April 2018

Strategic Research and Innovation Agenda 2021-27

European Technology Platform NetWorld2020

“Smart Networks in the context of NGI”

2018

Table of Contents

1. Introduction

The preparation of Framework Programme 9 (FP9) as part of the next Multiannual Financial Framework of the EU is progressing. The NetWorld2020 European Technology Platform (ETP) and 5G Infrastructure Association (5G-IA), organisations representing more than 1000 entities, representing 5 % of European GDP, are contributing to the definition of research areas especially in the domain of communication systems and networks.

ICT in general and networks (mobile and fixed) in particular is a fundamental enabler of a modern society. The Smart Networks of the future will be the nervous system of the Next Generation Internet and other commercial networks and are the platform for driving the digital transformation. Future communication systems and networks (Smart Networks) are the foundation of the Human Centric Internet. They provide the energy-efficient and high-performance infrastructure on which NGI (Next Generation Internet) and other digital services can be developed and deployed. Smart Networks will apply intelligent software (Artificial Intelligence and Machine Learning – AI/ML)) for decentralised and automated network management, data analytics and shared contexts and knowledge. Such infrastructures are the enabler for the future data economy. By virtualisation and strict policies, they will foster a free and fair flow of data which can be shared whilst at the same time protecting the integrity and privacy of data which is confidential or private: Users should be able to control their environment in the Internet and not be controlled by the Internet.

The United Nations 2030 sustainable development goals [1] require Smart Networks in many different domains to support the digitalisation of society and economy in developing and developed countries. The United Nations Broadband Commission for Sustainable Development has set deployment targets for 2025 [2] to underline the importance of communication systems and networks.

Strong Contribution to the European Economy

The ICT domain contributes significantly to the European economy with about 5 % of GDP, which corresponds to a market size of about € 600 billion. The communication systems and networks sector (manufacturing including communication equipment and telecommunications) enables this market with

- about 28 % (1.76 million employees) of ICT employment [3],
- 40 % (€ 237 million) of ICT market size [3] and
- 49 % (€ 14.4 billion) of R&D expenditure in Europe [3].

These numbers do not reflect the multiplication factor of advanced communications in the economy. The Worldbank has shown that the availability of broadband access increases economic growth and employment [4]. Ecosystems connected to digital platforms and market places create value for all members and have the potential to disrupt entire industries and show significant economic and social impact. For example, the automation achievable by the Internet of Things across a broad range of sectors will lead to a potential economic impact in the range of $4 trillion to $11 trillion by 2025 [5]. A fully functional Digital Single Market could contribute €415 billion per year to the European GDP [6]. Overall, the digitalisation of society is still in an early stage. For example, Europe's Digital Progress Report 2017 [7] points out that only 20 % of the companies in the EU28 countries are highly digitised and there are still many opportunities to be exploited especially by SMEs. According to an Accenture study [8] the economic opportunity from digitalisation in Europe is over € 4 billion in value per day.

Smart Networks are of strategic importance as the enabler for basically all sectors in society and economy for jobs and economic growth.

Smart Networks Vision

5G is just the beginning of a new paradigm after the successful development of mobile communication systems such as GSM, UMTS and LTE. Further development is absolutely

crusial to address new challenges and requirements coming from many different sectors in society and industry. The smart network architecture will be software defined and provide features significantly going beyond connectivity: Multiservice and Mobile Edge Computing will allow to store and process data locally at the edges of the network to provide fast reactions and efficient use of network resources. Programmable aggregation and virtualisation functions as well as built-in security functions enabled (e.g. by the support of blockchains) will create a trusted environment for the Internet of smart things in which new applications and ideas can flourish. Future cost-effective communication systems and networks will increasingly be based on AI/ML and increased softwarisation in addition to requiring the continued development of classical communication technologies. Therefore, it is recommended to research future communication systems in close cooperation with these domains from an overall system perspective. The communication infrastructure will form the nervous system of the future Human Centric Internet and the digital transformation. It will intertwine distributed network, compute and storage resources to facilitate an agile composition of new services supporting a multitude of markets and industry sectors. From supercomputers and parallel computers, to data analytics, passing through cybersecurity, the Internet of Things (IoT), cooperative robots, or autonomous vehicles, it is universally agreed that every system and application must be interconnected to its peers, as well as to other related entities and systems. The interconnection of everything will be a distinguishable flavour of a competitive advanced society. Without network innovation the digital transformation is likely to fail. Therefore, it is vital that the Smart Networks area is adequately represented in Framework Programme 9.

The Smart Networks concept provides the necessary infrastructure and builds on scientific advances in the areas of physical and logical[1] sciences as well as key enabling technologies to provide a coherent framework supporting the future network designs. It is a combination of Smart Connectivity, Data Analytics (AI and ML), high performance distributed computing and Cybersecurity.

This Strategic Research and Innovation Agenda is summarising the different research domains to make the overall vision of Smart Networks happen.

[1] Logical science means the specialised logic and mathematical development applied in ICT and Computer Science.

2. Network Architecture and Control

Towards Smart Networks

Beyond 2020, we envisage a network- and human-centric world, based on a comprehensive network model and architecture for control which leverages the most innovative and promising research elements of the Network community of the last years.

This **Smart Networks** concept builds on the required scientific advances in the areas of physical sciences, logical sciences and key enabling technologies and aims to provide a coherent framework in support of:

- Integrated Connectivity, Computing and Control (the 3Cs).
- Converged fixed and mobile networks, integrating the 3Cs.
- Improved coverage and reduction of white spots.
- Support of hyper-converged overlays.
- Hundreds of trillions (10^{14}) of connected, active, devices and terminals.
- Massive numbers of tailored cost-effective services
- Automated and greatly cost reduced network operation.
- High societal, vertical, autonomous and cross-sector penetration: Energy, transport, health, entertainment, security, industry, aerospace and many other sectors relying on this infrastructure.
- Multi-sensorial interfaces, multi-environment and wearables.
- Reality enhanced with virtual and augmented reality.
- Support of drone fleets and autonomous vehicles.

Smart Networks will be a key enabler of all the other application domains in higher layers, which are built on top of high-performance communication systems and networks.

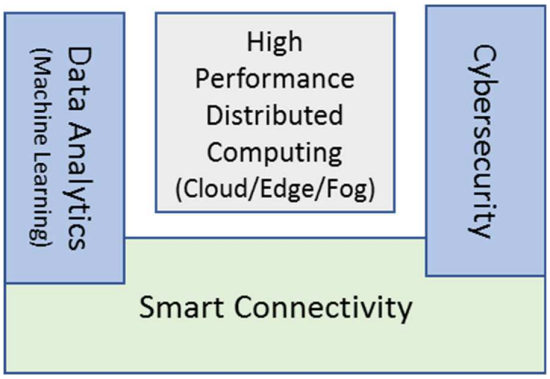

Figure 1 The Smart Networks concept

The vision of Smart Networks is a network beyond 5G, based on a single, unifying control framework that allows for instantiating and executing any control architectures, constrained by well-defined limits to the execution of each individual control architecture (Figure 1).

Key challenges the Smart Network control layer must solve are the aspects of control over multiple general-purpose, distributed, network control operating systems; the availability of powerful abstractions to resources to services; new naming schemes for virtualised resources; dynamic and automated discovery; intent-based open APIs and highly configurable policies to control the resource and service access and dynamics; isolation of application's execution environments and performances.

The future network infrastructures implementing the Smart Networks concept will make large use of technologies like AI/ML to implement data-driven closed control loops that can enable

cognitive (at first) and intuitive (then) network behaviour. The training and validation of such technologies require the availability of cross-technology and cross-sectorial datasets that do not exist yet. The networking research community needs to build those datasets, agreeing how they are generated, accepted and accessed.

For the Smart Networks we aim for comprehensive network control intelligence capable to handle the future communication network technologies and hence it is strategic to:

- **Build the cognitive and autonomic network service end-to-end orchestration** based on network and non-network functions and datasets (typically from the vertical application layer and beyond the mere communication aspects);
- **Allow dynamic pooling of local resources** from diverse participating devices;
- **Offer programmable analytics to the application** layer through open interfaces;
- Support and instantiate more and more **service intelligence at the edge as well as across cores**.

The following subsections discuss different fundamental problems that have been identified, as well as their related implications, in respect to future research challenges in Network Control.

2.1 Virtualised Network Control for Increased Flexibility

The Internet of the future will be a complex planetary system made of multiple physically interconnected elements, logically broken in separate islands, each possibly applying different security policies, routing mechanisms, access mode to application services. Resources will be configured and orchestrated dynamically to match the requirements of services running on the network.

A massive number of devices will be connected and will generate very large quantities of data. Useful insights will be generated based on the automatic analysis of all that data. The infrastructure that supports society will also be connected to the Internet, which will improve the effectiveness and efficiency of its operations.

The expected outcome of these trends will imply that the upcoming network will have to be interconnected to the Internet economy and to the cyber-physical infrastructure, and address security threats, in a world where AI/ML will be widely used. Open standards will be required, and governments will have to impose limits and regulations on the usage of all the data required to drive these new systems. In this context, overcoming the digital divide will be a key driver for technology evolution, and personal freedom and rights will need to be assured across all media.

Separate Decision from Enforcement

A key challenge is to separate enforcement (the ”how” part) from the decision (the “what” part), as well as to investigate the ways the control boundary evolves between the objective (i.e., a number of decisions at a given point in time) and its realisation (i.e., considering the operational limits of realising any decision being made). Future networks need enhanced flexibility in assembling service offerings at runtime by the (virtual) network operator. Therefore, the softwarisation of telecoms is crucial and future research efforts need to converge to consolidate the good results achieved by the scientific community so far.

To address this challenge in a ”not-always fully known environment”, we envision decision modules as software control elements, realising an adaptive control over the resources they manage. Changes in control objectives are reflected in the existing software, which, in turn, can establish additional software elements in order to react to changes in the control objectives. The enforcement, e.g., of flow handling or computation instalment, is realised by the resource owner, possibly self-constrained by objectives imposed by the physical

infrastructure and its operational environment. The overall system will nevertheless need to reliably fulfil the service requirements.

Control of Various Virtualisation Layers

In computing there has long existed a tension between isolation and robustness on one side, and performance on the other side. One example of this tension is the famous debate between Linus Torvalds – advocating for a monolithic kernel – and Andrew Tanenbaum – advocating for a micro-kernel approach.

Starting from these roots, the advances in recent years on cloud-based services and NFV architecture and platforms have moved the community focus on the performances and flexibility offered by compute virtualisation technologies (e.g., Xen, KVM, VMWare, Hyper-V, etc.) when running guests based on general-purpose operating systems. The virtualisation of network and non-network functions has many potential options today, starting from general purpose virtual machines based on Windows, Linux or FreeBSD to the more recent lightweight virtualisation technologies including containers, unikernels (i.e. specialised VMs with single-application function) and minimalistic distributions of general-purpose OSes (OS tinyfication).

The main challenges to solve in this area of research can be briefly summarised in:

- **Performance Area (SLA)**, i.e. design platforms that can support the specification, realisation and runtime adaptation of different performance metrics, taking into account workload type, size of the workload, set of virtual machines sharing the underlying infrastructure, etc. More in details, main challenges exist in:
    - *Optimising the VNF provisioning time* (including up/down/update) which includes the time to implement resource allocation at hypervisors, select the guest and host OS flavours, configure the need for hardware and software accelerators, etc.
    - *Increasing the runtime performance* (achievable throughput, line rate speed, maximum concurrent sessions that can be maintained, number of new sessions that can be added per second) for each virtual function.
    - *Efficient inter-VM networking solutions* that can support the achievement of the required performances.
- **Continuity, Elasticity and Portability**, i.e. service continuity of virtual functions can be interrupted due to several factors like upgrade progress, underlying hardware failure, unavailability of virtualised resources, software failures, etc. Some of the challenges to study and address include:
    - *Coupling of virtual functions and applications with the underlying virtualisation infrastructure*, e.g. in terms of hypervisor type support, new hardware capabilities (e.g. packets acceleration), etc.
    - *Service continuity*, i.e. to achieve efficient high availability of the virtualisation technologies in use for seamless (with zero impact) or non-seamless continuity of the supported services.
    - *Service elasticity*, i.e. to extend and rapidly adapt service coverage.
    - *Scalable continuous monitoring*, i.e. to collect and process state information from various sources and trigger new service optimisation strategies of the intelligent Smart Networks.
- **Security** of the virtual functions and applications and their states. Challenges in this area mainly concentrate around the support of multi-tenancy and secure slicing of the infrastructure resources, for which it is critical to
    - *Guarantee complete isolation* across resource entities (hardware units, hypervisor, virtual networks, etc.) and provide secure access between VM/container and host interface, VM-VM or container-to-container communication, etc.
    - *Quarantine* compromised entities while ensuring service continuity for other resources.

    - *Securely recover from runtime vulnerabilities or attacks* and restore the network functions to an operational state.
- **Management** of the operational aspects of the virtual functions and applications to implement centralised control and visibility, proactive management for dynamic resource allocation, auto-restart in HA model, audit trails, patch management, etc.

2.2 Integrated Fixed-Mobile Architecture

In the mobile and wireless networks, the demand for higher capacity can be achieved by new radio access technologies, diversification of the latter and by spatial multiplex, i.e. by reducing the cell radius and allowing frequency reuse. The latter leads to the well-known micro, pico and femto cell designs. The cell densification has been the main contributor to the tremendous increase of the wireless system capacity over the last generations. However, it also leads to the problem of backhauling, i.e. to the question of connectivity of femto/pico cell base stations to e.g. the mobile core network, the Internet or to other backend services. In principle, the backhaul provisioning has three alternatives, and in practice we expect all of them to be employed:

1. Major case: use optical and copper connectivity, if and as available (e.g. FTTH, xDSL). This is by far the major case for small cell connectivity, as it supports the full frequency reuse in the wireless cells.
2. Typical case for macro cells in rural areas: use a dedicated directed microwave links from tower to tower, until some dispatch centre to a fixed network is reached.
3. Specifically, at the edge: use macro-cell's fronthaul, i.e. another cell's mobile service, for backhauling. Alas, this approach strongly limits frequency reuse, as many small cells have to share the available bandwidth of the macro-cell, which also cannot be used within the small cells.
4. Rural areas and at the open water: use other possibilities, such as satellite connectivity or HAPS (high altitude pseudo satellites or high-altitude platform stations). This is currently a rare possibility and for the time being mainly in the study and experimental phase, but with the possible future use of LEO satellite based network services and the experiments of some Internet giants in the HAPS space, it could become a viable alternative. Note that the data from the satellite is typically sent back to a ground dispatch station, from where on fixed networking is used.

Since the technologies used for backhauling majorly are fixed network technologies, the backhaul problem solution requires a common management and control of both fixed and mobile networks.

Seen from the other direction, as was initiated by e.g. ETSI TISPAN, one could also consider opening the subscriber management and subscriber service systems, which usually exist as parts of mobile systems (e.g. EPC + IMS) and of network access control systems (e.g. NAS/DIAMETER), for any access network technology. In this view, the services made available by the operator over his Core network, should be transparently, i.e. with adequate or equivalent service quality guarantees, made available to the subscribers regardless of the access technology that the subscriber currently uses. This is usually referred to as *user-centric networking*, as the focus of service provisioning here always lies on the user profile, and never on any specific technology. Instead, any employed technology, be it in the access, the core or the service realisation, must be adjusted and operationally parameterised in a way to provide the same service with equivalent functional and extra-functional properties, as long as possible.

