time, they pave the way for a European-wide experimentation to demonstrate what technology enablers will meet the above requirements.

It is widely perceived that not all verticals of relevance to the society are covered by the activities performed in the context of 5G, and that experimentation of verticals will be a fundamental step for the development of beyond-5G networks.

The rest of this Section discusses some verticals that should be addressed to a larger extent in future projects, some key challenges that will have to face and some aspects underlying the relevance of an inter-disciplinary approach to experimentation of future technologies.

4.4.2 Public Protection and Disaster Relief (PPDR)

Communication state-of-the-art for PPDR is currently still limited to narrowband voice technologies such as TETRA, TETRAPOL etc. Data capability is currently limited to short data services and narrowband packet data (kbps). PPDR communications suffers a small economy of scale compared to consumer mobile, therefore limiting technological advancement, and yielding high cost equipment and services in a relatively low budget domain. Industry has foreseen the opportunity to transition PPDR communications services to utilise developing 4G technologies. The key motivation being to economically safeguard costs of devices and shared infrastructure. 3GPP standardisation activity is already in progress towards incorporating features for critical communications and PPDR into releases 13 and 14.

Requirements for PPDR include critical features such as:
- available resources needed to allow for highest resilience in period of stress;
- high dynamics, redundancy, coverage etc;
- effective sharing with consumers;
- appropriate QoS to accommodate the interests of both consumers and professionals;
- geographical coverage, not just by population;
- associated sharing of, or allocation of, dedicated spectrum;
- appropriate secure cross-border operation, as roaming between different EU states brings complexity associated with the variation of secure sensitivities, compound by the increasing dynamics of 5G networks and infrastructure sharing.

The European Commission are currently funding procurement related activities through the H2020 Secure Societies program. A Pre-Commercial Procurement (PCP) phase aims to implement the first critical 4G networks in the period 2018 - 2022. Timeframes for 5G development somewhat overlap the formation of a new market for critical broadband, which is expected to be catalysed by the PCP activity. It is therefore envisaged that after 2020 experimentation on a large scale of networks for PPDR will be required.

4.4.3 Mission Critical Applications

Mission critical applications include PPDR, but extend to further areas, like e.g. the control of industrial plants under challenging situations, train control in railway systems, or the patrolling of coasts.

Current mission critical communications are based on specific technologies, which must guarantee in any situation the basic services for exchanging relevant information among the operators. These basic services are called “mission critical services” and today are limited to voice and messages in various formats. Typical mission critical functionalities provided by these technologies (like TETRA for PPDR) include communications when no connection between the BS and the Core Network is available, direct mode connection among terminals, security, priority, preemption, low latencies, and many others. Operational applications are evolving towards large usage of data/video, deployment of interconnected devices, interoperability among different organizations in case of emergencies.

New technologies such as LTE are able to guarantee most of the basic technical performance required in mission critical applications, while others can be provided by a proper network deployment. 5G systems will provide heterogeneous networks in which new radio access technologies are implemented and coexist with those already available. Any specific service would thus be provided by means of simultaneous availability of different radio access technologies, terrestrial and satellite, with the most suitable being used for the current application, service and scenario. This will increase network resilience, guarantee graceful degradation in case of failures, simplify interoperability issues among different organizations.

The concept of Network Slicing will be at the basis of mission critical applications, and seamless usage of the various available bearers and networks and immediate switching of traffic from one to the other are of paramount importance. Other relevant issues are the characterization of radio access solutions and air interfaces that can be provided by 5G networks, network functions virtualization aimed at simplifying rapid deployments, security levels guaranteed independently from the type of access being used as well as dynamic identity and access management according to the specific operation.

In addition to the solution of some outstanding technical issues, the application of 5G architectures to mission critical systems needs to be proven under all operational requirements, from deployment to basic and emergency operations. Mission and non-mission critical services need to be assessed also looking at future evolutions, and reference platforms for integration and testing of the proposed solutions shall be defined. This calls for an experimental set-up and activities performed on a large scale to test new technologies provided in the context of 5G development with respect to the requirements of critical mission applications.

4.4.4 Testbeds for Safety Critical Verticals

The previous Sections emphasized the relevance of PPDR and mission critical experimental testbeds. They should involve a number of network nodes, and most of them would belong to the category of Ultra Reliable Machine Type Communication (uMTC) devices. They will demand reliability of at least 99.999% w.r.t. on-time packet delivery. Measuring through experimental activities such reliability is actually a great problem. The methodology applied must be capable to cope with correlated packet delays as typically encountered in networks due to buffer occupancies and fading channels. Once established, a likely result will be that experiments must run for very- or even infeasibly long time to achieve confident results as shown in Figure 9.

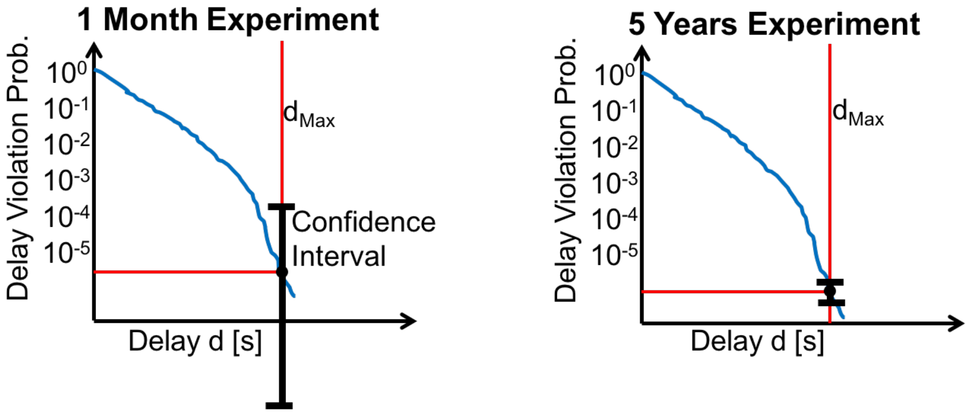

Figure 9 – Confidence interval vs duration of the experiment

The same problem is known in civil and mechanical engineering. Well-established methods exist to decrease experiment time without reducing result quality. The best known ones are climate chambers and shaker tables emulating the stress of an entire component lifecycle in much shorter time. Of course those methods cannot be applied in telecommunications since temperature and shocks are not in the focus of research when evaluating causes for packet delays and losses. Still equivalent methods must be found applicable for testbeds. Possible fields to search for solution are rare event simulation and importance sampling in empirical studies. All results and the context they were collected in, should be systematically saved and permanently evaluated against real life experiences once a system is broadly deployed. This is then used to permanently refine methods and models for network reliability evaluation.

Without well accepted methods to evaluate safety critical communication systems it is likely that many prototypically proven applications in verticals such as healthcare, energy supply and vehicular communication will never or only after very long time enter the market.

4.4.5 Industry 4.0 and 5G/Beyond5G Networks

Industry 4.0 introduces a number of technologically challenging requirements, targeting low latency and high reliable transmission to enable cooperative action of group of robots, production entities, cars and drones. Interworking of public (Operator) and enterprise (Industry 4.0, OTT) based solutions is also a challenge. Coexistence of end-to-end connections and radio links with extremely different requirements will be optimized and tested in this context. Finally, the target is to work on solutions for communication links to enable cooperative decision making. There are many similarities from the radio point of view between cooperative robots in industry 4.0, automatic driven cars and cooperative clusters of drones. Just few parameters are different such as speed and number of sensors/actuators embedded on the platforms.

Industry 4.0 wants get more flexibility to reconfigure the production processes on demand. The current solutions are supporting reconfiguration of modular subsystems. Each of these modular subsystems is equipped with manipulators, actuators and sensors but also with an integrated control system. Leading companies in the area of industry 4.0 are structuring the future in several steps in Plug & Produce Technology, Smart

Connected Equipment, and Smart Human Machine Interface. Topics in this context are Augmented Reality, Human Robot Collaboration, Cyber Physical Systems and moving the intelligence of the control functions to the cloud.

Research and Prototyping targets how to link robots, actuators and sensors directly to the centralized control functions (virtual SPS) in the cloud. In the optimal case all of the entities of a industrial fabrication unit are connected directly to the cloud and can be used as plug-and-work unit without any additional data cables. Just few control functions are remaining locally in the fabrication unit to maintain extremely short reaction times Microseconds.

There are robots moving between the fabrication units to carry materials between. Each robot is equipped with at least one manipulator arm to pick up materials. The robots are working in a cooperative manner to carry items together to the next production station. This cooperation is possible because of high reliable radio links between the robots and the virtual SPS.

Multiple production units are equipped with manipulators, actuators and sensors. Most of them are connected to a 5G radio subsystem. Others are connected via cables to a 5G radio hub. All provides a high reliable and low latency connection to the 5G base station and to the industry 4.0 cloud. A demo will show the reliable operation controlled by a centralized computer system (virtual SPS). Demonstration of flexible reconfiguration by adding actuators, sensor like cameras on demand. The virtual SPS in the cloud will automatically reconfigure the production processes based on the new capabilities. Moving robots will take together bigger items and carry it reliable to the next processing station. To do this the robots have to balance perfectly goods between them. It can be shown that the precision of the cooperation depends on the reliability of the 5G based communication links.

5G enhanced automatic driving: actually many companies spending huge effort in research on automatic driving. The current 3GPP is working on LTE based solution to support safety aspects with vehicle to vehicle and vehicle to infrastructure communication. Vehicular communications is a challenging use case for 5G because it could cover a wide area of applications such as traffic safety applications, automated driving, etc. Automated and connected driving will be a pillar of Europe’s industrial renaissance. The vehicle of the future is part of a connected world where superfast digital networks give access to communication, higher safety, improved environmental standards, entertainment, knowledge and personal contacts, to anyone, anywhere and at any time. Positive benefits are anticipated: increased road safety and lower fatalities; increased fuel-efficiency and lower environmental impact; reduction of traffic congestion; and higher comfort standards for users, congestion less traffic flows even if the traffic density increases. Experimentation on vehicular communications is tightly related also to mission critical communications. The reliability and latency in today’s communication systems have been designed with the human user in mind. For future wireless systems we envisage the design of new applications based on 5G communication with real-time constraints, enabling new functionalities for traffic safety, traffic efficiency etc. In such a use case, the challenge lies in reducing mainly the E2E latency while ensuring high accessibility and reliability of the communication services independent of whether they are based on cloud services or on direct communication between vehicles. Distributed cooperative maneuver planning and execution is enabled by highly reliable 5G radio links between cloud based services but also between the vehicles itself. It should be demonstrated that resources are dynamically used to guarantee the reliably required by the different services to run cooperative maneuvers. Resources are provided by means of fragmented frequency bands to support flexible feedback channels and scalable data rates but also by means of direct links and cellular links.

Addressed use cases for the prototyping are a) cooperative maneuver planning and execution, b) dense platooning of mixed types of vehicles, c) critical intersection with maneuver control by cellular nodes, d) connectivity of cars in sub level parking places, e) cooperative sensor sharing, f) packet forwarder based on automatic driving cars supported by 5G.