Trying to achieve either of both points above defines the question of *fixed-mobile convergence*, a slow yet unstoppable trend to fuse previously radically different fixed and wireless networking. Indeed, while the 2nd Generation mobile systems (e.g. GSM) mostly employed typical telecom technologies, and fixed networks were mostly used for data access using TCP/IP networking, this gradually changed with 3rd Generation capable (UMTS) of

providing TCP/IP services to the subscribers and the 4th Generation (LTE/SAE) employing the All-IP approach in its own realisation, where the TCP/IP networking represents the common technological base for system realisation and service provision. Besides the All-IP approach, mobile systems have also integrated other, previously fixed network technologies such as network security protocols and methods (e.g. TLS, IPSec, EAP, specifically 802.1X/EAP model, DIAMETER). Nevertheless, even 4G did not follow the early proposals for a user-centric network and remained an operator-centric technology. While it is therefore in principle possible to provide similar services e.g. over xDSL, WiFi and 4G (e.g. Internet access or voice call services) including using the same user identity and credentials, today, the service quality, attributes and features are still rather different. Mobile access, security guarantees, quality of service attributes e.g. are not homogeneously available and sometimes de facto restricted to the tightly integrated, native 3GPP RAN (LTE).

While the technological base therefore is already mostly shared, further convergence of mobile and fixed access would require common operational control of these heterogeneous infrastructures. Indeed, a dynamic and flexible provision of the service to the user according to the user profile would require capability to control the heterogeneous infrastructure elements and systems, so as to operationally change their runtime attributes.

2.3 Slicing and Orchestrators

Slicing promises a network-spanning (i.e., end-to-end), user-driven customisation of the basic, currently often invisible, network primitives. This translates to several new problem spaces, currently unaddressed, underestimated or completely overlooked in both the industry and academia.

Network Slicing versus Network Capacity Planning

As network slicing promises a sheer endless customisation of network-spread functionality, it becomes difficult to plan the capacity of network infrastructures in the same way as today. Whereas operators currently use their combined empirical knowledge regarding both infrastructure and the expected service (and its prices), network slicing turns this principle upside-down: while the infrastructure operator remains ignorant of or neutral to the service, the slice owner is expected to translate the service to capacity requirements, an exercise that lacks a reliable general methodology. Incapable of correctly translating service to capacity requirements, slice owners are likely to expect a cloud-like operation model: start small, expand or reduce contracts as you go. The elasticity of the slice therefore is a central requirement. This fact together with the required radical reduction of the service creation time (from 90 days to 90 minutes, as per 5G PPP KPIs) underlines the upcoming shift from planning of the infrastructure to continuous (and likely dynamically adapting) runtime operations on the latter. In simple terms, the network planning is misaligned to network slicing, as, under constraints of the physical deployment, it operates within a completely different time frame. What matters for slicing is runtime (continuous, real-time, hot) management and control. If network slicing wants to succeed in the above sense, the provided technologies must embrace this change and provide mechanisms and practices that feed runtime control over a longer timeframe back into the planning and investment cycle for network infrastructure.

Independently of scale, slicing renders the infrastructure usage and occupation much more diverse and more dynamic. This emphasises the requirement for continuous operation of the latter (real-time management or control). It means that infrastructure control and management are required to handle the dynamics in a new, currently unsupported manner. This includes handling node and service element loads, departures, additions, errors and the like.

Runtime management and control ultimately still drives the longer-term planning that we can see today in networks. In staying with our cloud analogy, the longer-term demand and supply pattern emerging from the many tenants of a data centre still drives the planning and therefore investment patterns for sufficient build-out of the cloud. Similar feedback must exist for slicing-

based infrastructure albeit situated in a many point-of-presence nature of resources, utilised over a possibly huge area of requirements on those resources.

Slicing Requires Novel Resource Control Means

With network slicing, the decoupling of the platform delivering the service and the service elements reaches a new level. While IP networking has decoupled services from network infrastructure by putting all services on the same technological foundation (the TCP/IP suite) and by pushing the service logic to the edge, slicing brings additional degrees of freedom in flow processing and combines edge and network in one logical entity; it is possible to have different flow processing logics active at the same time within the same physical infrastructure, usually in the form of software elements (different configurations, different active modules) deployed on top of more generically capable hardware resources. Whereas today's networks rely on specific flow processing machines (e.g., routers or switches), whose flow processing capabilities are intrinsically linked to the purpose of the device, network slicing breaks this barrier by allowing to define different flow treatments on the same network node and by concurrently reusing the given link for flows of different slices requiring different assurances. This immediately raises a completely new question of a service-independent control of resources per se: as all infrastructure capacities are, in principle, slice-independent, we need a means to make sure that the execution of a slice-specific element on an infrastructure element is durably possible. In other words, while a router routes and a switch switches, and there is hardly anything to verify about that; slicing will require to tell a node to route, while this node is possibly also doing other tasks at the same time. It must be verified that it routes correctly over time despite the task overlap. Classically, control was always something integrated in the service logic (on the respective OSI layer or abstraction level) and directly projected to resources dedicated to delivering (a part of) that service, as the existence and function of the node used to be the same, so was their control. For example, network service errors can be traced to network element errors, by using network service control means. With slicing however, this changes drastically: we need to understand resource control as a new, paramount domain: because a node or link generally does not have a single predefined function (see subsection 8.2), there is a new requirement to allocate, monitor, migrate and execute/run several service elements on a shared service-agnostic infrastructure node. An additional complexity arises from the insight that a slice function does not generally translate to a single infrastructure element, but it can be very well sustained by capacities distributed over the infrastructure. Due to scalability reasons, most network functions rely on hugely distributed realisations, causing the allocation, extension, monitoring or migration of a network function much more challenging than the question of copying a software state from one node to another.

Challenges on resource control in Network Slicing include:

- Resource control emerges as an initial glue that first allows operators to slice their infrastructures, i.e. as an initial new service that allows to allocate, monitor and remove service elements from sets of nodes and links. To avoid vendor lock-in and allow truly end-to-end slicing, it is exactly this glue that requires standardisation, and not any domain-specific management interface.
- Resource control must be able to reach out to all infrastructure resources and be capable to check the states and operations of all slice-specific elements on those resources. Besides, the realisation of the resource control itself should follow the insights from above, i.e. it must be distributed over all nodes and support elasticity of itself (compare with subsection 8.2).
- Because of the novel degree of decoupling service elements from the infrastructure, the central problem of slicing is not to make a blueprint, but to be able to execute any requested blueprint on top of a shared, distributed infrastructure composed of different capacities, occupied by loads from other executed slices. This distributed execution

under contention is extremely challenging and currently can only be solved on very small scales.

Slicing Efficiency is a Question of Scheduling

A different problem space, intimately related to the efficiency of slicing and the Total Cost of Ownership KPI, opens once one delves into the resource allocation question. Given a blueprint, one must find suitable resources in the infrastructure and make a reasonable long-term allocation of the blueprint on the selected resources (lifecycle as per slice lifecycle). This topic has received a considerable attention under the academic title of ”virtual network embedding”. As a result, both simplified greedy solutions and optimised heuristics (with tunable sub-optimality bounds) to this problem are available. However, the overall resource allocation problem of network slicing is twofold, and the second part is unsolved. This second problem is related to the question of elasticity of slices. More generally, to achieve slice properties not readily provided in the serving infrastructure (e.g., elasticity, but also availability, resilience, latency guarantees, etc.), slice embedding will be usually broader than the purely functional requirements of the blueprint. Therefore, for every entering flow, a simplified, yet more dynamic and online question of the resource allocation problem will arise: which of the suitable function-equivalent infrastructure resources should be involved into the treatment of that flow? This problem is one of job scheduling in the prepared infrastructure. (Note that this cannot be done within the slice, if the infrastructure owner promises (and sells) the extra-functional properties of the allocated slice; in other words, such provisioning will be done in the infrastructure, transparently to the slice owner).

The answer to the question of runtime networked job scheduling is paramount to address the Total Cost of Ownership (TCO) KPI, as a solution to this problem would allow to overprovision slices, without the need to overprovision the underlying infrastructure. The runtime networked job scheduling therefore is the answer to the elastic and dynamic network slicing questions, currently unsolved. Moreover, if an efficient solution to this problem can be found, then network slices can and, for efficiency reasons, should be implemented as dynamic scheduling.

Challenges in this area can be summarised in the following:

- The question of dynamic job scheduling in network slicing is paramount, as it permits both to provide superior extra functional properties of the supported slices and to lower the Total Cost of Ownership. Indeed, the TCO of a slicing implementation using only fixed-quota assignments (meaning that the sum of the resources consumed by all slice instances will define the necessary infrastructure resource footprint) would be horrible, comparable to hardware slicing. The resource assignment problem is a quest for a more efficient infrastructure sharing, including computing, networking and energy resources.
- The answer to the job scheduling in large networked systems requires a lot of fundamental research, to leverage the existing solutions from data centre research and to make them scalable and network-efficient. Because of the fundamental locks known from distributed systems research, the major goal should not be full optimality, but rather efficiency increase: given the size of the infrastructure, 1 % efficiency increase might translate in hundreds of millions of Euros.
- The elasticity of slicing has to increase towards subscriber level and even application level. Hence, an application should use different slices with inter-slice handover during its session in order to best utilise the network as well as provide superior quality of service with respect to slice offerings.

2.4 Evolution of NFV/SDN and AI/ML-based Network Control

Within the long-term target-picture, there are no network ‘elements’ anymore, but rather virtualised functions, realised by pure software for which the reliable controllability is key.

Network operators will aim to perform in fully automated manner:

- Instantiation of a complete end-to-end network that includes the RAN, mobile core, transport network, as well as the Data Network. This network may be logically separate and/or isolated for certain aspects like services, users, etc.
- Network services may be incrementally deployed in the operator's network in logically separated and/or isolated manner from the other already deployed services.
- Network services may be deployed and provided to other operators and/or service providers when requested, via open interfaces. This way, other operators and/or service providers can re-sell/extend the provided 5G network services.
- Fast lifecycle management (LCM) of the network automatically triggered based on vendor-independent FCAPS management.
- Plug & Play of new components into a live production network.
- Termination of one or more network service(s), or 5G networks as a whole.

The research challenge in this area is to develop a future network with ***Full Automation***, which reduces and tries to eliminate any human intervention, by leveraging on powerful AI/ML systems to realise a cognitive network.

There is a challenge that AI/ML is seamlessly applied to the network control, to run automated operations of network functions, network slices, transport networks, in an end-to-end scope.

Also, a particularly critical challenge is the possibility to implement predictive behaviours on the network, to make available a network control intelligence capable to prevent the impact of failures, the usage load, etc. and fast adapt network configuration to be always available at the target performance levels requested by the applications.

The call for AI/ML-based network control as a way to implement the concept of **fully automated Smart Networks is a must of future communication networks more than a nice-to-have**: in fact, the scale of deployments made possible by the function virtualisation, the extreme split in micro-/atomic-functions and the proliferation of more and more functions at the edge are creating network deployments of unprecedented complexity, impossible to manage and control with the actual human-driven decision support tools.

2.5 Media Access Control

Wireless technology has enormous potential to change the way we live, work, and play over the next several decades. Future wireless networks will support 100 Gb/s communication rates between people, devices, and the ”Internet of Things”, with high reliability and uniform coverage indoors and outdoors.

The shortage of frequency spectrum to support such systems will be alleviated by advances in massive MIMO, mmWave (and perhaps nmWave) and small cell technologies. Caching and computation at network edges (e.g., in base stations and access points) will reduce latency and increase energy-efficiency, enabling real-time data analysis, control and automation. Wireless technology will also enable smart and energy-efficient homes and buildings, automated highways and skyways, and in-body networks for monitoring, analysis and treatment of medical conditions.

Breakthrough energy-efficiency architectures, algorithms and hardware will allow wireless networks to be powered by tiny batteries, energy-harvesting, or over-the-air power transfer. Finally, new communication systems based on biology and chemistry to encode bits will enable a wide range of new micro and macroscale applications. In short, key areas for the future are including: Utilising more spectrum (mmWave) and (Massive) MIMO; rethinking cellular system design, with increased hierarchies and IP support; software-defined wireless networking; and ”smarter” and more agile (cognitive) radios for energy-constrained networks plus including energy harvesting designs. Furthermore, AI/ML will become an essential technique to better exploit the frequency spectrum in use.

The main challenges to be addressed in this field can be summed up into the following top priorities:

- Consolidate the initial research results done in the area of joint use of diverse spectrum, investigating new generation of WiFi in the THz band (e.g. for sensing and communications).
- Strengthen the maturity of Visible Light Communications, which are becoming more promising due to LED lights transmitters available in many places (cars, traffic lights, etc.).
- Model the communication channels in a more comprehensive approach, and address also the uplink, together with the downlink, making use of AI/ML techniques to learn environment and movement patterns and automatically adjust the channel performances
- Optimise network features of the radio and fixed part collectively, in order to efficiently support device to device communications, in-band backhauling, multi-casting/broadcasting to reduce interference via advanced network coding, improve scheduling with proactive scheduling, manage access to the medium to consider context awareness.

2.6 Network-Based Localisation

Location-Based Services (LBS) and Real-Time Location Systems (RTLS) market is significantly growing, stimulated by the various networked applications offered to the users by the current networks. Nevertheless, localisation aspects (and especially the business exploitation of both localisation information and derived knowledge) have never been considered thoroughly in the network evolution, but have rather been addressed as a valuable, but still aside, add-on to the main communication services that networks are called to provide.

We call for the ambitious challenge of **realising Smart Networks to incorporate by design technologies and APIs to enable location/context-based services and powerful business analytics** on top of them as a way to fully respond to the needs of the vertical applications implementing new personalised services for the end-users.

Key challenges in the area of network based localisation include:

- **Terminal localisation with sub-meter accuracy**. This precision could be required by applications like personal security, infrastructural monitoring (e.g. structural monitoring of buildings, roads, bridges, etc.), etc. it is critical to consolidate the *integration of localisation technologies designed into specific subsystems* (Wi-Fi, GNSS, Bluetooth, visible light, inertial, etc.) and to enable the *collection, interfacing, and fusion of location-based information coming from heterogeneous technologies and subsystems*.
- **Device-free localisation.** The challenge here is to properly design and use a network of sensor radars which are coupled with functions of spatiotemporal analysis of signals backscattered by single and multiple device-free targets (persons, things, and vehicles) and can allow to derive the position information (localisation and tracking) of the target. The work to be done is not only in the integration and processing of the various signals, but also in *waveform design to properly obtain localisation* accuracy in a given context of propagation, bandwidth, and application. It would be useful to consider mmWave technology to assess the achievable gain in tracking accuracy, as well as to develop *innovative algorithms for single and multiple target tracking* which make use of signals of opportunity, both radio (such as LTE, DVB, and DAB) and non-radio (i.e. acoustic and visible lights), massive MIMO, etc.
- **Spatiotemporal analytics**. Analytics are key to provide verticals with elaborate knowledge learned from localisation data. Such analytics will primarily leverage basic spatiotemporal features of individuals or crowds such as presence, position, heading, velocity or trajectory. It is needed to develop analytics that take full advantage of the localisation accuracy and precision to derive useful information on the physical

behaviour of individuals and connected objects to support business intelligence, smart intuitive buildings, intelligent transportation, smart management of the parking spaces, or network demand prediction.

- **Multi-modal Analytics.** In many domains where localisation is a driving technology, the individuals to be localised are associated with a multitude of data (e.g., accelerometer data, mobile application usage, imaging information activity patterns from the network such as HTTP(S) request sequences, etc.). The availability of additional data sources is an important opportunity to complement and enrich analytics, developing more comprehensive AI/ML models. There is the need to develop novel AI/ML models to combine the various data sources, build efficient representation models, and thus discover/detect collective anomalies. Hierarchical architectures for these analytics efficiently splitting the data engines between the core and edge of the network are key to guarantee low-latency, computationally efficient and scalable analytics processes.

3. Radio Technology and Signal Processing

3.1 Spectrum Re-farming and Reutilisation

Allocated frequency spectrum is one of the main factors that determines the system capacity. But radio spectrum is a scarce resource. Especially the lower frequency bands are precious and tightly regulated. In order to satisfy the high bandwidth demands of upcoming generations of mobile systems, it is crucial to reutilise the existing spectrum resources. While the traditional approach allocates a dedicated spectrum to each radio access technology (RAT), spectrum reutilisation between RATs offers a more efficient utilisation of resources and greater flexibility, e.g., for load-balancing. Spectrum reutilisation, also known as spectrum sharing, can be applied to licensed but also unlicensed bands.

A straightforward approach to inter-RAT spectrum reutilisation is *spectrum re-farming*. Re-farming performs static allocation of spectrum resources to different RATs. This method was already used to clear GSM spectrum to make it available for 3G. Because of its static nature, it has poor spectrum utilisation.

A more efficient utilisation is achieved by dynamic inter-RAT resource scheduling with optimised multi-RAT handover and interference coordination. Preferably, this is based on a centralised multi-RAT radio resource management. The signalling overhead can be reduced by decentralised strategies.

For the joint utilisation of licensed and unlicensed spectrum, adaptive strategies are required such as cognitive radio concepts, in which co-existence with existing (e.g. analogue) services is studied.

Spectrum reutilisation is supported by multi-RAT connectivity, which allows the UE (User Equipment) to choose the best RAT depending on the link qualities. This added diversity not only increases the performance due to better spectrum utilisation, it also makes the network more robust and resilient towards shadowing effects, hence improving the reliability and availability.

Future networks will support different services, enabled by network slicing based on a multi-RAT radio access. Multi-RAT connectivity can also make flexible use of licensed and unlicensed bands. E.g., data and voice traffic can be offloaded to WiFi or LTE small cells operating in unlicensed bands as an enhanced mobility concept. Hence, utilising unlicensed bands is important and technologies to bring the quality to the licensed spectrum level are open to study. This not only increases the overall throughput but also enables low latency.

3.2 Millimetre Waves

Millimetre wave (mmWave) have attracted large research interest in recent years due to the huge available bandwidth required to fulfil the today's traffic demand. This is reflected in WLAN and WPAN standards: in the license free 60 GHz band, the IEEE802.11ad WLAN standard provides rates up to 8 Gbps and the upcoming IEEE802.11ay WLAN standard will provide rates up to 30 Gbps. The fifth generation (5G) wireless networks aim to use mmWave in mobile networks, where the transmitter/receiver nodes may be moving, channels may have a complicated structure, and the coordination among multiple nodes is difficult [9]. This year's Winter Olympics in Korea already provided first glimpse at the 5G services powered by Korea Telecom with support from global equipment makers. This show case included a 28 GHz mmWave backhaul network for moving hotspots, such as buses. Additionally, the mmWave band in combination with mobile edge computing (MEC) is highly suitable for on-demand content (multi-media) delivery services, hence enabling the enhanced mobile broadband (eMBB). This combination of mmWave and MEC is the only way to satisfy both extreme communications requirements: ultra-high speed and low latency. Beyond 2020, MEC is expected to enable automated driving using mmWave based V2X/V2V links. This requires, however, cooperative perception and the exchange of HD dynamic map information between

vehicles and radio units, to enhance the visibility area. The automated driving use case can be considered as the most important application of mmWave and MEC, which requires both ultra-high speed and low latency [9].

Beyond 5G, it is expected that the data traffic due to mobile nodes (smart phone and tablets) will be more than 100 petabytes per month by 2023 [10], which is 10 times of the traffic in 2017. Alone in Western Europe, the data traffic is expected to be as high as 12 petabytes per month, which amounts to 56 terabytes per person per month [11], hence, offering a huge potential to exploit mmWave bands even in Terahertz range (not considered so far by 5G). It is projected in [12] that the volume of the traffic generated from smart phones will be 86 % of the global data traffic by 2021 and among this more than 50 % data will be offloaded to the fixed networks by means of Wi-Fi devices and small cells each month, while remaining will be covered by the cellular networks. Based on this projected increase in demand for data rates in short range communications, the multi-Gigabit/s WLAN standard provides a peak data rate of 8 Gbps (IEEE802.11ad) or 38 Gbps (upcoming IEEE802.11ay). Anticipating the requirements for short range communication beyond the year 2021, a very high data rate new Wi-Fi is inevitable. In order to achieve high data rates, one would require a large amount of contiguous bandwidth suitable for communications over short ranges, that is to be found beyond 100 GHz, for example around 140 GHz. The use of these frequency bands provides an excellent opportunity, since many antennas can be packed in a small area to direct a beam to the intended user.

An important business case for mmWave is in so-called '*smart factories – Industry 4.0*'. Due to its ability for spectrum re-use that enables multi connectivity for high reliability, mmWave provides a complementary solution to low frequencies. Additionally, due to the high penetration losses, mmWave is inherently more secure against eaves-dropper and is a suitable candidate for industry environments. An additional feature of mmWave is sensing/positioning with high accuracy. This allows detection with higher spatial and velocity resolution that is suitable for both V2X and industry automation scenarios.

Small cells are to play a key role to cope with the increasing traffic demands in mobile network. These small cells connect to the core network via wired or wireless backhaul links. The dense deployment of small cells and a variety of services offered by the RAN having diverse requirements on throughput, latency and reliability, poses new challenges on backhaul links. One way to address these challenges is self-backhauling using mmWave, i.e., the access and backhaul share the same wireless channel. 3GPP stage 1 in its Release 15 [13] outlines the requirements for the self-backhauling in 5G networks. Among these requirements are the flexible partitioning of resources, autonomous configuration, multi-hop wireless connectivity, topology adaptation, and redundant connectivity.

One of the main challenges will be to manage the different network features introduced in 5G and developed beyond the first release and optimise them collectively. Diverse network components need to be integrated, such as D2D, self-backhauling, multi-casting/broadcasting. While these technologies will be already available in 5G, the new challenge consists of extending them by advanced massive MIMO techniques, which are dynamically coordinated, considering interference and mobility.

3.3 Optical Wireless Communication

Despite the tremendous improvements due to the small cell concept and the allocation of new radio frequency (RF) spectrum, the continued exponential growth in mobile traffic [14] means that it will be inevitable that the RF part of the electromagnetic spectrum will not be sufficient to be able to drive the 4th industry revolution which is centred around data-driven economies and data-driven societies [15].

It is, therefore, natural to consider the infrared and visible light spectrum both of which are part of the electromagnetic spectrum for future terrestrial wireless systems. In fact, wireless systems using these parts of the electromagnetic spectrum could be classified as mmWave

wireless communications system in relation to Section 3.2. Light based wireless communication systems will not be in competition with RF communications, but instead these systems follow a trend that has been witnessed in cellular communications by inspecting all the generations developed during the last 30 years. Light based wireless communications simply adds new capacity – the available spectrum is 2600 times larger than the entire RF spectrum.

An important advantage is that off-the-shelf optical devices can be used to harness these unregulated and free transmission resources. By using advanced devices, lab demonstrations showed 8 Gbps from single light emitting devices (LEDs) and 17.6 Gbps using laser diodes (LEDs) [16]. Recently, a record of received data rates of 500 Mbps by using a single solar cell has been demonstrated. The use of these types of 'data' detectors has the appealing advantage of achieving simultaneous energy harvesting and high-speed data communication – a feature that will become ever more important in mobile machine-type communication (MTC) [17].

Networked and cellular wireless networks which are based on visible light communication (VLC) are referred to as LiFi (Light Fidelity) [18]. LiFi enables, bi-directional networked communication including multiuser access and handover (please refer to Figure 2 for a taxonomy of different light communication approaches). The blue arrow in Figure 2 indicates that the major research efforts in the last 15 years has been focused on enhancing link data rates of intensity modulated (IM) / direction detection (DD) optical wireless communication systems. With the advent of LiFi the research focus has begun to shift to challenges related to networking issues using light.

As in RF networks, there are issues surrounding interference management and interference mitigation in LiFi networks. However, since, for example, there is no multipath fading because the detector sizes are much larger than the wavelength, techniques developed for RF systems may only be sub-optimum. There are also fundamental differences as a result of IM/DD, in that signals can only be positive and real-valued. Consequently, new LiFi-bespoke wireless networking methods must be developed. Moreover, because light can be confined spatially by using very simple and inexpensive optical components, interference can be controlled much easier. This feature also allows step-change improvements of the small cell concept as single cells might cover sub-m^2 areas.

Furthermore, due to the extremely small wavelength, the active detector sizes are very small, and massive MIMO structures can be implemented at chip-level. Edinburgh University, for example, has developed a massive MIMO LiFi chip composed of 49 avalanche photodiodes (APDs) on CMOS which, as a major breakthrough, requires only very low-voltage negative biases while achieving at least 10 dB APD gains. The size of the 49-dector die was merely 2.8 mm x 2.8 mm. This property can be used to develop unique and LiFi-bespoke MIMO systems, networked MIMO approaches, and new angular diversity techniques in conjunction with low computational complexity cooperative multipoint (CoMP) systems. Diversity techniques in LiFi systems are especially powerful to combat random blockages that naturally occur in a mobile scenario.

Moreover, the spatial confinement of signals in LiFi enables the development of radically new physical layer security concepts.

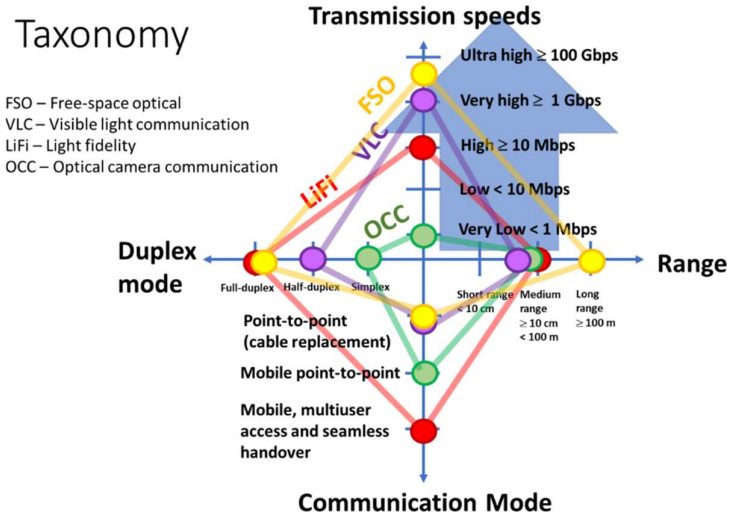

Figure 2 A taxonomy of emerging light communication technologies

Free-space optical (FSO) is point-to-point long range optical wireless communications with target data rates of tens of Gbps primarily using laser diodes and coherent transmission. VLC has been used in the context of line-of-sight high-speed point-to-point communication, primarily using LEDs in conjunction with IM/DD. VLC systems are usually designed for ranges less than 100 m, as well as for bi-directional communication. Optical camera communication (OCC) in contrast is simplex communication using embedded CMOS camera sensors as data detectors. Due to the use of CMOS sensors, the achievable data rates are well below 1 Mbps. OCC is primarily used for indoor navigation, asset tracking and positioning. These applications assume some user mobility.

LiFi is currently being standardised in a Task Group within IEEE 802.11. The new LiFi standard has received the following reference: IEEE 802.11bb. Similarly, VLC is being standardised in IEEE 802.15.13, while OCC has been standardised in IEEE 802.15.7r1.

Convergence with 3GPP access: LiFi is bi-directional communication. Due to the abundance of optical spectrum, typically the visible spectrum is used for the downlink by piggy-backing on lighting systems, while the infrared spectrum is used for uplink transmission. The simplicity of IM/DD in conjunction with advanced layered modulation techniques [19] enable highly energy-efficient transmission systems suitable for the uplink. The high energy-efficient and spectrum efficient modulation techniques are based on multicarrier approaches. Therefore, it could be argued that a tight interaction between radio and optical components should be considered at the level of baseband processing. Since OFDM transmission (e.g. 5G waveforms) is feasible on a free-space IM/DD optical link, it is definitely worth investigating the use of the same basic waveform and protocol stack for radio and LiFi systems. This would allow for a common baseband processing platform in both the small-cell transmitters and terminal receivers. Moreover, the 3GPP access-layer protocols are perfectly adapted to the use of downlink-only component carriers.

3.4 Terahertz Communications

Wireless data rates have doubled every eighteen months for the last three decades. Following this trend, Terabit-per-second (Tbps) links are expected to become a reality within the next five years. While mmWave communications are a step in the right direction, the total consecutive available bandwidth in such systems is less than 10 GHz. Consequently, supporting Tbps would require a physical layer efficiency of 100 bit/s/Hz, which is several times higher than the state of the art.

In this context, **Terahertz-band (0.1–10 THz) communication** is envisioned as a key technology to satisfy the need for much higher wireless data rates [20]. This frequency band, which lies in between mmWave and the far infrared, supports huge transmission bandwidths: from almost 10 THz for distances below one meter, to multiple transmission windows, each tens to hundreds of GHz wide, for distances beyond several tens of meters. However, this very large bandwidth comes at the cost of a very high propagation loss. Moreover, for many years, the lack of efficient ways to generate and detect THz signals has hampered the use of the THz-band in practical communication systems.

To date, different technologies are being considered to close the so-called THz gap. In an *electronic* approach, the limits of silicon CMOS technology [21], silicon-germanium BiCMOS technology [22], and III-V semiconductor HEMT, mHEMT, HBT and Schottky diode technologies [23] are being pushed to reach the 1 THz mark. In a *photonics* approach, uni-travelling carrier photodiodes [24], photoconductive antennas [25], optical downconversion systems [26] or, more recently, quantum cascade lasers [27] are being investigated for high-power THz systems. In both approaches, fundamental device limits are being reached, as the frequency is ”too high” for electronic devices and the photon energy is ”too low” for photonic devices to efficiently operate at *true* THz frequencies.