4.4.6 Integrated Moving Networks

With 5G and its evolutions, users will expect the connected society to be available with no limitations, and users will make use of bandwidth-demanding services like augmented reality and virtual office applications, also when on the move. In this context, future vehicles and transportation systems may play an important role in wireless networks by providing additional communication capabilities and becoming an *integrated* part of the communication infrastructure to improve capacity and coverage of operator driven mobile networks. That is, in order to serve vehicular users effectively, one promising solution is to deploy moving base stations on the vehicles to form moving networks.

One of the purposes of the moving base stations is of course to effectively serve in-vehicle users, which is becoming more and more demanding for high data rate and low latency services with modern well insulated vehicles that have a very high penetration loss (>=25 dB) in combination with high (above 6GHz) carrier frequencies ranging up to millimeter-waves (<=100 GHz). But another very important opportunity is to enable moving base stations to act also as cooperative ad-hoc small cell base stations in the heterogeneous mobile networks in order to serve also out-of-vehicle users.

Figure 10 – New connectivity opportunities

Thus, there is a large unexplored potential to integrate moving base stations as ad-hoc network elements into the heterogeneous mobile networks with mobile operator controlled network nodes to form *integrated moving networks*, but there are also several key open research topics. To name a few, tracking a large set of mobile channels at high speed to enable advanced spectrally efficient and robust closed loop (massive) MIMO schemes in the moving backhaul links, designing closed-loop and cooperative interference coordination techniques in ultra-dense heterogeneous networks, resource allocation and resource slicing for versatile QoS services to meet key performance targets on outage, throughput, latency and energy efficiency, and enabling efficient mobility protocols in such integrated moving networks.

To design such closed-loop cooperative transmission and resource allocation schemes efficiently in hybrid heterogeneous networks consisting of fixed and moving base stations, there is a vast unexplored potential to take advantage of various kinds of side information, like road infrastructure information, driving route information, positioning and social networks. By looking into such sources of information, there is also a potential along the way that a lot of new services with associated business models could emerge. Key challenges here are also how to guarantee privacy, security and implement authentication and owner protection of these information sources.

If successful, such integrated moving network nodes can also be used to realize ultra-reliable communication links between vehicles and between mobile devices of traffic participants. These components are not part of intelligent transportation services for traffic efficiency and safety, such as pedestrians, cyclists, playing children on the streets, pets, etc. They can be used to take advantage of the mobile communications networks to assist traffic safety & traffic efficiency protocols, as well as autonomous driving, that so far are using dedicated V2X communications solutions.

Another unexplored opportunity is to benefit from the fact that modern vehicles are moving multi-sensor systems that are constantly collecting information that could be very useful to support development of smart cities, such as sensing air quality, need for road maintenance, monitoring of noise levels, weather forecasts, traffic congestion levels for route optimization of critical transports, etc. - information that municipalities can use to optimize the resource efficiency in the cities towards sustainability, and to implement a better city life in crowded cities.

There is currently no real effort to put together large scale testbeds to analyse the potential of moving networks.

4.4.7 Drone Based Networks

The concept of moving networks can be extended by including unmanned aerial vehicles (UAVs: drones, quadcopters, air balloons, etc.). Nodes like base stations (Remote Radio Headers, RRH, for example) could be moved while being connected to the rest of the network using different technologies like millimetre waves, Visible-Light Communication (VLC), etc. Furthermore, it could be possible to transport devices with computation capabilities to support virtualisation, in order to deploy virtual servers (web servers, CDN nodes, DNS servers, SIP proxies, etc.) and NFVI (Network Function Virtualization Infrastructure) for computing, networking and storage.

UAVs can be used to move network capacity, while at the same time providing additional services to the users and to the network providers, owing to their privileged position above the ground; for instance, for network monitoring purposes or for offering better connectivity to the users.

Because of energy, weight and size restrictions imposed by aerial vehicles, the design of the network nodes to be moved is constrained and requires to use resource-constrained platforms. Other constraints are related to the envisaged future limitations in terms of flight endurance, and to the evolution of regulations regarding UAV flights in congested areas. These technological and regulatory constraints and trends, require an integrated view, which can be achieved only by means of proper large scale experimental platforms.

4.4.8 Exploiting the Users' Social Structure

5G networks are envisioned to support up to 100x more devices and 1000x more traffic compared to the state-of-the-art. Resource allocation problems and wireless network optimization, at different layers or cross-layer, have been studied extensively in the wireless networks literature. More recently, there has been ample evidence in the literature of the role of the rich and inherent social structure of mobile users, which is largely untapped, yet, holds great promise for optimized network operation and enabling new services. However, to date, this social structure has not been fully exploited for the benefit of the network operators and users. This research direction deserves more attention from respective research communities due to its inherent multi-disciplinary nature which cuts across multiple areas, namely social studies, wireless networks and big data analytics. Device-to-Device Communications is a core component of 5G networks due to its key role in boosting capacity, traffic offloading as well as enhanced user experience. Thus, developing policies, algorithms and protocols at the intersection of Device-to-Device communications, social-aware networking and content caching at the edge, is a promising research direction for future 5G networks. Such techniques should prove useful in large venues with high user density, e.g., sports games, fairgrounds, etc., where the number of users exceeds the network capacity, giving rise to congestion. In such scenarios, socially-aware networking will offload network traffic to device-to-device communications.

New mechanisms are needed to exploit the wealth of anonymized data about the mobile users that are largely ignored in state-of-the-art wireless systems. This is a software-intensive approach to enhancing the traffic carrying capacity of wireless networks, without additional infrastructure or major investments. For instance, big data analytics should reveal valuable insights about the spatio-temporal characteristics of the mobile users' traffic, inferred from their daily behaviour and social relationships, in order to direct precious network resources where it is mostly needed to best match the user needs. State-of-the-art systems waste huge resources, unnecessarily, due to the largely static resource allocation mechanisms over a geographic region as well as ignoring the sharp spatio-temporal variations in the user traffic demand, over the day and across different days of the week. This, in turn, gives rise to the new research paradigm of socially-aware dynamic resource allocation and wireless network optimization.

Another important trend in content delivery in 5G networks is caching at the edge. It aims at reducing the content delivery load on the cellular infrastructure, offloading it to local device-to-device communications, which eventually results in reduced content delivery delays and enhanced user experience, accordingly. In this context, the social structure of the users, e.g., similar content interests, will play a key role in optimizing various caching paradigms, namely proactive, cooperative and opportunistic content caching. Proactive caching predicts user behaviour and caches content on the mobile device before demand, based on prior spatio-temporal resource availability (e.g., battery charge and wireless connectivity) and content demand patterns.

In summary, the users' social structure is extremely relevant to communication networks, and it has not received sufficient interest so far, from the viewpoint of experimental testbeds.

4.4.9 Testbed infrastructure for future converged satellite terrestrial networks research

Current activities in progress take advantage of existing and emerging 5G testbed infrastructures and investigate synergies and availability between the existing terrestrial implementations and satellite testbeds. Testbed infrastructures developed in the scope of the European framework programmes (such as FIRE testbeds and e-Infrastructures), are being made available for experimentation to enable the trial and evaluation of service concepts, technologies, system solutions and business models to the point where the risks associated with launching of these as commercial products will be minimised. In addition, also ESA/ARTES testbeds supporting proof-of-concept are being implemented considering 5G technologies. Finally, member states have contributed significantly to the development of such testbeds through national funding and have great interest in the proposed activity to leverage past investments. There is a need to streamline and coordinate separate developments such that use of funding is optimised and large scale testbeds are deployed supporting meaningful experiments. Future activities shall analyse the collaboration models of the existing networks and testbeds (terrestrial and satellite). For development purposes an advanced 5G based terrestrial testbed infrastructure shall be selected as baseline, and which shall (a) be a reference instantiation of a 5G oriented infrastructure capable of supporting various verticals, (b) be a playground showcasing best practices and key technologies and APIs necessary to realise the 5G vision, (c) support network and vertical applications function virtualisation, (d) be based and implement emerging industry standards and (e) be as close as possible to a production network. The developed testbed infrastructure shall be used to assess converged network architectures for different scenarios/services. Some scenarios that could be validated are: Terrestrial network backhauling; Professional and mobile enterprise networks (including service for sparsely populated areas); IPTV in combination with multicast/broadcast; Massive M2M, e.g. in logistics/transportation; Massive broadband; Critical machine-type communication; Disaster relief; Emergency and safety.

The development shall provide the basis of future experimentation and subsequent developments of enhanced features and validation capabilities. Additional outcomes of the activity shall provide (a) extensive performance results on advanced features of the future converged satellite terrestrial network architectures to support coverage, mobility, security, Quality of Experience (QoE) etc., (b) guidelines for use of converged satellite terrestrial infrastructures and technologies, (c) identification of problem spaces to be further investigated by research and (d) ultimately support meeting the 5G-PPP Key Performance Indicators as published by the 5G-PPP association.

5 Roadmap

Technology	T<=2017	2017<T<=2022	T>2022
SDN and NFV	• (Separate) Standardisation activities • NBI definition based on intents.	• Virtualisation improvements for network functions, specifically for resource-constrained devices. • Interfaces and support for hierarchical multi-tenancy. • Definition of open interfaces for a complete control of network and user devices. • Large-scale carrier-grade trials for SDN and NFV.	• Definition of algorithms to automatically configure networks based on user demands. • Deployment of the first commercial software networks.
Security	• Isolation of workloads from different tenants in the same host.	• Security integration in NFV and SDN standards. • Solution of security issues in virtualisation techniques.	• Security is integrated in all network components.
UAVs	• Regulations in place	• Definition of scenarios where UAVs improve network KPIs. • Optimisation of usage and performance • NFV/SDN and drones: test beds, field trials	• UAVs are active part of the commercial networks
Moving networks	• Consolidation of the concept	• Identification of data carriers	• Moving base stations are active part of the commercial networks
Radio networks	• 3GPP study item on 5G Radio network phase I completed	• 3GPP 5G standards phase-I and phase-2 completed. • Commercial network deployment in 2020, with initial focus on eMBB devices followed by IoT, 5G spectrum allocated at WRC'19 • 5G in unlicensed bands commercialized and inter-working with 5G in licenced bands	• 5G standards evolution, e.g. SON, SDN etc. for ultra-dense deployment completed • Terahertz communications moving to foreground from academic research to industrial R&D and lab demos (tbps wireless communications)

SatCom systems as part of next generation networks	• Scenario definition completed • Feasibility studies launched	• Detailed studies on integration and performance • Testbed and demonstrations • Standardization and regulation actions	• Implementation and integration
Optical Networks	• Technologies for slicing the optical spectrum • Spectral slicing • Exploring additional physical dimensions like space and spectrum (beyond C and L bands) • SDM with multiple SMFs	• Automated spectral slicing networks • Advanced time-slicing (TDMA) systems in Metro and Core • Technologies for S, S+ etc band • Technologies for SDM (multi-core fiber)	• Automated time-slicing networks • Automated multi-core SDM systems

6 Recommendations

6.1 Virtualised networks and services

Regarding the virtualised networks and services area, these are the recommendations extracted from all received contributions:

- *Security is a must*. New security mechanisms to isolate workloads in the same physical host have to be considered, to be included in the core of the designed architecture.
- Design a *scalable organic network architecture*. The new paradigm of software networks implies new concepts and approaches that have to be taken into account. Furthermore, it is extremely important to perform large-scale carrier grade tests to extract new results that can be analysed to understand the performance and feasibility of this new approach.
- *Push SDN and NFV to their limits*. There are several areas where these technologies have to be further studied:
    - Unify the northbound interface between applications and the network controller, to provide a standard interface to configure the control plane based on intents.
    - To boost current standardisation efforts to improve multi-tenancy support in NFV and SDN. All related standardisation bodies have to include and develop multi-tenancy concepts in their working documents. It is essential to define clear interfaces and mechanisms to offer multi-tenancy services from virtual providers to other (probably virtual) providers.
    - Promote research on specific virtualisation techniques for virtual network functions. Because network functions have different requirements than traditional virtualised services, it is recommended to optimise the virtualisation environment for this type of functions. Aligned with this goal, it is also recommended to boost research on energy reduction for network function virtualisation.
    - Include specific paragraphs in project calls to encourage research related to storage, computation and virtual networking on resource-constrained devices.