More recently, the use of **graphene to develop novel plasmonic devices** for THz communications has been proposed. Graphene is a two-dimensional (2D) carbon-based material that has excellent electrical conductivity, which makes it very well suited for propagating extremely-high-frequency electrical signals [28]. Moreover, graphene supports the propagation of THz Surface Plasmon Polariton (SPP) waves at room temperature. SPP waves are surfaced-confined electromagnetic waves generated by the global oscillation of electrons. By leveraging the properties of graphene, nano-transceivers [29] and [30] and nano-antennas [31] and [32] have been proposed and are being developed. These devices are intrinsically small, efficiently operate at THz frequencies, and can support very large modulation bandwidths. Moreover, graphene is ”just the first” of a new generation of 2D materials (such as MoS2 or Hb-N), which can be stacked to create new types of devices and leverage new physics.

In parallel to the development of new device technologies, there is a need to understand and model the **THz-band channel**. In the case of line-of-sight (LoS) propagation [33], the main phenomena affecting the propagation of THz waves are the spreading loss and the molecular absorption loss. The *spreading loss* accounts for the attenuation due to expansion of the wave as it propagates through the medium and is common to any wireless communication system. The *molecular absorption loss* accounts for the attenuation that a propagating wave suffers because a fraction of its energy is converted in vibrational kinetic energy in molecules (especially water vapor). In the case of non-line-of-sight (NLoS) propagation [34], in addition to the two aforementioned phenomena, high reflection loss, diffused scattering and diffraction by obstacles need to be captured. Ultimately, stochastic multi-path channel models are needed to statistically characterise the channel.

In light of the capabilities of THz devices and the peculiarities of the THz-band channel, there is a need to develop new communication algorithms and networking protocols, tailored to THz communication systems. At the physical layer, new types of modulations are needed. For short-range communications (below one meter), the use of impulse-radio-like communication based on the transmission of one-hundred-femtosecond-long pulses following an on-off keying modulation spread in time has been proposed [35]. Such very short pulses are already at the basis of many THz sensing systems and can be generated and detected with current technologies. For longer communication distances, new **dynamic bandwidth modulations** [36] are needed to not only overcome but even leverage the unique distance-dependent bandwidth created by molecular absorption.

Independently of the modulation, and similar to any wired or wireless Tbps communication system, physical-layer synchronisation (both in time, frequency and phase) becomes a major

challenge. The front-end non-idealities, e.g. non-linearity and phase noise, can severely impact the achievable throughput. Going to Tbps throughputs implies increasing the bandwidth to tens of GHz. This is another challenge for implementations for two reasons: first, ADCs and DACs in the tens of Gsamples/s are needed; second very wideband analogue baseband circuits are needed (half the RF bandwidth). The lack of digital-to-analogue and analogue-to-digital converters (DACs and ADS, respectively) able to handle multi-GHz bandwidth signals, limits the application of traditional digital signal processing and motivates the research and development of new mixed (digital and analogue) techniques where some traditionally digital functions such as synchronisation or equalisation can be moved to the analogue domain. Additional challenges include new channel coding strategies, which leverage the uniqueness of the THz-channel, or physical layer security schemes for THz-signals. Very generally, efficiencies become dominant bottlenecks: at 1 Tbps, an efficiency of 1 pJ/bit (impossible today if we consider the whole PHY) translates into 1 Watt of power consumption; similar considerations about implementation efficiencies in silicon technology (area efficiency (bit/s/mm^2) and power density (W/mm^2)) show huge challenges at Tbps rates.

Similarly, many challenges arise in the higher layers of the protocol stack. At the link layer, novel **MAC protocols** are required for THz-band communication networks, since classical solutions do not capture the peculiarities of this band. The very large available bandwidth almost eliminates the need for nodes to contend for the channel. The transmission of very short signals also minimises the chances for collisions. All these come at the cost of more complex synchronisation schemes between devices. Ideas to be explored for new MAC protocols include, among others, the development of receiver-initiated transmission schemes to ensure that the transmitter does not waste resources when the receiver is not available, especially when highly directional systems are used. Additional challenges also include packet size optimisation and adaptive error control strategies.

At the network layer, new **routing mechanisms** could be developed that take into account the availability of both classical active relaying nodes as well as novel passive dielectric mirrors, which can direct the signal towards its final destination. In addition, new routing metrics that consider the channel molecular composition and its impact on the available distance-dependent bandwidth need to be explored. At the transport layer, as wireless multi-Gbps and Tbps links become a reality, the aggregated traffic flowing through the network will dramatically increase. These will introduce many challenges at the transport layer regarding **congestion control** as well as end-to-end reliable transport. For example, we expect that a revision of the TCP congestion control window mechanism will be necessary to cope with the traffic dynamics of THz-band communication networks.

For the validation and refinement of the developed solutions, new **experimental platforms** and integrated testbeds will be needed. For the time being, these are mainly focused in the sub-THz windows (300 GHz, 650 GHz), but systems at *true* THz frequencies will be required. Finally, in parallel to all the scientific developments, work needs to be done towards regulation and standardisation of the THz-band [37].

3.5 Ultra-Massive MIMO

The grand challenge for mmWave, THz-band and optical communications is posed by the very high and frequency-selective path loss, which easily exceeds 100 dB for distances over just a few meters in the presence of LoS (line-of-sight) and becomes even worse in NLoS (non-line-of-sight) conditions. As a result, high-gain directional antennas are needed to communicate over distances beyond a few meters.

Similarly, as in lower frequency communication systems, antenna arrays can be utilised to implement MIMO communication systems, which are able to increase either the communication distance by means of beamforming, or the achievable data rates by means of spatial multiplexing. In the last few years, the concept of Massive MIMO has been introduced and heavily studied in the context of 5G systems [38], [39] and [40]. In such schemes, very

large antenna arrays with tens to hundreds of elements are utilised to increase the spectral efficiency to communicate over a large distance. This approach has been proved to be very useful for mmWave communication systems [41] and [42]. When moving to the THz-band, antennas become even smaller and more elements can be embedded in the same footprint. However, linearly increasing the number of antennas is not enough to overcome the much higher path loss in THz-band.

In this context, the concept of **Ultra-Massive (UM) MIMO communications,** enabled by very dense plasmonic nano-antenna arrays, has been recently introduced in [43] and [44]. Instead of relying on conventional metals, nanomaterials and metamaterials can be utilised to build plasmonic nano-antennas (see Section 3.4), which are much smaller than the wavelength corresponding to the frequency at which they are designed to operate. This property allows them to be integrated in very dense arrays with innovative architectures. For example, even when limiting the array footprint to 1 mm × 1 mm, a total of 1024 plasmonic nano-antennas designed to operate at 1 THz can be packed together, with an inter-element spacing of half plasmonic wavelength. Such plasmonic nano-antenna arrays can be utilised both at the transmitter and the receiver (1024×1024) to simultaneously overcome the spreading loss problem (by focusing the transmitted signal in space) and the molecular absorption loss problem (by focusing the spectrum of the transmitted signal in the absorption-free windows).

By properly feeding the antenna array elements, different operation modes can be adaptively generated. In **Ultra-Massive Beamforming**, all the nano-antennas are fed with the same plasmonic signal, as in conventional beamforming. This mode can effectively overcome the very high attenuation at mmWave, The THz-band and optical frequencies enhance the communication distance. Moreover, beamforming has the benefits to avoid co-channel interference while exploiting the angle diversity by steering the narrow beam dynamically to the targeted angle directions. In **Ultra-Massive Spatial Multiplexing** (UM-MIMO), physically or virtually grouped array elements can be assigned to communicate with an individual user. This mode uses multiple streams on a single carrier to increase the capacity per user, and can be most effective when radio links operate in a high SNR regime and are bandwidth-limited. This mode improves the network throughput by the means of spatial multiplexing, given that the UM-MIMO channel matrix is well-conditioned, or equivalently, provides the sufficient diversity and rank. Obviously, any combination in between UM Beamforming and UM Spatial Multiplexing is possible.

In addition, to maximise the utilisation of the mmWave- and THz-channel and enable the targeted Tbps-links, more than one spectral window could be utilised at the same time. In this direction, **Multi-band UM-MIMO** enables the simultaneous utilisation of different frequency bands by leveraging the electrically tunable frequency response of graphene-based plasmonic nano-antennas. By tuning (virtually) grouped sub-arrays to different frequencies, a single UM-MIMO system can simultaneously cover multiple transmission windows. One of the key advantages is that the multi-band approach allows the information to be processed over a much smaller bandwidth, thereby reducing overall design complexity as well as improving spectral flexibility. In this direction, advanced **space-time-frequency coding and modulation techniques** need to be developed for the UM-MIMO systems to exploit all of the spatial, temporal and frequency diversities, and hence, promise to yield remarkable performance improvements.

Besides the challenges related to the plasmonic nano-antenna array technology, the realisation of UM-MIMO communication requires the development of novel **accurate channel models** able to capture the impact of both plasmonic nano-antenna arrays in transmission and reception, as well as, the behaviour of a very large number of parallel THz-waves propagating in space. Existing MIMO or Massive MIMO channel models for lower frequency bands [38], [39], [40], [45], [46] and [47] cannot be utilised because they do not capture the peculiarities of the THz-band channel, including the frequency-selective absorption loss or the very high reflection loss. Similarly, the few THz Massive MIMO channel models developed to date [48] and [49] do not take into account the capabilities of plasmonic nano-antenna arrays,

such as the sub-wavelength size and separation, and the opportunities this brings. Therefore, a 3D UM-MIMO channel model for ultra-broadband communications is needed.

3.6 Non-orthogonal Carriers

Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM) has been adopted in several wireline and wireless standards such as ADSL, Wi-Fi, LTE, and has recently been proposed for 5G [50]. CP-OFDM divides the bandwidth into several orthogonal subcarriers. The orthogonality is preserved as long as the transmitters are synchronised to each other. Fine time and frequency synchronisation is then required to maintain the subcarrier orthogonality. However, strict synchronisation is limiting in certain scenarios. For example, sporadic access in Internet of Things (IoT) and Machine-Type Communications (MTC) requires relaxed synchronisation schemes, in order to limit the length of the signalling overhead [51]. Ideally, the massive number of devices could just transmit their messages asynchronously; being only coarsely synchronised [51]. This could also be advantageous for low-latency communications. However, in multi-user asynchronous access, the CP-OFDM subcarriers are no longer orthogonal, which introduces high inter-carrier interference [52]. Therefore, CP-OFDM is no longer viable in such scenarios.

Several waveforms, e.g. Filter Bank Multi Carrier (FBMC), Generalised Frequency Division Multiplexing (GFDM) and Universal Filtered Multi Carrier (UFMC) may be more suitable since their subcarriers are better localised in the frequency domain, and therefore limit the inter-carrier interference. A good frequency localisation may also be beneficial due to other reasons, e.g. sensitivity to phase noise in mmWave, required accuracy of frequency-synchronisation, etc.

The waveforms differ in whether they are orthogonal, whether and how they employ a cyclic prefix, and how the subcarriers are filtered to make them well localised in the frequency domain [53]. FBMC is quasi-orthogonal, performs a per sub-carrier filtering and eliminates the cyclic prefix, but care must be taken in the implementation since contrary to OFDM, GFDM and UFMC, it uses Offset Quadrature Amplitude Modulation (OQAM). GFDM also performs per-subcarrier filtering, and reduces the overhead of the cyclic prefix by employing it for several symbols, instead of per symbol as in OFDM. However, its non-orthogonality introduces self-interference even if the transmitters are perfectly synchronised. This requires a more complex receiver using e.g. successive interference cancellation. UFMC eliminates the cyclic prefix and applies a filtering for a sub-band consisting of several subcarriers, where the subcarriers within a sub-band are orthogonal to each other but the sub-bands are non-orthogonal, introducing less inter-carrier interference compared to GFDM. Numerous comparisons between those waveforms have been made regarding implementation complexity, spectral efficiency, robustness towards multi-user interference (MUI) and resilience to power amplifier non-linearity etc., see e.g. [54] and [55].

Even if they have not yet been adopted in 3GPP, these post-OFDM waveforms are promising schemes, especially in asynchronous multiple access for massive IoT scenarios. Therefore, application-oriented research on algorithms and proof-of-concept implementations are needed to make them more mature.

3.7 Enhanced Modulation and Coding

Channel coding can be regarded as one of the most complex parts of the baseband transmission chain, and aims to correct errors to establish reliable communication. For decades, researchers sought for channel codes with good error correction performance approaching Shannon's capacity limits with manageable complexity. Modern channel coding schemes such as Turbo, LDPC and Polar codes with excellent performance made their way into several communication standards after advancements in semiconductor technology. However, as the decoders for those codes are very complex, there will be implementation bottlenecks (w.r.t. computational complexity, algorithm parallelisation, chip area, energy

efficiency, etc.) to be addressed for high throughput (e.g. when throughput is over multiple Gigabits per second) and/or low latency applications are targeted by future communication standards.

Even though these modern coding schemes show near-capacity error correction performance for many channels (e.g. binary input additive white Gaussian channels, BI-AWGN), their combination with higher order modulation schemes (such as QAM) can lead to a sub-optimal performance. One reason for this degradation is the so-called ‘shaping loss’ caused by the probability distribution of the transmitted symbols [56]. In order to approach capacity, the transmitted symbols need to have a certain probability distribution (e.g. discrete Gaussian distribution is needed for the transmission over AWGN channels), and using uniformly distributed symbols results in a performance loss, which can be up to 1.53 dB on AWGN channels.

Several solutions for constellation shaping are proposed to compensate this loss. One option is to optimise the locations of the modulated symbols in the constellation diagram to obtain non-uniform constellations (NUC), as adopted in the ATSC3.0 standard [57]. This scheme is also called geometric shaping and shows improvements compared to uniform signalling. Another approach is the so-called probabilistic shaping [58], [59] and [60], where a shaping encoder is employed to encode messages in a way that the transmitted codewords have a non-uniform probability distribution, resulting in a capacity achieving distribution when combined with simple QAM symbols. This approach is shown to perform close to channel capacity. Another feature of probabilistic shaping is that the probabilities of transmitted symbols can be changed to adapt the transmission rate without changing the FEC code. This is of particular importance since a single FEC code design is sufficient for rate-adaption. Considering the diverse requirements of future communications systems, several shaping encoders suitable for both high throughput and ultra-low latency (short blocks) have been proposed in the literature [59] and [61]. However, hardware implementation of efficient shaping encoders and decoders need further investigations.

Constellation shaping provides significant improvements in terms of error correction performance. In general, signal shaping is a fundamental and important technology to further improve the spectral efficiency of wireless and wireline communication systems, as the shaping loss may be considered as one of the last gaps between Shannon’s information theory and the practical communication systems to be bridged.

3.8 Improved Positioning and Communication

Especially in the massively connected world of the ”Internet of Things” (IoT), it is getting more and more important to be aware of where all these ”things” are located. Mobile radio based technologies envisaged for a future system will play an essential role in providing high accuracy positioning of the ”things”. State-of-the-art communications systems like 4G LTE support positioning in a non-cooperative way, i.e. in the downlink, several base stations send reference signals and the UE measures them and in the uplink, a UE sends reference signals and several base stations measure them. This is good enough to support the requirements imposed by the FCC for localisation in case somebody makes an emergency call (so-called E-911), where an accuracy in the order of 50 m can be required [62]. There are use cases for future mobile communications, e.g. V2X, smart factory and others, however, where a higher localisation accuracy is required. For instance, for V2X vulnerable road user discovery, an accuracy as high as 10 cm may be required (see e.g. [63]). Currently, 3GPP is considering a positioning accuracy of 50 cm for the future 5G NR standard, due to the higher frequencies and large signal bandwidths, dense networks as well as improved device-to-device communications available for 5G.

Cooperation can boost the positioning accuracy [64], especially in massively connected scenarios. In cooperative positioning, the user equipment (UEs) can send and receive signals, and exchange their position relevant information. If the density of UEs is large, it is likely that

there are line of sight (LOS) propagation conditions to each UE from several UEs, which is significantly increasing achievable localisation accuracy. There are two different approaches to position calculation, a centralised approach where a central entity calculates the position and a decentralised approach where UEs calculate their position based on the position estimates of the UEs in their vicinity.

While it is by now known that MIMO systems improve spatial diversity and result in spatial multiplexing gains, their power in improving positioning accuracy has not yet been fully exploited. Large antenna arrays at the BS (base station) result in very fine angular sampling, which can be leveraged for positioning methods. Further, existing positioning methods only work well in strong LoS environments in general. Many environments, however, experience strong multipath which cause performance degradations and reduces position accuracy. For that reason, the existing methods need to be revised or new methods need to be developed to accommodate multipath propagation. Such methods can additionally leverage the presence of large antenna arrays at the BS [65]. Clearly, having multiple antennas at the UE can improve cooperative positioning.

Accurate positioning can be leveraged to enable location-aware communications [66], e.g. design of narrow beams targeted towards the intended user in traditional cellular systems, facilitate autonomous driving, etc. Furthermore, accurate positioning is a prerequisite for emerging industrial and factory applications. Therefore, in contrast to legacy systems, positioning has a big impact on the operation of future communication systems. For these reasons, investigating new positioning paradigms, e.g. for joint communication and positioning, is essential, as it can further improve spectral efficiency, energy efficiency, and reduce latency.

3.9 Random-Access for Massive Connections

The future vision of IoT envisages a very large number of connected devices, generating and transmitting very sporadic data. The challenge here is how to coordinate such a network without spending the whole network resources and node energy in protocol overhead. Modern information theoretic research has formalised this problem as follows: consider a number of nodes, each of which makes use exactly of the same code, which is hardwired into the device for system simplicity and cost reasons. These nodes access a common transmission resource at random in a very sporadic manner. The receiver (e.g., a base station) must decode the superposition of codewords without knowing a priori who is transmitting [67]. After decoding the messages (payload), the ID of the transmitter can be found as part of the message, if necessary. For example, in some applications it is important to know the transmitter, but there are applications in which it is important to get the data and not the identity of the transmitter. The challenge now is to design such new random-access codes for which the superposition of up to K distinct codewords can still be uniquely decoded.

This new random-access paradigm is inherently related to **group testing**: a set of statistical procedures for which it is possible to identify the presence of certain individual agents by sampling combinations thereof [68] and [69]. A related setting consists of coded slotted Aloha, where sparse codes with iterative message passing decoding are developed along multiple random transmissions, to effectively eliminate interference by a sort of low-complexity successive interference cancellation [70].

A related problem consists of activity detection using a receiver with a large antenna array: in this case, users are given unique signature sequences and transmit at random in a completely uncoordinated way. The base station has multiple antenna observations, and must identify the ”active set” of users that are transmitting. This problem is related to **compressed sensing** where the sparse vector to be estimated is the vector of 0s and 1s, denoting ”absence” or ”presence” of the transmitters. Modern techniques based on Approximated Message Passing (AMP) can be used for this purpose [71] and preliminary research results show the exact trade-off between the length of the signature sequences (protocol overhead) and the number of

active users, such that the probability of identification error can be made as small as desired [72] and [73].

In both cases the massive random-access and the activity detection problems, a significant research effort must be made in order to bring the abovementioned theoretical ideas to practice and to solid and principled system design. Furthermore, even the basic theory needs to be extended, for example, to encompass asynchronism and presence of unknown parameters, such as phase and frequency offsets, and random fading coefficients, for which the current theory has only partial answers.

In a second step, this line of research should consider waveforms adapted for low-latency sporadic access for the cyber-physical systems characteristic of the tactile Internet [74]. Here, sub-ms latencies may be required in order to control moving or even flying objects (passenger drones) or other similar scenarios requiring the combination of ultra-reliable communication with centralised control systems. Similar mechanisms will also be required for evolved Industry 4.0 applications [75]. It is envisaged that the physical-layer transport mechanisms will be associated with real-time cloud computing (mobile edge computing) in proximity to the radio network to implement the necessary control loops. This concerns primarily sub-6GHz access for the uplink and massive connectivity of objects to wireless infrastructure. The objective is to provide solutions for the evolution of cellular IoT uplink waveforms and protocols that scale to huge number of connected devices with stringent energy and potentially latency constraints.

3.10 Wireless Edge Caching for Further Increased Throughput

Wireless communication networks have become an essential utility for citizens and businesses. Wireless data traffic is predicted to increase by 2 to 3 orders of magnitude over the next five years [12] and [76]. The implications of these trends are very significant: while continued evolution is to be expected, the maturity of current technology (e.g., LTE-Advanced for cellular and IEEE 802.11ac for WLAN) indicates that the required orders of magnitude throughput increase cannot be achieved by an incremental ”more-of-the-same” approach. As far as wireless capacity is concerned, the forthcoming 5th Generation (5G) of standards and systems is focused to a certain extend on the traditional view of ”increasing peak rates” [77]. In contrast, it is widely recognised that a major driver of the wireless data traffic increase is on-demand access to multimedia content (Wireless Internet) [12] and [76]. Peak rates do not necessarily yield an improved user Quality of Experience (QoE). For example, typical video streaming requires rates ranging from ~ 400 kbps (standard quality) to ~2 Mbps (high quality). What really matters for the end user QoE is the availability and stability of such rates, so that a video can be played anywhere, at any time, and without interruptions. Also, we observe that the users' content consumption pattern and the operators' data plans are dramatically mismatched. For instance, a standard monthly data plan in the EU includes ~ 3 Gbytes of LTE traffic at a cost ranging between 15 and 50 EUR, while a single movie requires ~ 1.5 Gbytes of data, such that the whole plan would be depleted by streaming ~2 movies.

In light of the above considerations, a novel content-aware approach to wireless network design is needed. Such novel approach should support the paradigmatic shift **”from Gigabits per second to a few Terabytes per month for all”**. More precisely, the special features of on-demand multimedia content can be leveraged in order to deliver a target of ~ 1 $TB/month of content data to each user in a scalable and cost-effective manner. This target is far more challenging than achieving Gbps peak rates, which have been already demonstrated by various ”5G-ready” experimental platforms [78] and [79].

Meeting this challenge requires a **profound and non-incremental advance** in the information theoretic foundations, in the coding and signal processing algorithms, and in the wireless network architecture design, in order to exploit the potential gain of content-awareness.

Recent research in information theory and wireless communication has shown that content distribution over a wireless network (e.g., on-demand video streaming) can be made much more efficient than current state-of-the-art technology by caching content at the wireless edge

[80], [81], [82] and [83]. This means pre-storing segments of the content files at the base stations, at dedicated ”helper” nodes, and also in the user terminals.

Traditional caching (e.g., prefix caching) decreases the transmission load by the fraction of data already present (pre-cached) at the destination. With these novel modern techniques, based on extensive use of network coding, it is possible to show that a constant (non-vanishing) per-user throughput can be achieved while the number of users grows to infinity. We refer to this behaviour as ”full throughput scalability” [84]. For the sake of concreteness, consider the analogy with conventional TV broadcasting: in this case, leveraging the broadcast property of the wireless medium, an infinite number of users can be served with a finite transmission resource, i.e., a finite bandwidth and transmit power. For example, this approach is taken in the so-called enhanced Multicast-Broadcast Multimedia Service (eMBMS) in 4G networks. Now, the reason for which eMBMS turned out not to be a huge success is that users do not consume wireless multimedia as they used to consume traditional live TV: they wish ”on-demand” services, to access what they want at the desired time and location, and not at the time decided by a TV broadcaster. With on-demand delivery, the broadcast nature of the wireless medium cannot be exploited in a direct and trivial manner. In fact, streaming services today treat the on-demand traffic as unicast individual traffic, as if the content was individual independent data.

Treating on-demand content as unicast traffic is highly inefficient, since it does not exploit the huge redundancy inherently contained in the users’ requests, which concentrate on a relatively small set of very popular files, especially in video-server services where the library of popular movies can be controlled by the service provider, and can be updated at a relatively slow pace (e.g., the library is refreshed every day/week/month). Such redundant requests arrive to the server in an asynchronous way, such that the probability that many users wish to stream the same file at the same time is basically zero. Coded caching techniques have the ability of turning the unicast traffic (on-demand streaming) into a coded multicast traffic, such that again the scalability of broadcasting a common message is recovered and full throughput scalability is achieved.

Beyond these very compelling theoretical results, a significant knowledge gap must be filled to make these ideal of practical value. Therefore, a significant research effort must be made yet in the following areas:

1. Coding (e.g., combining edge caching with modern multiuser MIMO physical layer schemes);
2. Protocol architectures (e.g., combining edge caching with schemes for video quality adaptation such as Dynamic Adaptive Streaming over HTTP (DASH) [85]);
3. AI/ML based content popularity estimation and prediction, to efficiently update the cached content.

4. Optical Networks

Within the next decade, the world will go digital. European citizens will enjoy a new quality of life, industrial productivity will flourish, and digital value chains will create a vibrant economy while addressing inequalities in education, labour market, wealth creation and civil rights. We will enter a new era in which trillions of things, billions of humans, and millions of connected autonomous vehicles, robots and drones will generate Zettabytes of digital information. All this information needs to be transported, stored and processed. AI/ML will free us up from routine tasks and boost human creativity and product innovation.

Smart connectivity will be the foundation of this new digital world: Always available, intrinsically secure, and flexibly scaling. A programmable network infrastructure will be the nervous system that the digital society, industry and economy will heavily rely upon. Delivering the required performance, resilience and security levels, while satisfying cost, energy efficiency and technology constraints, presents a formidable research challenge for the next decade.

Optical networks have long been the solution of choice for submarine, long-haul, and metro applications, thanks to the unparalleled capacity, energy efficiency and reach of optical fibre transmission. In recent years, optical network technologies have conquered inter- and intra-data centre networks and have created tremendous growth in this sector.

Building a mobile network will require optical connectivity to each radio antennas and a powerful optical network behind. So, the mobile networking we all expect everywhere and which forms the basis for IoT and other smart networking applications will continue to rely on progress in the optical infrastructure to higher capacity, lower latency, increased programmability, increased environmental hardening and significantly reduced power consumption.

As capacity and latency requirements increase, optical solutions will also become the de-facto standard also in the access network. Fibre-to-the-x (FTTx) technologies will displace copper and radio technologies wherever mobility is not required and fibre can be made available.

Additional benefits such as their reliability and EMI (electro-magnetic interference) immunity make optical network technologies attractive also for applications such as critical infrastructures, factories of the future, private enterprise networks, and vehicular networks. Overcoming the challenges in scaling electronic interconnect speeds together with the remarkable advances in electro-photonic integration will pave the way towards a new generation of optical networking and IT equipment. Combining the advantages of optics and electronics and leveraging mature micro-electronics packaging is the way forward to deliver unprecedented functionality, compactness and cost-effectiveness.

Seven out of the top 20 network operators are headquartered in Europe while six out of the 20 largest optical equipment manufacturers have major R&D centres in Europe. By revenue, they represent more than 30 % of the global optical equipment market. Two of the top 3 component manufacturers have operations in Europe and more than a hundred SMEs and universities provide complementary innovation on network, system, or component levels. Optical technologies leverage a telecommunication infrastructure market of 350 Billion EUR and impact more than 700,000 jobs in Europe [86].

From ground-breaking discoveries such as optical fibres and EDFAs (Erbium Doped Fibre Amplifier) over products such as WDM systems and 100 Gbps transponders to global standards such as SDH (Synchronous Digital Hierarchy) and OTN (Optical Transport Network), Europe has been at the forefront of optical communications R&D for many years. Yet, innovation cycles are fast and competition is fierce. New research challenges require a continued effort to defend and strengthen Europe's leading position.

4.1 Flexible Capacity Scaling

The question of data traffic growth in optical networks is coming up periodically in the discussions of analysts. A few years ago, it was claimed there was a slowdown in the pace but these observations failed to notice that there was a massive transfer of data from public Internet to the private Intranets of cloud providers. Overall, the global data traffic has been doubling every 2-3 years over the past 15 years and it will continue to increase at an impressive rate. Extrapolations from today's traffic predict rates of 10 Terabit/s for opto-electronic Interfaces and over 1 Petabit/s for optical fibre systems by 2024 [87]. Symbol rates more than 100 GBd, although they are challenging from the technology point of view, are crucial to maintain the transportation per bit cost at the necessary levels to stimulate innovation.

This evolution stumbles upon the most fundamental limits of physics that are: the Moore's law on Silicon integration and Shannon's limit on fibre capacity which both are tremendous barriers to future growth. These limitations are already slowing capacity increase and will become gating items just a few years from now so urgent research efforts are necessary to avoid a system gridlock. There is a clear danger that a two-fold increase in the requested capacity will require doubling the amount of optical/electronic hardware. This will increase cost in a linear fashion and threaten future capacity growth. Obviously, disruptive approaches are now needed in optical networks to push Shannon's and Moore's limits out further.

First, recent successful innovations will be exploited far beyond current status. It can be predicted that optical communications are moving to coherent everywhere. Once viewed as prohibitively expensive, coherent technologies will massively expand from long-haul systems into all fields of optical communications: to support the new generations of wireless systems beyond 5G, to offer enhanced broadband access, to cope with the growth of inter data centre communications, to make edge cloud a reality for all and even to allow a new breed of intra-data centre networks. Coherent is the most promising technology to bridge the gap that is caused by the Shannon limit, leveraging ”shaped” modulation formats, flexible rates, and increased density WDM.

Contemporary digital coherent technologies were essentially born in Europe in 2010 and Europe is in a good position to keep its leadership if it continues to innovate at fast pace. A mutualisation of research efforts to establish and drive the upcoming standards is however required. We argue that a new flagship initiative ”Open Coherent Communication Everywhere” tackling reach, capacity and cost considerations across all applications based on a common technological approach would be key in federating Europe's strengths.

To expand network capacity beyond the Shannon's and Moore's limits, given by current fibre and integration technology, we need to exploit all dimensions in space and frequency, opening new optical wavelength bands and space division multiplexing.