- Advance the integration of terminals in SDN to include their resources in the network. This way, communications like device-to-device or access sharing could be supported naturally by networks.
- Networks should be truly open to provide the kind of information required by some services, to receive notifications about flows and the state of the network.
- Related with the previous recommendation, network providers should provide mechanisms to accept new customers on demand and just in time.
- Stimulate research on SDN applications capable of handling user mobility on ultra-dense deployments, and forecast her position based on tracking and location algorithms.

6.2 Radio Networks and Signal Processing

Section 4.2 addresses the development of more and more efficient radio networks based on the following common research directions:

- Network layer optimization, including cross layer optimization, to support multi-services, muti QoS/QoE, and heterogeneous configurations;
- New radio protocols and signal processing for efficient backhauling and fronthauling in heterogeneous networks;
- Channel modelling for TeraHertz and Visible Light communications, as well as, for new network paradigms such as integrated moving networks;
- Exploitation of side information for performance optimization;
- Social-Aware network management and configuration;
- Design of Satellite Communications components;
- Integration of satellite networks in future networks;
- Development of Software Defined and Cognitive Radio tools for the new frequency bands;
- Development of open-source tools for radio network optimization.

6.3 Optical networks

Optical networks will play a central linking role in the creation of a ubiquitous Digital Environment. Notably, optical networks are performing the heavy duty task of providing the necessary interconnection between the various heterogeneous local networks either at close range or between remotely located places. Optical networks will be a key enabler in integrating heterogeneous Access platforms, spurring technology-agnostic service creation, and facilitating dynamic Clouds through the convergence of Datacentres and Networks in Core and Metro.

From the analysis presented in section 4.3, the following challenges need to be addressed in the optical networking domain:

- *The need for capacity:* All evidence we have today is telling us that the multitude of pervasive technologies in access will result to an explosion in the volume of the created data. Optical networks in the Campus, Access, Metro and Core domain should scale up to match "supply" to "demand".
- *The need for dynamicity:* a striking feature is that the volume of data is not only increasing but also there will be a considerable differentiation in the requested

response times expected from the optical networks. This dynamicity creates the need to come up with optical networks amenable to dynamic reconfiguration at different time scales, to support services with diverse performance requirements that run over the same infrastructure.

- *The need for homogeneity:* concerning processing and storage, future optical networks will mainly be built from whiteboxes something that will turn *virtualization* into a global event, while concerning optical technology, which is employed for creating, manipulating and steering light, we will still have to employ purpose-built subsystems due to the different physical processes involved. However, through the advances in manufacturing and interoperability, and the deployment of commodity signal processing systems, there will be a seamless integration of processing and transmission facilities at different scales: starting from within the node this integration will be extended to include local, regional and finally global networks.
- *The need to operate at different scales:* the integration of the local and the global, to allow operating seamlessly at different geographic scales, is unthinkable without a ubiquitous control and management platform based on open software. Even if the optical and electronic industries are still evolving in parallel paths, virtualization - homogenization in general- will allow to slice, integrate and orchestrate a pool of dissimilar resources in a generic way. These are essential steps towards implementing full programmability, leading to automation of the orchestration tasks from Access to Core. However, the ability to demonstrate coordination at such different scales, ranging from intra-node to global, is still an elusive task.
- *The need for novel architectures:* fresh approaches are needed to implement these advances. A rigid distinction between network segments and the excessive use of aggregation stages becomes a bottleneck to faster service creation and to lower CapEx/OpEx systems. Novel architectures that drive site consolidation, foster the convergence of the heterogeneous systems in the Access, and lead to Access-Core integration, are needed. Optical network architectures exploring balanced distributed and centralized processing schemes for more effective service creation are also sought. Last, but not least, efficient architectures deepening the convergence between networks and Datacentres, fusing intra-node with global operations, are also of primary importance.

Other important research topics include: encapsulation and framing formats to transport dissimilar protocols and serve applications with different latency and packet loss requirements; multi-layer monitoring techniques to implement cross-layer troubleshooting and to allow the use of the obtained data, in conjunction with data analytic techniques, for cognition and application awareness. The latter are essential ingredients of a fully automated optical network. Finally, optical technology, that has been the enabler of the optical networking revolution, is expected to play an important role in 5G networks as, for example, in fixed/mobile convergence platforms and miniaturized pluggable interfaces in Metro.

6.4 Experimentation with verticals

In order to support next generation networks development, the following aspects related to experimental and vertical sectors shall be addressed:

- Inter-disciplinary research should be fostered
- De-verticalisation of experimental facilities should be sought
- Large scale testbeds for Public Protection and Disaster Relief applications are needed
- Large scale testbeds for Safety Critical applications are needed
- The impact of sociality in radio networks should be better investigated.
- The paradigm of moving networks, where base stations are positioned over vehicles, or drones, should be deeply investigated to achieve better coverage, and augmented capacities.
- Future activities shall analyse the collaboration models of the existing networks and testbeds (terrestrial and satellite).

References

[1] ITU-T P.912, Subjective video quality assessment methods for recognition tasks (2008)

[2] http://www.eena.org/download.asp?item_id=153

[3] http://www.indect-project.eu/

[4] http://www.foi.se/en/Customer--Partners/Projects/ARENA/ARENA/

[5] Leszczuk, Mikołaj, et al. "Recent developments in visual quality monitoring by key performance indicators." Multimedia Tools and Applications (2014): 1-23.

[6] De Mari, M., Strinati, E. C., & Debbah, M., Matching Coalitions for Interference Classification in Large Heterogeneous Networks, PIMRC 2014

[7] Felicia Lobillo, Zdenek Becvar, Miguel Angel Puente, Pavel Mach, Francesco Lo Presti, Fabrizio Gambetti, Mariana Goldhamer, Josep Vidal, Anggoro K Widiawan, Emilio Calvanese, "An architecture for mobile computation offloading on cloud-enabled LTE small cells and Methodology and Tool for Energy Consumption Modeling of Mobile Devices" IEEE WCNC 2014 workshop CLEEN, Istanbul, Turkey, 6-9 April 2014

[8] Jessica Oueis, Emilio Calvanese Strinati, Sergio Barbarossa, “Distributed Mobile Cloud Computing: A Multiuser Clustering Solution,” International Conference on Communications (ICC), 2016 IEEE, 23-27 May, 2016).

List of Contributors

Contributors	Company/Institute
Arturo Azcorra (SRIA editor)	IMDEA Networks and U. Carlos III
Jaime Garcia-Reinoso (chapter editor)	University Carlos III
Alexandros Stavdas (chapter editor)	University of Peloponnese
Raul Muñoz (chapter editor)	CTTC
Jörg-Peter Elbers (chapter editor)	ADVA Optical Networking
Alessandro Vanelli-Coralli (chapter editor)	University of Bologna, Italy
Roberto Verdone (chapter editor)	University of Bologna
Olav Queseth (chapter editor)	Ericsson
Maziar Nekovee (chapter editor)	Samsung
Jean-Sébastien Bedo	Orange
Pedro A. Aranda Gutiérrez	Telefónica I+D
Diego López	Telefónica I+D
Raffaele Bolla	University of Genoa
Roberto Bruschi	University of Genoa
Franco Davoli	University of Genoa
Felipe Gil Castiñeira	Universidade de Vigo
Francisco Javier González Castaño	Universidade de Vigo
Carlos Pérez Garrido	Universidade de Vigo
Jorge Fontenla González	Universidade de Vigo
Adam Flizikowski	ITTI Sp. z o. o.
Krzysztof Samp	ITTI Sp. z o. o.
Piotr Tyczka	ITTI Sp. z o. o.
Nikos Koutsouris	WINGS ICT SOLUTIONS LTD
Kostas Tsagkaris	WINGS ICT SOLUTIONS LTD
Panagiotis Demestichas	University of Piraeus
Artur Hecker	Huawei ERC
Dirk Trossen	InterDigital UK
Jon Crowcroft	University of Cambridge
Mike Bursell	Intel
Christos Bouras	University of Patras and CTI
Giada Landi	Nextworks
Giacomo Bernini	Nextworks
Josep Mangues	CTTC
Ramon Casellas	CTTC
Ricardo Martínez	CTTC
Ricard Vilalta	CTTC

Michela Svaluto Moreolo	CTTC
Josep Maria Fàbrega	CTTC
Dimitra Simeonidou	University of Bristol
Yan Yan	University of Bristol
Anna Tzanakaki	University of Bristol
Reza Nejabati	University of Bristol
Sergi Figuerola	i2CAT
Eduard Escalona	i2CAT
Oded Raz	Eindhoven University of Technology
Nicola Calabretta	Eindhoven University of Technology
Nick Parsons	Polatis
Salvatore Spadaro	Univeristat Politècnica de Catalunya
Josep Prat	Univeristat Politècnica de Catalunya
Ernesto Ciaramella	CNIT
Antonella Bogoni	CNIT
Emma Lazzeri	CNIT
Nathan Gomez	University of Kent
Mikhail Popov	ACREO
Domenico Siracusa	CREATE-NET
Tinku Rasheed	CREATE-NET
Marco Ruffini	Trinity College Dublin
Nicola Ciulli	Nextworks
Jean-Charles Point	JCP-CONNECT
Dirk Breuer	DEUTSCHE TELEKOM
Jose Torrijos	TELEFONICA
Stéphane Gosselin	ORANGE
Annie Gravey	Institut Mines –Telecom
Björn Skubic	ERICSSON
Achim Autenrieth	ADVA Optical Networking
Stefan Host	LUNDS UNIVERSITET
Achille Pattavina	POLITECNICO DI MILANO
Tibor Cinkler	BME
Peter Olaszi	AITIA
Christos Matrakidis	University of Peloponnese
Juan Pedro Fernández-Palacios Gimenez	Telefonica
Andrew Lord	British Telecom
Marco Schiano	TelecomItalia
Giuseppe Ferraris	TelecomItalia
Peter Vetter	Nokia

Sebastien Bigo	Nokia
Thomas Pfeiffer	Nokia
Jean-Pierre Hamaide	Nokia
Gabriel Charlet	Nokia
Rene Bonk	Nokia
Peter Ohlen	Ericsson
Fabio Cavaliere	Ericsson
Gemma Vall-Llosera	Ericsson
Patryk Urban	Ericsson
Yabin Ye	Huawei
Gernot Goeger	Huawei
Changsong Xie	Huawei
Ronald Freund	HHI-Fraunhofer
Colja Schubert	HHI-Fraunhofer
Nicolas Perlot	HHI-Fraunhofer
Johannes Fischer	HHI-Fraunhofer
Volker Jungnickel	HHI-Fraunhofer
A.M.J.Koonen	Eindhoven University of Technology
Ioannis Tomkos	AIT
Dimitrios Klonidis	AIT
Luis Velasco	UPC
Didier Colle	University of Ghent
Wouter Tavernier	University of Ghent
Carla Raffaelli	University of Bologna
John Mitchell	UCL
Stephane Berghmans	Commscope
Jordi Domingo-Pascual	Polythecnic University of Catalonia
Evangelos Kosmatos	Wings
David Lund	HW Communications
Andreas Timm-Giel	Technische Universität Hamburg-Harburg
Maciej Muehleisen	Technische Universität Hamburg-Harburg
Tommy Svensson	Chalmers University of Technology
Shahid Mumtaz	Instituto de Telecomunicações, Aveiro, Portugal
Jonathan Rodriguez	Instituto de Telecomunicações, Aveiro, Portugal
Ana García Armada	Univ. Carlos III de Madrid, Spain
Alejandro de la Fuente Iglesias	Univ. Carlos III de Madrid, Spain
Thomas Zemen	Austrian Institute of Technology
André Bourdoux	IMEC

Nassar Ksairi	Huawei Technologies Co., Ltd.
Ingmar Land	Huawei Technologies Co., Ltd.
Ton Koonen	Eindhoven Univ. of Technology
Egon Schulz	Huawei Technologies Duesseldorf GmbH
Tamer ElBatt	Nile University, Giza, Egypt
Amr El-Sherif	Nile University, Giza, Egypt
Mohammed Nafie	Nile University, Giza, Egypt
Fadel Digham	Nile University, Giza, Egypt
Andreas Georgakopoulos	WINGS ICT Solutions Ltd
Evangelos Kosmatos	WINGS ICT Solutions Ltd
Voula Vassaki	WINGS ICT Solutions Ltd
Ioannis-Prodromos Belikaidis	WINGS ICT Solutions Ltd
Alexander Markhasin	Siberian State University of Telecommunications and Information Sciences
Anwer Al-Dulaimi	EXFO Inc., Canada
Carlos Mosquera	University of Vigo, Spain
Michele Luglio	University of Rome, Tor Vergata, Italy
Rainer Wansch	Fraunhofer Inst., Germany
Christian Tohde	Fraunhofer Inst., Germany
Thomas Heyn	Fraunhofer Inst., Germany
Alexander Hofmann	Fraunhofer Inst., Germany
Christoph Mecklenbräuker	EURASIP, Special Area Team "Signal Processing for Communications"
Giuseppe Caire	EURASIP, Special Area Team "Signal Processing for Communications"
Ana Perez	EURASIP, Special Area Team "Signal Processing for Communications"
Carla Fabiana Chiasserini	Politecnico di Torino, Italy
Matti Latva-aho	University of Oulu Finland
Alessandro Guidotti	University of Bologna, Italy
Giulio Colavolpe	University of Parma, Italy
Giovanni E.Corazza	University of Bologna, Italy
Daniele Tarchi	University of Bologna, Italy
Emilio Calvanese Strinati	CEA-LETI, France
Anastasius Gavras	Eurescom GmbH
Adam Kapovits	Eurescom GmbH
Maria Guta	European Space Agency (ESA)

Annex A Other contributions

A.i Drivers and Tussles of Digitization

The knowledge-based economy: An economic epoch is defined by three factors,

- *The production systems and the tools these systems are employing*
- The *means of communication means and those for information creation*
- The type of *energy (power) systems* engaged.

In each epoch there is a certain relationship and interdependence between these three factors and so far each of these factors was founded over a certain technology which was uncommon in these three factors. The characteristic of the knowledge-based economy we have entered, as a clearly identifiable new economic epoch, is the dominant role of information and communications systems which are emerging as the common ingredient all the three factors have. This presupposes the *digitization* and the *standardization* of all human activities and of all processes in their lives. As the matter of fact, the digitization and standardization allow the *convergence* of the three factors into a new framework. Moreover, digitization allows *biotechnology* to emerge, for the first time, as the fourth factor in the definition of an economic epoch.

The distinctive feature of the knowledge-based economic activity is that it provides a new framework for planning, organizing and coordinating resources (human or material) to an unprecedented scale, while removing all barriers associated with *location*, *distance* and *skills:* we are entering the era where all technology-related restrictions which confine humans to implement their social functions at certain locations only, are being removed. This opens up the opportunity to select, mobilize and utilize the most diverse parameters and assets needed to complete a particular task regardless of their actual physical location and regardless of the nature of these tasks. In this way, social functions, like the deployment of new production tool systems in an industrial complex, the handling of agricultural robots in the fields, the employment of integrated health-care systems, the operation and control of distributed and renewable energy systems, scientific research, education, recreation and cultural development, etc will be rooted to the same processes.

We acknowledge that the *knowledge capital*, which knows no geographical boundaries, is an indestructible entity in striking contrast with its first or second industrial-wave counterparts that were often subjected to dissipation. Moreover, unlike physical resources, the knowledge capital is not exhaustible; it can be almost instantly transferred anywhere and can be employed or used simultaneously from an unrestricted number of human users or machines. In parallel, this new era brings its own challenges: the *coordination* of diverse and complex operations is an extremely challenging and critical task. The knowledge-based economy is ushering a revolutionary wealth system but it also includes many unchartered directions, the implications of which are not fully understood and accounted for in all respects at this moment, while we are not fully prepared for the many alternative outcomes.

The Instruments: We have passed the point where national economies are integrated to regional ones transforming the previously isolated markets into vast market clusters. Today, regions within a single nation, countries, unions and even whole continents are immensely and inextricably linked creating new opportunities while the economic interests are shifted to new directions. In this framework, global economy is characterized from interdependency and cooperation on one hand and from competition on the other.

This evolution has created global reach enterprises that may leverage the features of the knowledge-based economy to become completely location-independent, capitalizing on an automated connectivity to and between any process, person or thing; moreover, they are designated from their ability to scale to meet any demand. This ability to scale to worldwide demand but still be able to provide for the delivery of services or goods locally requires a certain organizational and technological framework. On the organizational front, global alliances or consortia are formed to establish global reach with ubiquitous local access in each geographic market. Each of these alliances will likely be anchored to one or more large global industry leaders or service providers. In such a dynamic world, the enterprises with the most automated operations and the alliances with optimized global-local reach operations are those that will benefit the most and will be able to leverage their dynamic agility to win in the marketplace.

However, the organizational framework alone is not sufficient to safeguard better, lower-cost, value-added digitization and service creation. A common, standardized technology platform, overcoming the current picture of a plethora of over-fragmented and single-purpose technologies, is a necessity that will bring into life the potential of a knowledge-based economy matching the ability of the over-the-top enterprises to coordinate on both “global” and “local” scale. This is the point where the next generation networking is expected to play a central role.

The scope of this section is to present a comprehensive strategy to address the communication/networking challenges of this new economic epoch and through this to ensure that the European leadership is fostered. To do so, excellence in science has to be accompanied by steps ensuring the rapid adoption of innovation in the business world. Actions in this direction are the even closer interaction of all actors in the stake-holder chain and a shift in emphasis in the service and application creation from the leisure-time to the activities that directly influence economic activity e.g. to the activities proclaimed in Industry 4.0.

A.ii Flexible and Suitable Infrastructures

Today, any user needing a telecommunication service can get it through a network operator that will try to satisfy the user requirements at the best leveraging on its infrastructures. In case no network operator has the capability to fully satisfy user requirements, the user must accept the service from the operator that best fit the request. Moreover, if the user requirements are dynamic over the time there is no possibility to change the architecture and the service level just for the required time interval. Flexibility both from architecture point of view and as a function of time is a primary and very innovative need for the user in the scenario determined by current and future Internet services and paradigms.

To satisfy target QoS/QoE, it may be necessary to set up of a heterogeneous network or to select a particular technology, that implies enhanced infrastructure flexibility, even dynamically exploited, and introducing or enhancing flexibility in capacity management and booking (in case of single operator but more important in case of multiple operators). This process must be transparent to the user. The selection of a particular technology (LTE, Optical fiber, Satellite, WiFi) or of a particular operator can be driven by QoS/QoE satisfaction.

QoS/QoE control needs two types of actions: i) monitoring and measuring relevant parameters, ii) reacting to reach target performance. The latter phase, driven by the processing of information collected in the former, can imply modification/improvement of relevant protocols (from resource management up to application) and modifications of the architecture (including optimum segment selection to be implemented in real time). To achieve real global coverage (from the ocean to the basement of a building, from the remote mountain peak to the narrow street of a big city surrounded by very tall skyscrapers) the network shall be composed of many segments (terrestrial fixed, terrestrial mobile, satellite, UAV) realistically operated by several different operators and some of them not operated at all (NFC, Ad hoc networks). To get service-level awareness in real time implies that the different networks are really integrated at management level and not only simply interconnected. Full network flexibility (including real time protocol and architecture reconfigurability) will be an irreplaceable value added if we want to ensure the maximum level of QoS/QoE satisfaction.

An innovative concept to be proposed is the ‘Infrastructure Composition Awareness’ and ‘Dynamic Infrastructure Set up’. In fact, different operators can implement up to date network management solutions (SDN, NFV, etc.) but with very slight difference with respect to one another and not at the same time. As a consequence, different network composition will provide different performance and the situation can vary from one day to the day after.

The goal can be to set up a framework to manage different segments operated by different operators or even managed by the single final customer (for example the home or office local network or the NFC connection). The framework must have the capability/entitlement to manage up to the single modem or to give orders to networks elements with the aim to guarantee the target QoS or QoE.

Several key parameters of a telecommunication network, which are important for performance evaluation, must be identified according to the scope. They span from lower layers only (e.g., physical and network performance) to higher layers (e.g., Quality of Experience - QoE, protocol timeouts, efficiency). In general, each parameter plays its role (directly or indirectly) for the success and quality of the resulting internet service, but a direct relation among them and as final QoE is not always straightforward. A possible list of key performance indicators (KPI) in an Internet based advanced network and their application context is reported in **Table 1**.