The exploitation of new wavelength bands will require advances in a wide range of technologies ranging from optical amplifiers, tailored to these new bands, to a wide range of opto-electronics devices and sub-systems; namely, tuneable lasers, optical multiplexers, couplers, optical mixers, photodiodes, wavelength selective switches and other optical switching solution. System design guidelines will also have to be revised and updated taking into account the new physical impairments that will undoubtedly come up in the new bands. Intensive research efforts are necessary along these lines.

In parallel, we must invest in space division multiplexing. This approach can offer several orders of magnitude capacity increase, either by multiplying fibre count in cables, or by introducing multicore or multimode fibres. Here again, new node and system architectures, new digital signal processing, new space division multiplexers, new switches, new optical amplifiers are needed along with the new fibre types needed. For space division multiplexing to become a cost-effective reality, a change of scale in component count per square millimetre will be required.

4.2 New Switching Paradigms

Future applications, such as autonomous driving, augmented/virtual reality and augmented workspace, will severely change the architecture and dynamism in optical networks. New network architectures with edge clouds close to the end user and centralised clouds with flexible distribution of network and application functions will be required. Cloudlets at the edges can be viewed as ”data centres in a box”, that can be flexibly deployed at the network edge to meet the capacity or latency constrains required by the applications.

The optical transport network then can be seen as a programmable network fabric that can dynamically provide slices of distributed network, compute and storage resources to applications and tenants. Switching can be accomplished on layer 0-3 depending on technology availability and service requirements and may cross multiple vendor or operator domains. Emerging technologies such as flexible Ethernet (FlexE) or flexible OTN (FlexO) add further degrees of freedom. With the trend towards disaggregated switch platforms, multi-layer switching functions often need to be orchestrated across multiple platforms. Technologies such as software-defined WAN (SD-WAN) move intelligence to the end-points of a connection and can use multiple transport options in-between to optimise cost, performance and survivability.

Flexgrid technology on the optical layer allows the introduction of a spectrum-as-a-service model offering the opportunity for a flexible network slicing in the wavelength domain. An operation over multiple wavelength bands and spatial dimensions requires new switch and transponder architectures that have not been discussed in great detail yet. Some applications may require network resources only for a very short time. Consequently, approaches enabling a faster reconfiguration (< 1 ms) on the optical layer and taking into account concerns such as amplifier power transients need to be developed.

While optical switching in commercial applications has so far limited to circuit switching, advances in photonics integration could allow optical flow or packet switching approaches to become practical, which were previously considered too costly or complex to implement. This opens a range of new applications and use cases.

Research is required to investigate the benefits and drawbacks of different switch architectures on the network and application layers, to develop novel switching architectures and new routing protocols within, and to develop new semantic description and information models allowing the control of new devices by an SDN controller platform. New languages could be used to programmatically define switch configurations so that formal verification and rule checking becomes possible on a global network scale.

4.3 Deterministic Networking

While a lot of the success of the Internet relied on a best effort traffic paradigm, the digitalisation brings a multitude of applications in which reliability, latency and often also a certain throughput and signal quality need to be guaranteed (e.g. URLLC = ultra-reliable low latency communication in 5G). Examples range from mobile fronthaul traffic over critical control applications in the vehicular and industrial space to high-resolution machine vision or augmented/virtual reality applications. Some of the most challenging requirements discussed today are < 75 µs latency (including fibre transmission which adds 5 µs/km), < 8 ns timing error, and several tens of Gbps throughput for CPRI signals.

Networking assuring end-to-end deterministic-performance is in the centre of attention that may include, but is not limited to, controlled physical layer performance, guaranteed throughput of high-priority services, and upper bounds in QoS parameters such as latency and jitter. Typically, a precise timing solution is required to provide a time reference to all network nodes with sufficient accuracy. A central traffic management is desired to avoid an overbooking of the network and to be able to provide service-level assurance for the services

running over the network. In addition, deterministic networking should also be revisited in the context of network and central office (CO) virtualisation. Research on how to achieve deterministic QoS targets while using function chaining over shared compute and network resources should be addressed. This includes study on hybrid use of electronic and optical switching as well as studies on the scalability of guaranteeing deterministic QoS for a large number of flows/applications. A redundancy concept should ensure continued operation in case of an equipment or link failure. Some of the applications will have to run over a public network. In other scenarios private network builds are also possible or sometimes even required.

Multiple options and technologies are being debated in standards bodies such as ITU-T, IEEE, and IETF and range from OTN over FlexE to time-sensitive Ethernet or IP approaches. However, the overall picture is not clear yet and the achievable performance of different methods needs further investigation. Further research is required to determine the optimal solution set for a diverse range of applications, to develop the necessary planning and provisioning tools as well as the means for service assurance and performance verification.

4.4 Optical Wireless Integration

5G will deeply transform the underlying transport network, due to several concurrent causes: the end-user capacity is increasing, the coverage of the mobile network becomes denser, and a split architecture in the radio access network is introduced, where different functional splits between baseband and radio units will be supported by the same transport network. The latter means that the traditional distinction between fronthaul and backhaul networks is blurred. Since it is impossible, for cost and operational reasons, to design dedicated networks for a considerable number of heterogeneous last-drop technologies, the adoption of a shared network infrastructure that makes use of common transmission and switching platforms is unavoidable. This is not a trivial task and will be the big network design challenge for several years from now.

In principle, several candidate technologies exist for enabling the coexistence of fronthaul and backhaul networks. In practice, all require a redesign and a redefinition of their application space. For example, packet switching with new packet friendly fronthaul interfaces is likely to be implemented in scenarios where many users generate a low amount of traffic data each. However, the need to meet tight latency constraints (which can be as low as of 100 μs) and deterministic delays across several packet switches may require the development of new framing, multiplexing and synchronisation techniques that can guarantee the requirements of time-sensitive networking without losing the advantages of statistical multiplexing. The convergence of radio and fixed access, the necessity to support dedicated enterprise connections, and the transport of high bit-rate fronthaul signals corresponding to the high-level split options defined by 3GPP or the CPRI consortium will foster the adoption of interchangeable multi-layer switching platforms supporting various switching granularities that range from packet, to time-slot and wavelength channel level. This will lead to a further level of convergence, where a layered control plane offers to operators the ability to set up services in short time, while being unaware and totally decoupled from the underlying transmission and switching technology. However, the current transport networks based on OTN and DWDM (Dense Wavelength Division Multiplex) are, in some respects, inadequate to support the requirements posed by the new generation mobile systems.

In the emerging new ”mobile transport network”, where it is of particular concern to lower costs as well as to mitigate the jitter effects caused by complex justification mechanisms, the adoption of a single time-division multiplexing hierarchy across all network segments (access, aggregation and the core) would be highly desirable. Moreover, 25 and 50 Gbps channels, corresponding to high-level radio functional split interfaces, should be multiplexed and transported in an efficient (i.e. with minimal overhead) and in a scalable way. At a physical layer, DWDM already has the formidable aggregate capacity capable to support 5G broadband services in densely populated areas, scaling up to several hundreds of Gbps/km^2.

However, breakthroughs in technology are still necessary to bring down the total cost of ownership, something that is seen as the necessary condition to justify the large-scale deployment of DWDM in x-haul (i.e. the network segment where mobile front- and backhaul as well as fixed access transport converge). Examples of enabling technologies here are: cost-effective 50 Gbps (and beyond), optical interfaces with direct detection receivers capable of reaching distances up to 40 km in the 1550 nm window, possibly employing new modulation formats; tunable photonic integrated devices for compensating the chromatic dispersion; ”lite” coherent transceivers with low DSP complexity as well as new technologies for developing cost-effective and compact optical amplifiers. A new generation of coherent transceivers and optical amplifiers would also allow to implement fronthaul networks exploiting broadcast-and-select DWDM Passive Optical Networks (PONs) architectures. These technologies are to compensate the attenuation of passive splitters in the optical distribution node, which is a key issue for bit-rates 50 Gbps and above. This scenario is of considerable importance as it enables the convergence of fixed access and fronthaul on PON fibre infrastructure that is capillary deployed, at least in big cities.

Another technological breakthrough would be the design of ”fully colourless” DWDM networks, namely networks based on port agnostic devices that are offering the advantage to greatly simplify network provisioning, with lower installation and operational costs. Fully colourless networks require both tunable transmitters and reconfigurable optical add-drop multiplexers (ROADMs). Although ROADMs are widely used in wide area DWDM networks, current Micro Electro-Mechanical Systems (MEMS) or Liquid Crystal (LCoS) technologies can hardly scale their cost down enough to be applicable to an access network. Novel system-on-chip devices based on silicon photonics could potentially reduce the cost by two orders of magnitude.

All aforementioned transport technologies deal with digital radio split interfaces. Analogue Radio over Fibre (A-RoF) is a well-known alternative technology for the distribution of wireless signals: ideally, an A-RoF system acts as a mere medium converter, creating in optical fibre an exact copy of the radio signal on air, without further processing, with obvious benefits in terms of hardware complexity and power consumption. In practice, many performance issues remain to be solved before A-RoF can be deployed to the variety of scenarios where digital units are used today. Examples of challenges to be addressed by A-RoF systems are: noise mitigation techniques that compensate for the absence of equalisation and forward error correction mechanisms implemented in digital systems; and linear modulation and photodetection devices to decrease the effects of high peak-to-average power ratios and inter-modulation in multi-carrier wireless systems. Those two aspects are especially critical in 5G, due to the high order modulation format (256-QAM) and the high number of subcarriers (2048 OFDM) used in wireless systems. Furthermore, the introduction of millimetre waves necessitates to development of linear devices with high signal bandwidth, up to 100 GHz. Centralised architectures, where the baseband processing unit is shared by a certain number of remotely placed radio units, meet operators’ plans for a reduction of the number of network nodes, and consequent saving of operational costs, but pose further challenges due to the increased link budget and the accumulated chromatic dispersion caused by the higher distance between antenna unit and central office.

Finally, an important aspect to be addressed is the definition of an end-to-end control system, able to manage the interface with the packet switched core network and to monitor the quality of service for the whole link, encompassing A-RoF and packet switches. For example, one of the issues brought about by massive densification of cells is the high traffic fluctuation they experience due to the lower number of users served. As new fronthaul technologies capable of restoring the relation between the variable cell traffic and their fronthaul transport data rate (e.g., through functional split or variable-rate fronthaul), the control plane needs to coordinate statistical multiplexing of the allocated resources across the wireless, access transport and central office domains, merging priority and best effort applications over the same shared infrastructure.

4.5 Optical Network Automation

The new flexibility and technology advances in optical networks, including increasing programmability and remote configurability at the device level, require advances in network control, automation, and autonomicity. Software defined optical networks (SDON) based on logically centralised control and management provide the basic capabilities for new operational paradigms and network automation. With SDON, a centralised controller supports programmable flexibility of optical links in multi-layer, multi-domain optical networks with tuneable wavelengths and variable modulation schemes. It also allows multi-vendor ease of integration in the business and support systems. However, autonomic transmission and networking need to be further investigated.

Adopting SDN principles, including centralised deployments and the systematic adoption of unified data modelling frameworks, languages and open Application Programming Interfaces (APIs) allows an easier integration with Operation and Business Support Systems (OSS/BSS). This includes, in particular, the possibility of adopting dynamic and flexible pricing schemes and improved integration with billing systems. Open and standardised model driven control can integrate the optical network with OSS and payment systems to achieve new revenue streams for network infrastructure providers. The use of smart contracts will also allow cross-country end-to-end dynamic programmable connectivity.

As stated, the industry is slowly adopting a unified approach to device information and data modelling. The first initiatives mainly cover modelling low-level optical systems and devices such as transceivers, open line systems and ROADMs, in view of having common models across-vendors and, in particular, easing network control in the so-called disaggregated deployments, where individual components can be deployed and upgraded independently. However, modelling activities need not to stop at the optical device level and should cover other aspects related to network operation, including, for example, topology and inventory management or service description. This implies the need for a gap analysis of current modelling languages, identify shortcomings and limitations and extend them where required. This research and development activity should be carried out in view of subsequent standardisation, avoiding the complexity of having multiple, often overlapping and incompatible models, frameworks and languages. This also involves finding a clear balance between de facto and de jure standards, combining the outcomes of Standards Defining Organisations (SDOs) and those of Open Source initiatives and projects, typically more agile and based on frequent releases, continuous integration and continuous deployment (CI/CD).

Control and Management solutions (in terms of actual products) tend to be monolithic, vendor-dependent and strongly coupled to the underlying infrastructure. However, due to operators evolving requirements and market pressure, and coupled to the softwarisation trend and the adoption of unified modelling, a related industry trend is the disaggregation at the software level, and the progressive evolution towards more micro-services based architectures, where basic, reusable functional elements and modules are clearly defined, with standardised interfaces and reference points and can be composed and allocated depending on the actual needs.

Smart Networks in the context of NGI requires additional solutions and innovations for optical network automation beyond simple programmability. Global reach and optimised local service delivery capabilities need to be combined in highly flexible and highly granular ways and should be available on-demand for the value chains of web-based software and IoT platforms.

Intent-based zero-touch provisioning of network services must be enabled for Smart Networks. Another key aspect for zero-touch control of optical networks is the introduction of network resilience in the transport domain, which is one of the key aspects for network cybersecurity.

The advances in optical transmission and switching technology create new challenges for network automation and autonomicity. Spatial division multiplex (SDM) introduces an

additional dimension of flexibility in routing and spectrum allocation. Margin-optimised transmission to maximise the fibre capacity requires dynamic re-routing and network re-optimisation in case of aging effects or additional splicing loss after fibre repairs.

The complexity of the underlying optical technology, resulting in a large number of interdependent configuration parameters, requires cognitive networks powered by streaming telemetry, real-time network measurements and AI/ML. Optical devices can potentially produce a huge amount of operational and monitoring data, and operators are eager to process and use such data, in view of the potential and promised improvements in network operation and automation procedures. However, this challenge involves the development of flexible architectures and protocols for streaming telemetry, for which it is commonly accepted that current methods are limited in terms of encoding efficiency, imposed overhead and sampling rate. Along with the new architectures and protocols, consensus needs to be reached which parameters are key indicators, so that common procedures and algorithms can be used in heterogeneous scenarios. It is also worth mentioning that telemetry does not only rely on retrieving data from one (or multiple) monitoring points, but also involves measurements at the network level, potentially requiring active network probes and joint processing of data measured at different places and times (spatial and temporal diversity). The design of such probes is also a challenge, since it is well known that it should not be disruptive affecting existing services or consuming an excessive amount of resources.

Coherent transponders inherently provide monitoring and sensing capabilities. Streaming telemetry and continuous performance monitoring provides the data for AI/ML. Using model-based approaches and statistical analytics, SDN controller can observe the network behaviour, analyse the performance, detect anomalies, and autonomously act, re-configure and optimise the network. However, the simplistic approach of ”throwing” some AI/ML algorithms on a large set of data will not achieve this goal. To fully utilise the expertise from big data analytics generally available, dedicated research activities for data curation and analytics, AI/ML in optical networks are required.

Given the support that the optical infrastructure provides to 5G-based services, optical performance monitoring and analytics based on such data seems to be critical to understand the base-line performance of such upper layer services.