Table 1: KPI general definition

Layer	Technology/Standard	KPI
Physical	PHY of wireless and wired IEEE standards 802, 4GPP, 5GPP, radio propagation models, fiber propagation, Modulation, link budget	SNR, Availability, Coverage, propagation delay, attenuation, spectral efficiency, Collisions, power.
MAC	MAC wireless and wired IEEE standards 802, 4GPP, 5GPP, BoD, FEC, QoS, Synchronization, SDN, Security.	BER, Access Delay, Contention, queuing time, Overhead, framing efficiency, switching time, frame loss, frame priority.
Network	IP, IPv6, NAT, Flow Classification, IPSec, mobility, handover, Proxies, Network Coding, Firewall, SDN,	PER, Inter-networking adaptations needed, integration effort, goodput, APIs functions for operator inteactions, QoS mapping.
Transport	TCP, UDP, other protocols, Proxies, Security (TLS), firewall, CDN	RTT, protocols efficiency, throughput, delivery time, channel utilization, fairness, friendliness
Application	SIP, HTTP, FTP (and all other main internet protocols), Streaming, Proxyng, SCADA, Shaping, DPI, firewalling,	Timeouts, QoE, overhead, protocols malfunctions, application blocks, quality stepping and degradation, service availability
N.A.	Other aspects of network integration, deployment, costs assessments,	System Capabilities, costs, complexity

In more details, Bit error rate (BER), packet error rate (PER), transmission capacity, resource utilization, end-to-end delay (RTT) are important to evaluate each single segment independently. They can be measured quite easily (although problems of scale associated to value collection, elaboration and fusion can be relevant) and are related to networks performance. On the other hand, QoE is the most complex, difficult to define (Experience is a subjective human-related factor, while KPI shall be strictly defined and measurable) and important indicator. It is not easily and directly measurable from lower layer indicator or through the interaction with the end-nodes.

As concerns congestion evaluation, detection through packet loss must be replaced by proper channel capacity estimation. Congestion prediction can be achieved through TCP acknowledgment reception statistics (in literature several techniques that can be improved) or other methods that shall be designed. Since, as above stated, a reasonable level of congestion indicates optimum capacity utilization, even a certain minimum level of congestion can be identified as a target performance.

To assess the target new service levels, specifically in realtime, higher layer KPI are more meaningful but lower layer indicators, which directly and indirectly affect the overall network performance are easier to be measured. At present there is no closed form relation between lower and higher layer KPI (as an example, it is not easy to predict, by BER and RTT measurements if a YouTube video streaming will suffer buffering/freeze). A valuable research target could be the identification of such kind of relationship.
Performance assessment may be carried out leveraging on networking measures and not mandatorily requiring the collaboration of the end-user (who may be unaware). New techniques based on heuristics, generalized traffic analysis, deep packet inspection, generation of QoE samples from networking statistics data fusion shall be investigated. This is a major challenge and an essential step to ensure operations of future networks.

The first beneficiary of this kind of introduced flexibility would be the user, who would always be able to get the best service through the best infrastructure, optimizing resources both in terms of performance and costs. Anyway, also the operators would get benefit from such an innovation because they would give value to their capability to ensure QoS/QoE through their infrastructure even to users of other operators, although for limited time.

A.iii Facilitating the exploitation of network and service virtualization and cloudification

As mobile broadband services have started to be part of most people’s everyday lives, network traffic is exploding, which is of course good for Communication Service Providers (CSPs), but at the same time the danger for offering services at a bad Quality of Experience (QoE) that doesn’t meet the expectations of the subscribers is emerging. CSPs want to tackle the need to expand their bandwidth capacity while avoiding any dramatic increase of CAPEX. They also desire to be in position to adjust the use of network and service resources on demand and scale resources up or down based on their business needs. In a couple of words, CSPs basically wish for agility and elasticity. And this is why they are expressing a very strong interest on adopting solutions based on Software Defined Networks and Virtual Network Functions (SDN/NFV), namely on virtualization. But a careful expert would argue that agility and elasticity are rather benefits of the cloud, which is a concept not interchangeable with virtualization.

According to wikipedia, virtualization refers to the act of creating a virtual (rather than actual) version of something, including, but not limited to, virtual computer hardware platforms, operating systems, storage devices, and computer network resources. It enables the optimized utilization of resources, as more applications and services can be packed onto the infrastructure. On the other hand, cloud computing offers through a broad network access, a pool of resources that can be assigned dynamically and on demand, while their usage can be monitored, controlled, optimized and reported. So, what the CSPs actually need is the Network Function Cloudification. To fully exploit the merits deriving from a virtualized cloud environment, it is required to go further than just porting network appliances and services from running on bare metal to running on Virtual Machines (VMs).

The main obstacle to overcome is the inherent statefulness of most of the NFVs and the lack of a clear separation between data and functionality, where only the former includes states that are stored in the cloud. Since the performance of CPU, network and storage I/O is constantly enhanced, such a separation allows a much better failure recovery, as well as better scalability. Any costly proprietary hardware capable to perform a single dedicated function will evolve to a set of software functions running as VMs on low-cost COTS hardware. Even the network topology is going into software through the SDN architectures and the network's control plane is separated from the underlying data plane. The vision is to design and implement platforms that will provide any functionality, including storage or resource management, as a micro-service or an "X" Virtual Function (XFV), facilitating the creation of innovative and sophisticated applications in a customizable manner. Each XFV will be offered as a discrete component, with specific functionality, probably packed as an individual image e.g. a unikernel. The unikernel includes the minimal set of libraries that are necessary for the execution of the related XFV, leading to multiple advantages in terms of security, efficiency, flexibility etc.

The creation of services and applications will be done by chaining XFVs according to a service graph in a dynamic and re-configurable way. This allows more easily to meet the desired QoS and QoE requirements, as well as to adapt to changing environment (network and software ecosystem) conditions. But at the same time, this entails the necessity of defining appropriate orchestration mechanisms, based on intelligence distributed in all components. Real time monitoring of information, contextual analysis and learning capabilities of the XFVs will enable the network to predict what, where and when something is needed. The decisions will be made based on context- model-driven inference and in a totally autonomic way, taking into account specific meta-data, namely policies, which can guide the behaviour of the execution containers. It is important to support policy creation at all levels, starting from the developer of a component, continuing to the provider of a service and the provider of the infrastructure, and reaching the final user of an application. There are several ongoing research and standardization activities to this direction, like e.g. ETSI NFV MANO (Management and Organization) and OASIS TOSCA (Topology and Orchestration Specification for Cloud Applications) as well as commercially available tools like e.g. Puppet, Chef, Ansible and Salt.

The design and implementation of frameworks or platforms that support virtualization, cloudification and the micro-service / XFV concepts for resource management, service orchestration and application deployment, will generate new business models for the communications' sector, by enabling almost any device, from simple IoT sensors to powerful compute servers, to be part of the infrastructure and managed through virtualization. It will also broaden the range of the involved stakeholders, from multi-national vendors to SMEs or small organizations that can provide highly specialized intelligent solutions. Innovation will find more ways to be injected into the market, while the whole ecosystem of networking (manufacturers, providers, users) will become more competitive and at the same time robust and efficient in most aspects, i.e. security, cost, QoS / QoE. Especially CSPs will be enabled to offer to almost anyone, high-end services currently available only to the largest enterprise subscribers, and most importantly, in a rapid, cost-effective and finally profitable way.

A.iv Dual Connectivity using LTE-LAA Primary and Small Cells

The large data traffic in the fifth generation (5G) networks requires overcoming the spectrum limitations by aggregating the licensed bands with the low-utilized unlicensed spectrum. The Long-Term Evolution-Licensed Assisted Access (LTE-LAA) is identified as the technology that utilizes the unlicensed bands using small cells (Scells), as shown in Figure 11. These LAA Scells operating as supplementary downlink (SDL) for the LTE primary cell (Pcell) and share the unlicensed band in coexistence with the Wi-Fi. A major consideration in the design of LTE-LAA is the temporary transmissions of the Scells that increases the demand for new synchronization solutions to address the wide network time alignment. In particular, employing dual- connectivity (DC) using licensed and unlicensed bands require to define new methodologies for distributing accurate timing reference across the network for symbol-time and carrier- frequency offset.

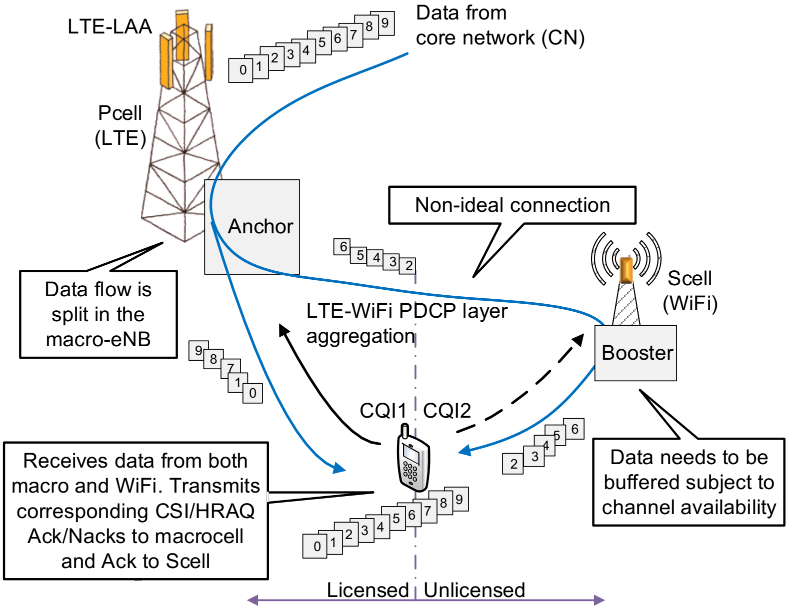

Figure 11 - Layout of data coverage for an LTE-LAA user connected to Pcell and Scell.

There is a strong need to design new schemes for performing synchronization in LAA consists of Scells under the umbrella of Pcells. The challenge is that a user equipment (UE) operating in LAA network model is actually connected to multiple number of BTS with signals propagating in different bands with different QoS characteristics, as shown in Figure 12. Therefore, delivering a load or performing a SIP calls cannot be guaranteed and the multi-connectivity might be terminated at any time due to the lack of sync and any changes in the wireless environment. One solution is to employ the user equipment (UE) as an anchor between the Pcell in licensed band and the Scell in unlicensed band to distribute accurate reference timing. This solution can be extended to include frequency offset between neighbor Scells by using the UE for sharing time stamps.

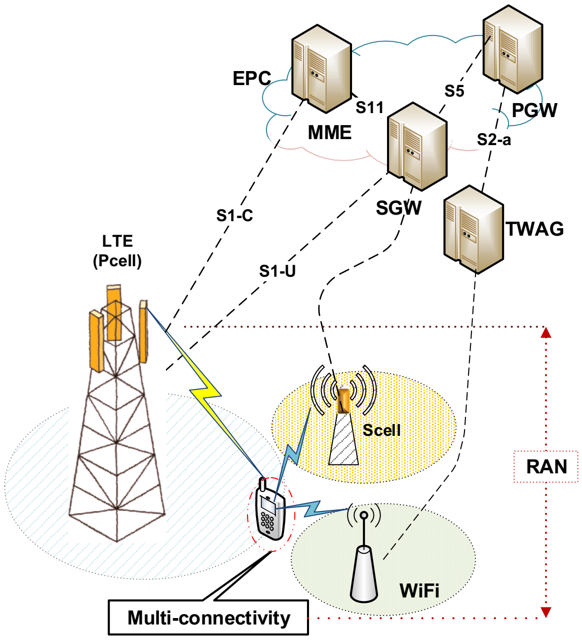

Figure 12 - UE with multi-connections of different QoS characteristics.

Since Sept 2015, the ETSI started to consider designing new coexistence mechanisms for LTE to operate in unlicensed spectrum. Together with the allocations of new bands for cellular communications across Europe specially the UK; there is a significant interest in operating the LAA soon as a milestone towards 5G. The requirements of the 3rd Generation Partnership Project (3GPP) for HetNet synchronization accuracy of 500 microseconds in 5G is a strong technical challenge that has not been identified carefully in the EU market. This benefits of using radio coordination features to support coupling small cell transmissions in unlicensed band with the cellular operations is of significant importance to the wireless market and any new ETSI standard.

A.v Quantum-Assisted Design of Wireless Systems

The potentially excessive complexity of numerous optimal classic communication schemes, such as the Maximum Likelihood (ML) Multi-User Detector (MUD) prevents their practical implementations. The design challenge becomes even more grave, when we consider the ML-detection problems of multi-user, multi-cell network-MIMOs operating in practical dispersive scenarios. In this context the powerful parallel processing capability of Quantum Search Algorithms (QSA) can be beneficially exploited for solving large-scale wireless optimization problems. Many of these classic problems can be more efficiently solved with the aid of QSAs, but this requires substantial further research from the broader community. This is particularly the case, when solving challenging multi-component optimization problems by finding the Pareto-front for jointly optimizing numerous parameters, such as the BER, transmit power, delay, etc under time-invariant channel conditions.

- This research is in its infancy owing to several factors, amongst others due to the extremely short coherence-time of the available quantum circuits, which hence impose both quantum bit-flips as well as phase flips. A potential remedy to this problem is to use the family of quantum-codes, synonymously also referred to as stabilizer codes.

- Promising opportunities open up by communications over quantum channels, such as their increased capacity potential. Classically the mutual information between the channel's input and output has to be maximized. Naturally, in case of quantum channels the capacity has to be redefined, leading to diverse scenarios to be considered.
- 3) A natural distinction concerning the channel capacity definition is, whether we restrict ourselves to classic bits as the system's inputs/outputs or not. In case of classic inputs/outputs we encode the input symbols/states into quantum states, send them over the channel and carry out a decision at the receiver side, effectively constructing a 'classic-quantum- classic' processing chain. This is a natural approach, since humans can only process classic information. By contrast, if we do not restrict ourselves to classic inputs/outputs, we are capable of dealing with quantum channels within larger quantum systems.
- The most important question arising in this context is, whether quantum channels are capable at all of increasing the achievable capacity and if so, under what conditions. The answer is definitively yes. Moreover, as a stunning result, redundancy-free error correction is possible over noisy transmission media, at least for a specific subset of quantum channels.
- And the science-fiction saga still continues... one of the hot research topics in this field is referred to as super-activation of quantum channels. Naturally, there are numerous quantum channels, which have zero capacity in the context of classic information transmission. But stunningly, when considering two of these zero-capacity channels used in a parallel manner and, additionally applying a special decoder operating by obeying the quantum-domain rules, the output of the decoder starts to deliver information...
- EXtrinsic Information Transfer (EXIT) charts have to be redefined for quantum systems and used for designing radically new systems. Since the so-called 'observation' of the quantum states destroys their fragile quantum-state, this 'observation' can only be carried out after all computations have been concluded in the quantum-domain. This requires the employment of syndrome-based quantum-codes, which is capable of avoiding the observation of the qbits.

A.vi Inferential Networks

Beyond 5G networks can definitely extend the typical role of wireless networks devoted to secure and efficient communication. In particular, we foresee that they can serve as inferential networks to infer various types of processes (in addition to transmitted messages for multiuser communication) in a multidimensional space. Spatiotemporal signal reconstruction of stochastic processes from samples randomly gathered in a multidimensional space is a crucial problem for a variety of emerging applications (e.g., environmental monitoring, crowd tracking, dynamic objects control). Multidimensional signal reconstruction has to face with uncertainties in the observations (measurements and sensors positions), signal properties (signal spectrum and spatial correlation), and sampling properties (inhomogeneous sample spatial distribution and sample availability). The signals emitted by beyond 5G network units can serve as signals of opportunity for inferring positions of objects in the space and for controlling their movements without the need of dedicated infrastructures and with reduction of electromagnetic pollution.

- Multidimensional random sampling generalizes the reconstruction of a stationary random process in one dimension, which, for regular sampling, was addressed by Balakrishnan and Lloyd based on Kotelnikov and Shannon sampling theory, while, for irregular sampling, was described by a Levinsons theorem establishing the condition for perfect reconstruction. Many authors have worked on this since the end of eighteens century (mainly in one-dimension), but none of them have studied multidimensional random sampling and optimal interpolation in multiple dimension when nodes are randomly scattered in space according to an inhomogeneous spatial distribution and the observations are affected by noise and by imperfect knowledge of sample positions. The inference of sensors positions using signal of opportunities presents several open problems, especially in soft-decision localization algorithm design and clutter removal filtering.

A.vii Cost-Effective Ubiquitous Rural and Remote IoT/M2M/H2H 5G Communications

A dramatic economical challenge (fundamental economic barrier) of the rural 5G forthcoming communications represent the mission-critical discrepancy of the widespread centralized (i.e., star-like) hierarchical low-meshed (i.e., Node-centric, Uplink-eNB-Downlink connecting) architecture with respect to the fully distributed topology of numerous polytypic data streams/'streamlets'. Unacceptably high investments are required into deployment of the optic core infrastructure for a ubiquitous wide covering of the sparsely populated rural, remote, and difficult for access (RRD) areas by traditional centralized broadband communication techniques because the profitability boundary vs the population density exceeds a several hundred residents per square km. The really indispensable approach for overcoming of the RRDs 5G economical barrier represents an increasing of broadband cell's covering area.

Rural and Remote 5G PHY Disruptive Green Technologies

As in SRIA "5G: Challenges, Research Priorities, and Recommendations", August 2014, the urban 5G LTE traditional broadband technologies, on the one hand, and the DVB-2T broadcasting communications, on the other hand, relies on two alternative approaches because of economic reasons: respectively LTLP (Low Tower - Low Power) and HTHP (High Tower - High Power). To overcome an economical barrier, the profitable ubiquitous rural 5G IoT/ machine type communications is required a "golden mean" approach HTFP (High Tower - Fundamental Limits Power) expressed through fine green 5G PHY disruptive smart techniques for close "on-the-fly" of fundamental physical limits:

- smarter increase a SINR trough beamforming/compact antennas radiation efficiency / adaptive pattern shaping for antenna arrays/orthogonality gain instead rise of the transmitter power;

a) close the fundamental minimum of invariant power consumption ICPE in mWt/[(bit/c)]/Hz;

b) provide “on-the-fly” the profitability-power-efficiency-aimed fundamental trade-offs for rural green 5G networks in practical invariant variables notions for energy-saving and spectral-efficiency.

N.1.2 RRD Distributed 5G MAC Disruptive Multifunctional Technologies

The MAC technology fundamental challenge of the 5G rural IoT/M2M/H2H communications are conditioned by so named time barrier and dynamic barrier. They represent phenomena of long-delay wireless medium access control performance degradation (time barrier) and dynamic instability (dynamic barrier) which appear when the propagation time increases many times. The flexible QoS-guaranteed multifunctional fully distributed dynamical medium access control (MFMAC) ensures the mission critical qualifications and disruptive technologies for improving an economical, spectral, and energy efficiency of the RRD 5G MACs:

- higher throughput and minimal overheads both close the fundamental limits; high efficiency, tolerance, and lower latency;
- high controllability, reliability, stability, flexibility, differentiation, and guarantee of distributed dynamical (“on the fly”) control of bandwidth resources, priority traffic parameters, and soft/differ QoS/QoE;
- multifunctional and universality abilities rely on the dynamically controlled and adaptive ATM-like smart unified protocol MAC, i.e., MFMAC, through the entire wireless networking hierarchy - core, backbone, and access networks;
- fully mesh all-device-centric radio access architecture all_device-to-all_device (DmD, m >>2) rely on the multipoint-to-multipoint (MPMP) Virtual Space/Wireless ATM Hyperbus topology;
- fine flexible 5G MFMAC scheduler should be capable to adapt "on-the-fly" the superframe formats to offered massive traffic and optimally allocating offered traffic by equal ATM-like minimal bandwidth blocks without superframe overflow or redundancy;
- multifunctional “all-in-one” smart recurrent M-sequences RS-token MAC-addressing tools supported “on-the-fly” techniques.

Rural Distributed 5G PHY-MAC Invariant Criterions

Usually, the spectral (SE) and energy (EE) efficiency criteria expresses through the Shannon’s capacity of the continuous channels with additive white Gaussian noise (AWGN) $C = \Delta F_s \log_2(1 + P_s / P_n)$, where ΔF_s is bandwidth, P_s – signal power, P_n – noise power, $P_n = \Delta F_s N_0$, N_0 – signal-sided spectral power noise density, in Watt-per-Hertz. The channel’s output, or receiver input, powers characteristic P_s / P_n named as signal-to-noise-ratio (SNR).

However, the continuous channels throughput capacity C allow study only the potential efficiency PHY values depending directly from three his spectral-energy basic parameters. So named, *invariant criterions* of spectral, power and energy efficiency allows to solve an optimization or trade-off problem depending from set of real conditions and parameters of the radio channels, methods of signals coding, formation, modulation, transmitting, receiving, processing, decoding, etc. Two invariant efficiency criterions were introduced for the wireless physical layer with orthogonal spread spectrum *m*-ary signals[i] :

- the invariant criterion for spectral efficiency (ICSE) was introduced as the digital channel's Shannon capacity $C_m(g, B_s)$ per Hertz ((bit/sec)/Hz), i.e., $c_F(m, g, B_s) = C_m(g, B_s)/(B_s/2)$, where g is channel-side, or receiver input, signal power invariant variable expressed via signal-to-interference plus noise ratio (SINR), $g^2 = P_s/(P_i + P_n)$, P_s, P_i, P_n – accordingly signal, interference, and noise powers, B_s is frequency-time invariant variable named as signal's base, $B_s = 2\Delta F_s T_s$, T_s is *m*-ary signal duration, *m*-ary digital channel's Shannon capacity $C_m(g,B_s) = \log_2 m + [1 - p_m(g,B_s)]\log_2[1 - p_m(g,B_s)] + p_m(g,B_s)\log_2[p_m(g,B_s)/(m-1)]$ in bit-per-symbol, where $p_m(g, B_s)$ is *m*-ary symbol's error probability (SER) defined through invariant variable $h(g,B_s) = g\sqrt{B_s/2}$ expressed, in turn, through receiver's output ratio signal energy per symbol to signal-sided spectral power additional Gaussian interference plus noise density $N_{0_{in}} = N_{0_i} + N_{0_n}$.
- the invariant criterion for power efficiency (ICPE) was introduced as the signal-to-interference plus noise ratio (SINR) per *m*-ary digital channel's Shannon capacity per Hertz: SINR/[(bit/sec)/Hz] $w(m, g, B_s) = g^2 / c_F(m, g, B_s)$.