4.6 Security for Mission Critical Services

Ever increasing interconnectedness not only of people but also of devices starting from huge power plants down to billions of IoT devices like sensors or kitchen appliances does not only increase the dependence on the network infrastructure but also expand the threat surface and therefore the vulnerability of every individual and of the society as a whole. Important threats do not only include hacking and espionage, but also network outages due to natural catastrophes as well as terrorism and sabotage against critical infrastructure. Therefore, it is getting more important, in addition to protecting computer systems and personal devices, to also better safeguard our network infrastructure against data leakage and unexpected service outages.

A signal on an optical fibre can be easily tapped, once the physical access to the fibre is available. At this point, the data of millions of users and billions of applications is exposed to theft and manipulation. Therefore, encryption and integrity of the data is essential and also needs to be kept at a level playing field with increasing threat scenarios (crypto agility). Improvements need to consider quantum-safety, for instance by post-quantum replacements of current algorithms or by provable and long-term secure data transmission of highly sensitive information using quantum communication with photons. Also, novel research directions like physical layer security for optical networks should be explored.

Adding redundancy is the conventional, but also expensive way to improve the reliability and resilience of networks. Alternative concepts, that are high on the research agenda today, are increased flexibility, massive monitoring and software control of optical networks. Data

processing (e.g. by means of AI/ML methods) can help to detect upcoming problems early and counteract in advance with the available flexibility. It should be possible to employ this functionality across the borders of a single networking domain.

However, the higher flexibility of optical networks, enabled through software controlled network elements (software defined networking, SDN), also increases the vulnerability of such networks to various kind of attacks and therefore security and resilience aspects need to be part of the concepts from the beginning (including both the hardware and software layers of the network). More generally, the design of network equipment needs to employ modern security and reliability paradigms (security by design) and apply modern software technology to foster efficient and secure implementation of increasingly complex network elements.

4.7 Ultra-high Energy Efficiency

With data centre traffic consuming nearly 2 % of all electricity used today and the share of communications technology in overall world energy consumption growing over the last decade, there is an urgent need for a paradigm shift to greener IT technology.

Increasing use of optical technologies within the IT and communications industries is one key opportunity to limit the increasing energy consumption against the massive growth of overall data capacity that networks and data-centres are handling. Since light can travel vast distances through fibres, fibre optics consumes only a fraction of the energy used by conventional technology that transports electrons via copper wires.

This means that higher reach and higher capacity interfaces can significantly reduce the energy consumption of networks as they reduce the number of optical interfaces and regenerations needed.

Also, the inherently low power consumption of optics can be directly used to reduce power consumption overall, if optical functions replace more power-hungry electronics, e.g.:

- Functions may be turned off or switched into a low power mode if not in use
- Electronic processing can be bypassed more frequently through new control mechanisms that optimise traffic flows across network layers, particularly if combined with optical space and wavelength switching.
- Electronic interfaces of modern communications ICs are a strong driver of power consumption. Ways to replace those interfaces by lower power optical interfaces that are integrated or co-packaged with those chipsets can reduce power consumption substantially.
- The further drive to higher capacities and interface speed per electronic chip will make optical communication within chipsets as well as chip-to-chip a future necessity.

4.8 Optical Integration 2.0

Two developments, coming from different size scales, require progress in the field of optical and electronic integration: On one side, electronic integration advances, following Moore’s law, lead to increased interfacing requirements between electronic processing chips on a printed circuit board (PCB) or between PCBs. More numerous and higher speed lanes require the transition from high-power consumption electronic interconnects to optical interconnects to scale the increased processing power while limiting the power consumption. Here, integration of the optical components for transmission and detection of the high-speed signals into the electronic Silicon platforms is required. On the other side, increased data traffic in optical networks will require more and more high-speed optical interfaces, when exploiting wavelength and spatial multiplexing. To avoid scaling of cost and power consumption with the exponentially increasing data traffic, the development of standardised components for spectral and spatial unit cells are required [87].

Both applications point to Silicon Photonics (SiPh) as integration platform, enabling the use of Silicon mass manufacturing processes for optical applications. Also, standard electronic packaging technologies can be leveraged, eliminating the need for expensive gold boxes and special assembly processes. Known as a low-cost platform for non-hermetic and high-temperature operation, Silicon provides good passive optical properties for routing, modulation and detection of light. It is a natural fit to also integrate RF electronics technology into this platform for driving and controlling the optical SiPh components. Projects in this field are well underway. It needs to be mentioned, however, that while Moore's scaling of electronic memory and processors yields ever-smaller structures in Silicon, this miniaturisation is not feasible for optical components, where the telecommunication wavelengths on the order of a micro meter pose a size limit. Further integration of photonic and digital processing functions will require scale adaptation.

Work is also required to enable active functionality, like lasing or amplification, into the SiPh platform. Initial projects are underway to integrate and structure III-V materials (e.g. Indium-Phosphide) into the Silicon substrate as well as other passive materials (e.g. Silicon Nitride). It is crucial that this integration can be accomplished at wafer scale and without sacrificing the benefits of the silicon platform that are cooler-less, non-hermetic, high temperature operation.

Further performance gains will be achieved by adding organic materials into the Silicon platform, providing very high optical coefficients and reducing the required driving power [88]. This will ultimately be beneficial for the integration of optical functionalities on every size scale.

Advances in photonic integration will pave the way for a raft of new IT and networking devices in which optical, RF and digital electronic functions can be combined, initially in multi-chip modules (MCM) comprising highly integrated CMOS dies and high-speed optical engine dies on the same substrate. Appropriate package design allows a reflow soldering of these components under standard process conditions.

5. Edge Computing and Meta-data

5.1 Beyond Mobile Edge Computing

Research areas on Edge Computing are widely covering many aspects ranging from communication networking and IT topics. In addition, recently a consolidated set of standards (from ETSI ISG MEC to 3GPP) is confirming the industry interest on Edge Computing as a key technology enabler for 5G systems [89]. In particular, ETSI MEC specifications are defining a standard architecture [90] and a set of interoperable APIs introducing cloud computing capabilities at the edge of the network. This SDO started its work at the end of 2014, initially focusing only on mobile access networks, and increasing in 2016 the scope to multi-access, thus covering now also WiFi and fixed access. On the other hand Edge Computing is clearly part of the Release15 standard for 5G system architecture (as defined in 3GPP specifications [91] and [92]), and further work is expected soon for a holistic definition of management and orchestration aspects, also in NFV environment.

Still there are some open research aspects on MEC that should be studied, also to better drive future definition of communication systems beyond 5G. Some examples of topics are:

- Application offloading in advanced (multi-vendor, multi access) MEC systems.
- Computationally-aware cell association, i.e., evaluation of the proximity of processing resources in multi-RAT network deployments.
- Separation of data and application instances, and advanced edge computing aspects for mobility.
- Micro services, containers, integration of MEC with fog for different verticals.

Current solutions on application offloading do take into account the communication demand of the apps to be offloaded and the different communication resources available, and do exploit the opportunities for joint optimisations when performing the application offloading at higher network layers down to radio resource allocation (RRC) [93] and [94]. Moreover, the standard in MEC is already defining application offloading as a key use case in ETSI MEC phase 1 specifications (see [95] and [96]), where the MEC host executes compute-intensive functionalities with higher performance compared to mobile devices, improving user experience, and where consumers can use low complexity devices by off-loading compute capacity to the MEC host (Figure 3).

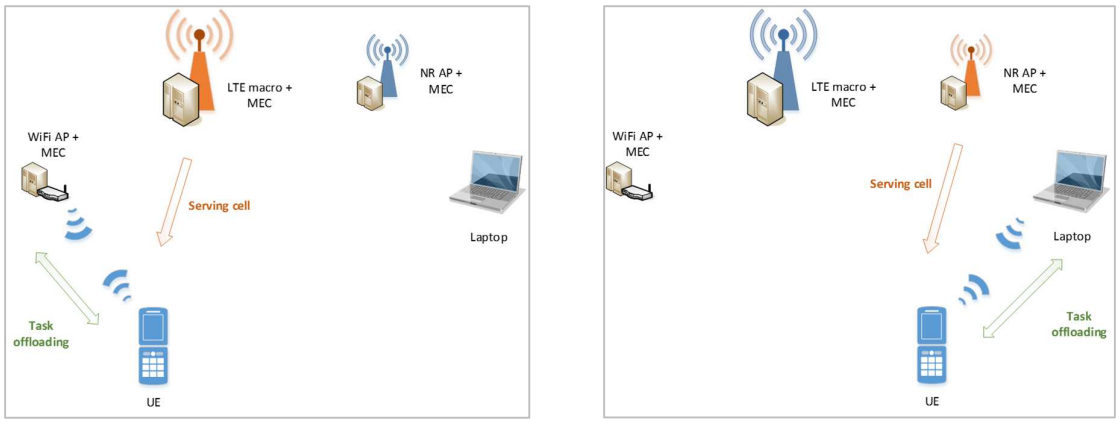

Figure 3 Examples of Task offloading opportunities

Nevertheless, many research aspects should be still studied, e.g. covering multi-vendor and multi access aspects, and from a holistic energy efficiency perspective. In fact, from one hand future systems will become more heterogeneous and complex (integrating many access technologies and networks); in addition, the evolution of communication systems poses increasing challenges from the energy point of view, as in perspective, computational tasks

will become very challenging at the terminal side, in particular from an energy consumption perspective (as an example, lets' consider a terminal connected with multiple RATs (WiFi, LTE, NR); and an increasing complexity of PHY/MAC for CA (Carrier Aggregation) above 5 carriers including LAA (Licensed Assisted Access), LSA (Licensed Spectrum Access), LWA (LTE-WLAN Aggregation) etc.). Also, from the application point of view, computation is increasing, but batteries do not evolve with the same speed. As a consequence, the distributed computing paradigm offered by MEC should be better studied, e.g. by offloading computationally demanding tasks in the terminals while taking into account energy consumption (of both computation and communication) and exploiting multiple RATs, in order to find further opportunities to offload in principle and kind of computational tasks, not only toward the MEC host but potentially among the different terminals (e.g. network functionalities, processing, and offloading coding/encodings, or differentiating traffic between NRT RATs to RT-RATs).

To evaluate the above-mentioned offloading opportunities, when a multi-RAT network deployment is assumed, the latency and energy-efficiency of the applied user-cell association rule needs to be revisited. In a typical communication system, cell association (initial selection and re-selection) is done by considering only the radio propagation and capacity aspects (e.g., biased or unbiased RSS / path-loss / distance-based cell association). Nevertheless, when a MEC system is deployed at the edge of the network, the application layer end-point (i.e. the MEC application instance) is hosted in the MEC platform, which can be collocated with the radio access point (AP) (e.g., eNB, gNB, WiFi AP, CRAN aggregation point, etc.). In these cases, usually, the MEC platform is running on top of a virtualised environment, together with the other Network Function Virtualisation (NFV) functions, sharing the same cloud resources. As a consequence, a performance-enhancing cell association rule should jointly take into consideration communication and computation resources, both static and dynamically changing – load dependent ones.

In addition, separation of data and application instances is a key aspect for the evolution of the network infrastructure. A clear example can be provided by automotive use cases, where the application of Multi-access Edge Computing (MEC) and the concept of Fog Networking are mixed together in the field of 5G automotive communications. The issue is to bring the application endpoints as close to the vehicular environment as possible in order to enable ultra-low latency and high bandwidth services, while considering also huge amount of data transfer in the system. The challenge is thus to allow for a spatial separation of i) Information source (e.g., a remote car providing sensor data), ii) the Information aggregation and processing and iii) Information sink (Figure 4):

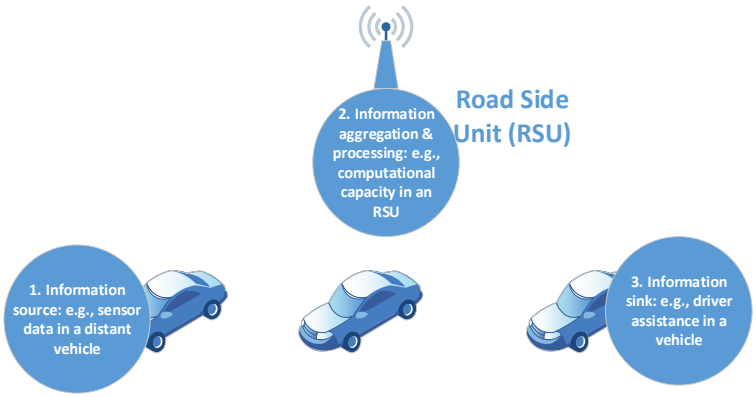

Figure 4 Automotive use cases

Therefore, future research and innovation should also focus on the separation of Data-, Application- and User-Objects in MEC / NFV environments, where these objects may be spatially distributed about network/roadside infrastructure and vehicles themselves.

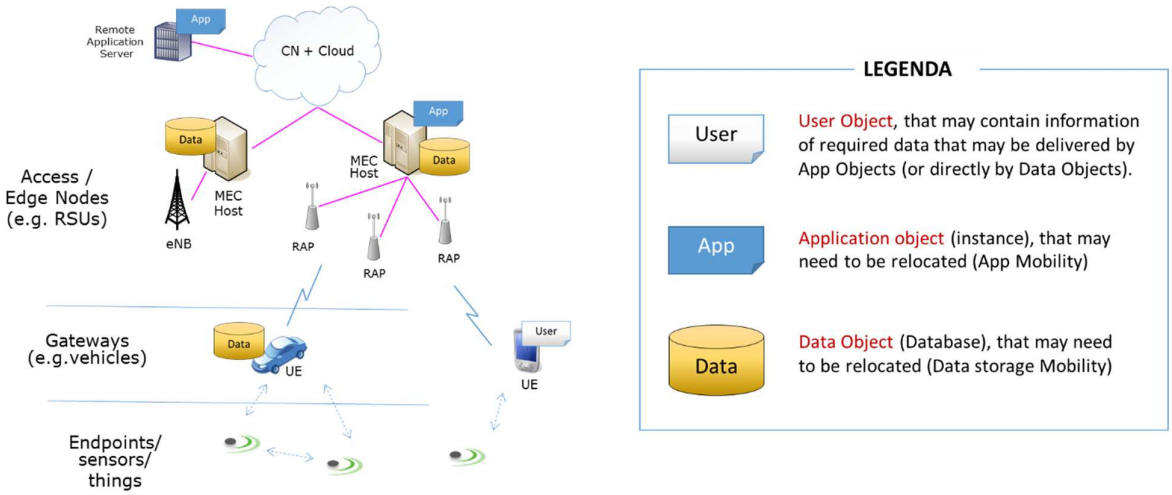

Figure 5 Example of system architecture for 5G automotive Communications based on Separation of Data-, Application- and User-Objects in MEC / NFV environments

Figure 5 is showing an example of a system architecture for 5G automotive Communications based on the separation of Data-, Application- and User-Objects in MEC / NFV environments, even if in principle this aspect should also be studied for other vertical segments (e.g. industry automation), that are expected to enable a mature 5G deployment and potentially drive future evolution toward the future generation. In particular, both academic and industrial communities should address the questions on i) how to design corresponding objects (User, App, Data) in accordance with current MEC 0 and 5G standards [91] and [92] (possibly by foreseeing impacts on future standardisation work) and ii) how to perform suitable “pathfinding”, i.e. how to ensure that the objects are suitably connected to each other in a dynamic environment.