Rural Distributed Green 5G PHY-MAC Fundamental Limits

For an ICPE invariant criterion, we can formulate the generally optimization problem $w(m, g, B_s) \to \min$ with constraints on the least permissible value of ICSEs $c_F(m, g, B_s) \geq [c_F]_{\inf}$ and, possibly, constraints on the greatest permissible values of cover efficiency ICCE and investment efficiency ICIE, i.e. profitability.

For an ICPE invariant criterion, was stated a fundamental power-consumption statement:

The minimal specific power consumption per (bit/sec)/Hz for fixed alphabet size m, and free g and B_s for Gaussian both the noise and the interference is an universal power constant which do depend neither on the signal base B_s, nor from free g, i.e., SINR.

For an ICSE invariant criterion, we can formulate the generally optimization problem $c_F(m, g, B_s) \to \max$ with constraints on the least permissible value of ICPE criterion $w(m, g, B_s) = \mathrm{const}(m) \equiv w_{\inf}(m, g^*, B_s^*) + o(w)$ where $o(w)$ is small Landau symbol.

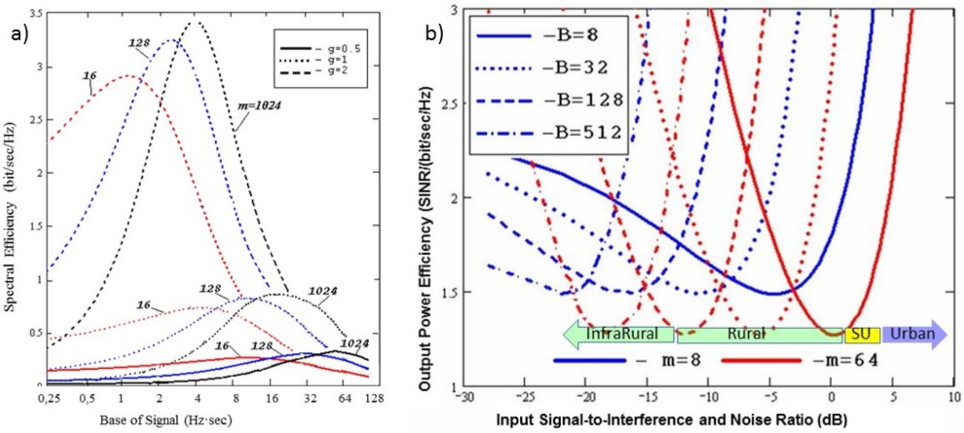

Figure 13 - The green invariant efficiency criterion generally optimization problems graphs: a) close upper limits of spectral efficiency ICSE (for urban); b) close fundamental minimum limit of power efficiency ICPE (for rural)

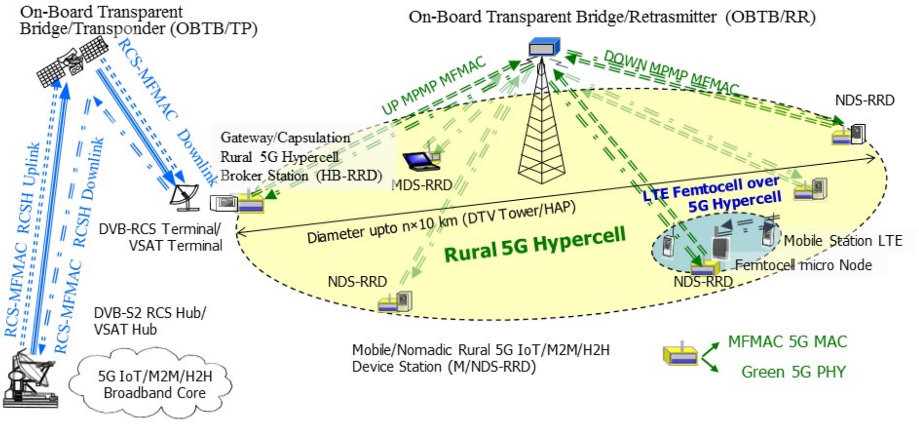

Figure 14 - Rural Green 5G Hypercelle Architecture

A.viii Technologies with significant impact in optical networking

Below we append a number of technologies that are expected to have a considerable impact to the evolution of Clouds:

- Silicon photonics that embed quantum dot laser technologies to bring baseball-size supercomputers to a reality
- Silicon photonic interposers that enable massive 3D integration of electronic chips
- Active silicon photonics, where silicon-lattice compatible lasers are volume manufactured
- Innovative and stable solutions for automated packaging of integrated photonic chips
- Electronics made from nanoscale tubes, wires, and sheets of carbon integrated with silicon photonics
- Electro-active polymers
- Nanophotonics to bring down the size of photonic integrated chips to a size comparable to electronic integrated circuitry

Last but not least, the gap between the proof-of-concept phase and volume manufacturing stage needs to be shortened for any integrated photonic platform (InP, SiO2, SOI, Si3N4, LiNbO3, Polymer) and combination of these ones with upcoming carbon based materials and other meta-materials.

A.ix Technologies to assist the convergence in Access

Massive MIMO

Massive multiple-input multiple-output (‘massive MIMO’) techniques are proposed as they can support picocell network architectures in an effective way. Many antennas need to be connected, and a fibre fronthaul network with radio-over-fibre techniques can be advantageously deployed to offload the comprehensive radio generation, modulation and signal processing functions from the antenna sites to one or more remote sites. The challenges are to devise efficient and robust radio-over-fibre techniques plus the added signal processing intelligence in order to realize cost-effectively such massive MIMO enabled picocell networks.

Radio beam steering

Within the radio picocells, additional ways to further increase capacity and reduce energy consumption can be realized by spatially-selective assigning radio capacity, i.e. by radio beam steering. The challenges here are to devise compact modules which can be co-installed with multiple-element antennas for remotely-controllable beam steering. Optical hybrid integration can offer powerful compact solutions for this, enabling also to do the beam steering under remote control from a network management site with the appropriate demand-adaptive intelligence.

High-transmission optical antennas

Technical details, importance and impact: High-speed (10 Gbit/s or more) wireless links over short distances (typically 500 m) can be provided by spatially coherent optical antennas. Typically the laser source is emitted as a single-mode beam through the air and is coupled, at the receiver, back into a single-mode fiber. This type of wireless links can benefit from all the technological advances in telecom fiber around 1550 nm, including signal processing and transmission formats. With a dispersion-free and polarization-maintaining atmospheric channel, transponders may even be simplified.

Concept and telecom sector affected: These optical antennas would meet the demand of high-speed wireless backhaul links in mobile-phone networks characterized by small-size coverage cells. Operating in the unregulated spectrum region, such systems would be free of interference and eye-safe. The high coupling efficiency between the two antennas would provide enough link power margin to compensate for strong attenuation (e.g. 30 dB) typically caused by fog. Meeting the user's link availability requirement is then only a matter of power-margin scaling.
The installation of optical antennas on building tops shall be similar to the one of microwave antennas: need for line of sight and alignment procedure in a reasonable time.

Innovation and challenges: Antennas with high spatial coherence are associated with high directivity and low transmission loss. The first step to guarantee high directivity is to keep the antennas in the right direction and perform a tracking of the wavefront tilt. Other degradations of spatial coherence generally occur as the laser beam goes through the antennas and the atmosphere. The inhomogeneity of the atmospehere's refractive index causes wavefront distortions with a typical coherence time of 10 ms. Although various adaptive techniques are conceivable and several of them have been recently implemented, they usually do not exceed TRL3. This is the case for the promising concept of optical phased array (OPA), which needs significant engineering effort to improve both performance and reliability. The OPA is an adaptive-optics technology without deformable optics, i.e. without mechanical parts. The spatially distributed array elements are phase-controlled to allow wavefront correction and/or beam forming (thus also beam steering). The OPA implementation can be fiber-based or integrated on a chip where waveguides and phase shifters are more compact, stable and scalable to many elements.

High speed visible light wireless communications

Although optical wireless is not a recent topic, it has seen a boost of activities since development and commercialization of today visible LEDs, which is becoming a pervasive technology. As LEDs are being deployed everywhere, and at the same time allow for fast modulation speeds, they can in principle be used as both lighting source and wireless data transmitters. The use of cheap devices of very long lifetime is also a key benefit. Research and industrial activities in visible light communications have demonstrated impressive achievements in the last few years (as example: using commercial LEDs, unmatched wireless speeds of 5 Gbit/s and wide coverage areas up to 20 m2.There is indeed a wide range of applications for this promising technology, which include primarily office/home high speed applications, but also vehicular networks, underwater communications, wireless in RF-free areas (hospitals, factories), board-to-board solutions for data centres, or centimetre level indoor positioning. Market potential is therefore impressive, but thus strong effort is required to exploit it fast and effectively.

The next step is indeed the evolution of VLC concepts, starting as point-to-point link laboratory technology, into true implementation design rules and specific realizations, i.e. a true LiFi network. First, full-fledged system implementations must be proven, with complete integration with existing lighting technology. However, higher bitrates, which are the final target of this technology, will require dedicated electronics and suitable modifications of the optics. Also, adoption of specific design and modelling has been proposed and is currently under study in IEEE P802.15.7r1 Short-Range Optical Wireless Communications Task Group. This is key to any future development since optical wireless requires detailed channel modelling, which is even more critical than in radio-wireless solutions, due to the much shorter wavelength. Network design based on the LiFi very small cells (around 20 m2) should be introduced and analysed.

Finally, following the trend of the Internet of Things (IoT), the future LiFi network should be able to communicate to a wide set of devices with heterogeneous capabilities. For instance LiFi devices may feature different types of optical transceivers, such as high speed photo-detectors, embedded CMOS camera sensors, and solar cells, which could be used both for harvesting and communications. Thus, the future LiFi network needs to be designed since the beginning while considering the unique requirements posed by future IoT devices. Proper addressing of these issues would require interplay of network operators, lighting industries, research centres and, most likely, industries working in consumer electronics.

Coherent optical systems in access

In the access low-cost solutions are essential. A technology that has attracted interest is coherent system for a PON-type access network where:

- A very large number of sites (residential, business, antennae) is served providing a wide variety of services with high flexibility thanks to the transparency to the transmission protocol.

- the existing infrastructure is re-used guaranteeing co-existence of 5G and legacy PON systems
- colourless operation for easy network dynamic reconfiguration is provided
- very high losses and long reach for consolidation purposes are supported
- UD-WDM can be used ensuring optimal exploitation of the optical spectrum
- the developed statistical wavelength multiplexing solution naturally performs flex-grid multi-format optics in a highly cost-effective form, providing fine granularity capacity management

This technology may also be used for metro back-hauling applications sine it allows systems with high sensitivity and high spectral density where needed, allowing the implementation of new metro-access network architectures more efficient in taking advantage of the existing infrastructure, more flexible in terms of service provision and with higher capacity.

Photonics for Comb Generation

Frequency combs, i.e. sets of equally-spaced lines in the frequency domain, are needed in several applications spanning from radio-frequency wave generation to metrology. In the optical communication scenario the applications where frequency combs are needed are mostly wavelength-division multiplexing and orthogonal frequency division multiplexing. Several methods for generating frequency combs for these applications are possible. An approach to comb generation exploits the spectral lines of a mode-locking laser (MLL) which guarantees a coherent comb but with a fixed line spacing. A gain switched distributed feedback (DFB) laser externally injected with a continuous wave (CW) laser can be also employed. Another approach is based on a single CW laser followed by a cascade of phase and amplitude modulators allowing independent tuning of the comb spacing and central wavelength. Continuous-waves seeding of a multistage mixer requires nonlinear passive components as optical fibers. The comb can be tuned acting on the seeding CWs, nevertheless the footprint is typically large. The combs generated with the different techniques can be enhanced by exploiting nonlinear effects in fibers and microresonators.

Despite the already mentioned solutions, photonics can still bring an added value to the research in this field, especially through innovation in terms of devices, materials and photonic integrated circuits design and development.

A.x Core and Metro technologies to accelerate the convergence in Clouds

Optical Transceivers

Beyond to what analysed in the section for Metro, low cost 400 Gbit/s system supporting more than 100 km transmission for metro network and inter Datacentre connections would be a key enabler. In particular, of interest are low cost direct modulation and direct detection technologies, design high intensity, low cost 400 Gbit/s transmission chips and modules.

Moreover, programmable and bandwidth variable transponders (BVTs) is a key enabling technology for future EON. Interconnecting heterogeneous access networks (e.g.

multiple wireless technologies) and edge devices with future EON which is also heterogeneous (comprising multiple technologies such as flexi WDM and optical OFDM) requires a new types of flexible transponders that can be programmed for interconnecting various edge devices and access technologies and also adapt their optical baud-rate and modulation format. The former can change the occupied optical bandwidth, while the latter can provide variable spectral efficiencies for different link distances. In addition, a cross-layer cross technology BVT would provide new network optimization and convergence solutions.

Transmission using the orbital angular momentum of light

To enhance the capacity of core networks, multiple switching domains can be exploited in multi-plane optical interconnection networks. One possible candidate is the Orbital angular momentum (OAM).

Orbital angular momentum (OAM) The concept of OAM communications arises from the fact that photons with orbital angular momentum are associated with certain classes of optical modes. It has been well known that the optical polarisation is associated with the spin angular momentum of photons, however it was only in 1992 that researchers realised that the OAM of photons is associated with the helical phase front of optical modes, with the the Poynting vector following a spiral trajectory. This kind of modes, however, was well known to exist in optical fibres. Known as ‘skew rays’, they had largely been a nuisance, in particular in graded index multi-mode fibres, where the reduction in inter-modal dispersion between the ‘meridian rays’ did not result in significant increase in bandwidth precisely due to the existence of these skew rays.

A classical picture of the cylindrical optical mode could be illustrated by the schematic of Figure 15 (left), where the components kz, kr and k_θ of the wave vector k are quantised by the boundary conditions, giving rise to mode indices [m, l]. m is the number of field nodes in the radial direction while l is the quantum number of azimuthal phase eigenfunction $\Phi=\exp(jl\theta)$. The orthogonal eigen-mode set supported by a fibre is characterised by a unique set of [m, l], corresponding to sets of k_r and k_θ

It has been physically proven that l represents the quantization of the orbital angular momentum of the photons in the optical modes, in that each photon has an amount of OAM of $l\hbar$ ($\hbar$: Planck Constant) and l can take positive, 0, or negative integer values.

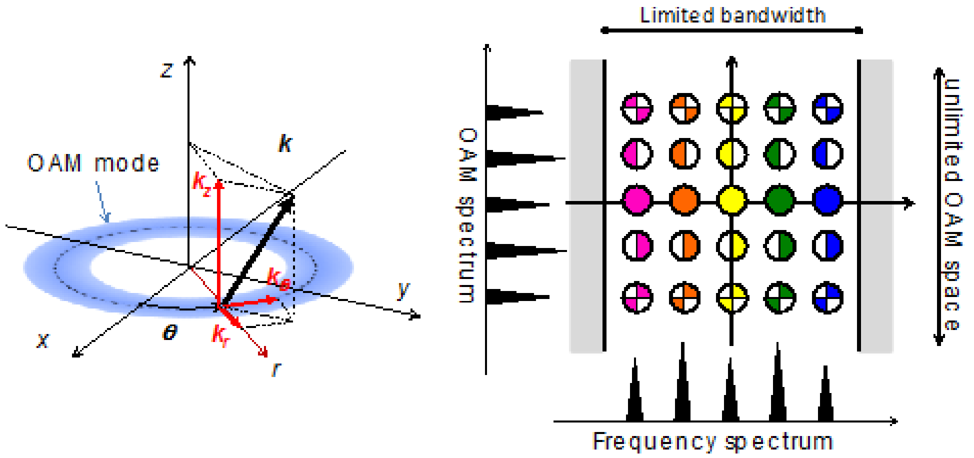

Figure 15 - Left: schematic representation of the structure and vector components of the cylindrical optical mode. Right: central concept of the project proposal illustrating the capacity benefits in exploiting the OAM space.

The central concept of the use of OAM in optical switching is illustrated in Figure 15. As the available spectrum for communications becomes fully populated, an additional dimension is made available by expanding each carrier frequency/wavelength into multiple OAM channels. This is equivalent to re-using the spectrum L times (L is the number of OAM states used), increasing the spectral efficiency by a factor of L. One can also immediately infer that this is made possible by the orthogonality between the OAM modes. It is also immediately clear that, in addition to increased transmission capacity and spectral efficiency, a new dimension of resources for networking, in addition to wavelength, time and space, has been made available by the ability to assign different OAM modes to the different data packets to be switched.

Progresses in OAM technologies will be a key issue to overcome the current state of the art and to foster the adoption of OAM based networking in real life.

A.xi Photonic technologies for Verticals

Remote sensing (RS) applications embrace wide range of sector, spanning from precision agriculture to maritime and land security and disaster prevention. The use of remote sensing systems, as a noninvasive solution, is well documented nowadays, and presents quite a lot of advantages compared to in situ observation in terms of area covered, type and quality of the retrieved data, realtime analysis and data processing.

The state of the art of RS present a wide range of solutions and are classified in passive (sensor retrieving radiation that is emitted or reflected by the object or surrounding areas) and active (the radiation is emitted by the sensor itself that then collects the radiation scattered by the target) systems. Among all the RS systems, RADAR (radio detection and range) and LIDAR (light detection and range) are mostly used to gather information in a noninvasive way.

The sensing systems above are currently limited in their further development by the following aspects:

- Active sensors make use of costly electronic hardware to generate the transmitted signal and their cost and power consumption increases with the frequency.
- Active RS are often designed and developed to cover a single frequency or a fixed set of frequencies
- Recently, the possibility to merge data derived from two or more RS sensors has been proposed. Data merging will help in analysing diverse target properties, or to refine the sensed data. However, difficulties in the data merging due to sensors different angle of view, timing, sampling grid, resolutions and so on, are arising, and are currently overcome by means of heavy data processing techniques.
- Electronic based generation of transmitted signal in active RS nowadays imposes the use of a single transceiver for each RS sensor, making coherence between sensors impractical.

To overcome these, research efforts in the field of microwave photonics systems, photonic integrated circuits design and development, and remote sensing systems will have to be carried out.

A.xii Use of frequency bands above 6GHz

To meet the steadily increasing throughput requirements of 5G and beyond 5G systems, the use of frequency bands above 6GHz, possibly going up to 90GHz, is widely perceived as a necessity. Actually, it appears that the high microwave band (say>20GHz) and the millimetre-wave band up to 90GHz will be part of many 5G radio access networks, for indoor connectivity, outdoor wireless access and even for fronthauling and backhauling densely deployed 5G small cells: millimetre waves are pervasive (see Figure 16). Furthermore, the boundaries between these applications are sometimes blurred e.g. when indoor connectivity can be used outdoors at low mobility or in the case of self-backhauling.

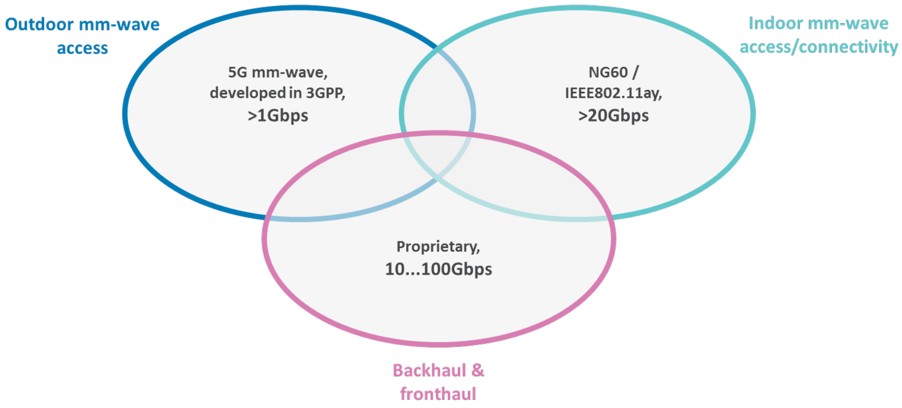

Figure 16 – Frequency usage in 5G

Very interestingly, the technology and challenges for these different applications (outdoor, indoor, front/backhauling) have many common challenges – requiring research and prototyping - such as: hybrid beamforming, wideband transceivers with very low EVM, large antenna arrays with interconnect losses and signal distribution over large distances, fast ADCs, DSP intensive digital baseband, etc... Hybrid beamforming in particular requires difficult trade-off to determine the optimum number of antenna elements (tens to hundreds), sub-arrays, front-ends, analog signals and digital streams especially in a multi-user and MIMO context. Challenges are present at the base station/access point side and at the terminal but with different constraints of size and power consumption.

Developing efficient hardware/software solutions for such millimetre-wave links will be key for the successful adoption and deployment of mm-wave links. In addition, additional challenges are beamforming solutions for link initiation, high user mobility, dense user scenarios, user/resource scheduling and multi-cell operation. The fact that the exact frequency bands for outdoor access is not yet clear (probably not clear until WRC-19) and that the outdoor mm-wave air interface/multiple access is not yet defined is an additional difficulty.

It is needed and urgent to bring mm-wave links or cells to higher maturity of implementation and demonstration (more real-time, higher throughput, larger number of streams/users, more integrated implementation, supporting higher user mobility, better system level performances...) to give European companies (equipment manufacturers, chipset manufacturers, telecom operators) a differentiating advantage when those mm-wave technologies become mainstream.

Linking the research/demonstration activities to verticals and use cases is also an important aspect that needs to be addressed so that the complete value chain is covered both horizontally and vertically. Integration in the complete network must also be addressed.

A. Markhasin, A. Kolpakov, and V. Drozdova, “Optimization of the spectral and power efficiency of M-ary channels in wireless and mobile systems”, in Proceedings of the Third IEEE International Conference in Central Asia on Internet. The Next Generation of Mobile, Wireless, and Optical Communication Networks – ICI’07, 26-28 September 2007, Tashkent, Uzbekistan, Paper # 177. pp. 1-5.