Moreover, as many research trends are already now considering Edge Computing implementation in VM (Virtual Machines) and containers, some further elaboration considering lighter solutions for micro services should be studied, e.g. by better integrating MEC with fog concepts for different verticals. Two key aspects should be considered as drivers for the evolution toward NGI (Next Generation Internet): the increase of number and heterogeneity of devices, and the continuous need to cope with security and privacy. In particular, from one hand the implementation of many vertical segments will introduce new devices (vehicles, drones, wearable, cameras, robots, smart objects …) that will need to be connected in an integrated MEC/fog environment. On the other hand, the definition of advanced solutions for security and privacy will be continuously needed, as huge amount of data will be exchanged in communication systems with an always-increasing level of complexity.

5.2 Future Directions for Fog Computing

5.2.1 Cloud Computing: Friend or Foe?

Cloud computing is playing an increasingly central role for businesses as well as individuals. Cloud computing allows individuals and enterprises, to consolidate, transform CAPEX into OPEX, and access an elastic fabric of computing and storage. It all sounds so nice, but the story is far from being idyllic, let’s try to understand why.

First and foremost, the main cloud providers are non-European. Additionally, these companies have a market penetration and a regional coverage that gives them an unfair and extremely hard to match advantage. As a result, these companies are storing and using – as we all learned from recent scandals, not always as we expect – personal and commercial data of European citizens and companies. In the age of data and information, the ability to gather,

analyse and exploit this astonishing large data sets keeps increasing the competitive advantage of non-European cloud providers.

Second, the raise of Consumer Internet of Things (CIoT) is further propelling Cloud Computing as cloud-centric architectures make it relatively simple to quickly create Consumer IoT applications. But once again, the companies that profit the most for this new raising economy are non-European cloud providers. The biggest win is not the short-term increase in revenues but the massive amount of data that they continue to accumulate and analyse.

Third, data centres deployed by today's cloud providers use an incredible amount of electricity and produce incredible heat. This is not just a consequence of their scale, but also stems from the data-centre philosophy on assuming that hardware is cheap and problems can always be solved by adding more hardware as opposed to optimising.

Thus, Europe should carefully ponder on the strategy going forward to get back in control of data and to reduce the dependency from non-European companies in this strategically important field. **Fog computing** is a potential means for that objective.

5.2.2 Fog Computing

The short-comings of cloud computing started emerging with the raise of the Internet of Things (IoT). Early applications adopted cloud-centric architectures (Figure 6) where information collected from things is processed in a cloud infrastructure and decisions are pushed back from the cloud to things.

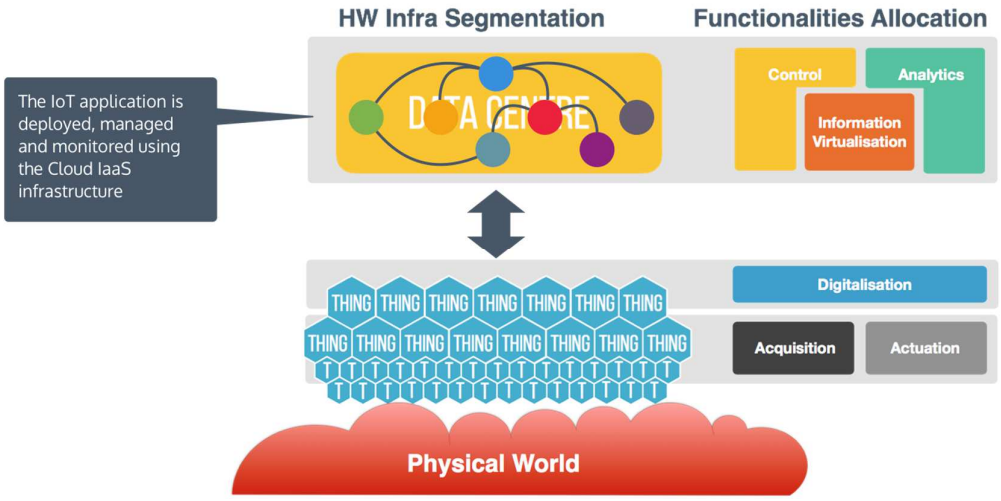

Figure 6 Cloud centric architecture.

However, the cloud-centric paradigm quickly shows its limitation in the context of Industrial IoT (IIoT) and in general with Cyber Physical Systems (CPS). More specifically, the following assumptions, at the foundation of cloud-centric architectures, are generally violated in IIoT applications:

- **Connectivity:** Cloud-centric architectures assume that things are sufficiently often connected. While this is mostly true for IoT applications, it is far from being the common case in IIoT applications. As an example, autonomous agricultural vehicles, or robots in a smart farm are often deployed in locations with very poor connectivity.
- **Latency:** Cloud-centric architectures assume applications can tolerate the latency associated with pushing data from things to the cloud, processing information on the cloud and eventually sending back some control information. This latency is orders of magnitude higher than the reaction times required by several IIoT applications, such as, autonomous vehicles, smart factories and smart grids.

- **Throughput:** Cloud-centric architectures assume that the throughput required to push data from things to the cloud may be massive when looking at the aggregate traffic, but it is generally composed by limited individual flows. In IIoT the situation is quite different as, often, there are data flows with high individual throughput and in several applications the aggregate volume is incredibly high. This makes it unfeasible or not very effective to stream these massive volumes of data to a data centre.
- **Cost of Connectivity:** CIoT commonly assumes that the cost of connectivity is negligible. This stems from the fact that consumers pay for connectivity – either via their mobile data plan or their home Internet connection. In IIoT the situation is completely different, it is the owner of the system that pays for connectivity. This cost is non-negligible in applications with large number of data streams, such as Smart Grids as well as for applications deployed in remote areas such as oil exploitation rigs that can only rely on expensive satellite communication where 1 MByte of data can cost as much as $ 8!
- **Security:** Cloud-centric architectures operate under the assumption that end-users are comfortable in giving away their data. While this may be true for CIoT applications, the situation is completely different in IIoT. Information is a strategic asset which the vast majority of companies, operating in an industrial, do not want to leave their premises.

Fog computing has emerged as an architectural approach to deal with the limitations exposed by cloud-centric architectures in the context of CPS and IIoT applications. These challenges faced by cloud-centric architectures are the opportunity for the EU and EU-based companies to innovate faster than the non-European cloud providers and play a central role in the future of CPS. Considering that CPS and IoT are projected to become a huge market, the EU should focus its resources on competing on this blue ocean as opposed to trying to catch up with non-European cloud providers.

5.2.2.1 What is Fog Computing?

Fog Computing aims at providing a virtualised infrastructure that allows to distribute computing, storage, control and networking functions closer to the users along a cloud-to-thing continuum (Figure 7). From this description it should be clear that **Fog Computing** can leverage cloud infrastructure or equally function without it.

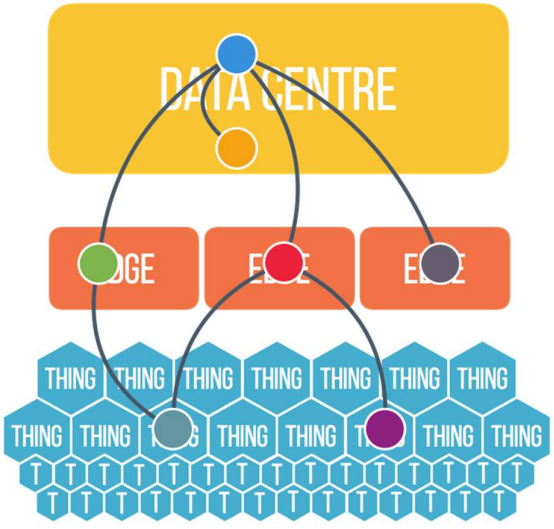

Figure 7 Fog Computing end-to-end resource virtualisation

Fog Computing makes available a set of abstractions to unify the compute, storage and communication fabric end-to-end, thus allowing applications to be managed, monitored and orchestrated across the cloud to a thing continuum.

In other terms, fog computing provides the same *abstraction* of an elastic compute, storage and communication fabric but in a decentralised manner. The main advantages of this architectural approach when compared with cloud computing is the ability to keep the data and the processing close to where it makes the most sense, which in many cases is close to where the data is produced.

5.2.2.2 Fog Computing and Multi-Access Edge Computing (MEC)

The concept of fog computing was introduced in the context of Cyber-Physical Systems (CPS) while Multi-Access[2] Edge Computing (MEC) is a very similar concept that emerged in the context of 5G architectures. Fog Computing and MEC share most of the requirements but not all. One of the main distinguishing factors of Fog Computing, when compared to MEC, is the need for supporting exotic I/O and accelerator aware provisioning, real-time, embedded targets as well as real-time networks such as TSN (Time Sensitive Networks) (e.g. IEEE 802.1).

Thus, most of the innovations and technologies implemented for Fog Computing will be reusable for MEC and vice-versa. Considered EU's historical strengths in telecommunication and embedded systems this is another reason why the EU should focus on this domain.

5.2.3 Fog Computing Research Directions

The non-European cloud providers are starting to position technologies that try to alleviate the problems of cloud-centric architectures described above. Examples are Amazon Green-Grass or Microsoft Edge solution. That said, the non-European cloud providers are not tackling the problem structurally, are just trying to provide point-solutions to alleviate some of the problems. Fog computing requires an infrastructure that is quite different from those found in data-centres. It poses different challenges and provides various opportunities for innovation. Below is a list of innovation actions that the EU should consider to competitively position in the Fog Computing market.

Action 1: Fog Computing Infrastructure

Fog Computing will be as successful as open and interoperable will be the foundational infrastructure. As such the very first action that the EU should undertake is an innovation action around the creation of the European Fog Computing platform. This platform should be designed ground-up with decentralisation, efficiency, interoperability, security and extensibility in mind. Here the EU commission should look for disruption as opposed to the usual incremental innovation over data-centre technologies such as Open Stack or other data-centre technologies.

Action 2: Fog Computing Platform Services

To propel the adoption of Fog Computing it is essential for some of the key platform services such as, data sharing, storage and analytics to be available. As such the EU Commission, along with the action around the Fog Computing Infrastructure, should create an action focusing on the creation of platform services designed ground up to address the decentralised architecture typical of fog deployments. Once again, this will require disruption as bulk of main stream technologies in data sharing, storage, and analytics are data-centre oriented.

[2] Originally named Mobile Edge Computing (MEC)

Action 3: Fog Computing and MEC Convergence

The EU should acknowledge that fog computing and MEC share most of the requirements, thus as much as possible should try to focus the investment on infrastructure that can address both. Most of the time the main barrier between communities is the terminology, thus one of the action that could be taken by the EU Commission is to unify Fog and MEC and ensure that the community build synergy and critical mass.

Action 4: Fog Computing and Blockchain

Fog computing is about decentralisation but also about resource sharing among different entities. For resource sharing to be successful some form or compensation should be devised for those who lend resources. The Blockchain could be the trusted distributed ledger that records these transactions. But there is another link between fog-computing and the Blockchain. The Blockchain was introduced as a way to provide a decentralised ledger, but the trend now is to centralise the servers in the cloud. This is against the philosophy and the original motivations for the Blockchain. Thus, another synergy between fog-computing and the Blockchain is that the former can be used to provide the decentralised but virtually unified compute, storage and communication fabric on which to deploy a Blockchain. An action should be envisioned to explore the synergies between fog computing and the Blockchain.

Action 5: Dynamic Resource Management and Monitoring

Dynamic resource management under constraints is an NP-Hard problem (a class of decision problems in computational complexity theory, NP - **n**on-deterministic **p**olynomial-time). There is background research on the subject, but very little of it has focused on the problem we will have to deal with fog computing. Thus, another action should be undertaken to research decentralised algorithm for allocations with well understood competitive ratios w.r.t. the optimal solution.

5.3 Massive IoT Services

Massive IoT generally refers to applications that require a huge volume of low-cost and low-energy consumption devices [97]. Such devices are usually deployed in large areas and connected using wireless technologies. Their architecture is characterised by reduced computation and storage capabilities, which significantly affect the set of functionalities that can be implemented.

Over the last decade, machine-to-machine communication has mostly focused on connecting relatively few endpoint sensor and/or actuator devices to central servers, but more recently business IoT projects have started growing from small pilots to global rollouts of tens of thousands of connected devices. The trend is clearly set: 12 % of companies embracing IoT now have at least 10,000 connected devices, according to a state-of-the-market report produced by mobile network operator Vodafone, and the proportion adopting IoT on a massive scale – over 50,000 connected devices – has doubled in the past year, from 3 % to 6 % [98].

Massive deployment of constrained IoT endpoints still raises however several challenges that can only be partially handled by current technologies and solutions, and therefore will need to be further addressed in the coming years, also considering the new envisioned IoT scenarios and applications.

5.3.1 Critical IoT services

The integration of endpoint IoT devices is usually a key enabler to introduce functionalities such as remote control and information exchange across different systems. Autonomous Vehicles (AV) or Unmanned Aerial Vehicles (UAV) solutions, for instance, are already being experimented as self-driving cars or autonomous drones and are all equipped with sensors and actuators interconnected among each other for coordination and control. On the other

hand, Virtual or Augmented Reality systems are at an early stage available as simple system such as smartphone applications. Though today such systems are at an early stage of experimentation, the need for their large-scale adoption in a near future has been already envisioned. As an example, there are many initiatives aiming at reshaping the future of urban mobility with massive adoption of AV solutions [99].

However, mass-scale deployment of such systems will represent a significant challenge for the network and IoT infrastructure. In many contexts, best-effort data delivery is sufficient. In others, instead, the need for real-time interactions between the services and the devices arises. These critical applications are characterised by strict Quality of Service (QoS) requirements, particularly in terms of latency and reliability, as a mandatory enabler.

Smart manufacturing systems [100], for instance, include closed-loop control and real-time analytics services [101] that require ultra-reliable and low-latency communications to deliver telemetry data and control commands, from sensors and to actuators deployed in the assembly line, respectively. Latency higher than a few milliseconds would result in the whole system to enter an emergency shutdown state, which might cause substantial financial repercussions. Similarly, remote monitoring applications of *smart health systems* do not only require bounded latency and reliability but also strict data prioritisation to differentiate periodic reporting data from critical information. The latter functionality allows prioritising critical data, such as alarm data, that require immediate delivery and process. *Multimedia Sensing* [102], such as smart surveillance systems or augmented reality systems, will demand for bandwidth requirements, to ensure the timed delivery of multimedia content traffic, characterised by a data rate significantly higher than the typical IoT applications. Future Smart Vehicles systems [103] will have, at their cores, services prevent collision or share emergency information. In this case, to guarantee low-latency and reliability, the network must implement advanced mobility management strategies, thus to ensure seamless connectivity.

Current research and standardisation initiatives on network technologies and communication standards are already focusing on **supporting real-time edge-communications**, e.g. 5G standardisation activities and the research efforts referred to as tactile Internet. All such initiatives aim at winning the so-called “1 ms-challenge” [104] for ultra-low latency applications, i.e. ensuring a round-trip latency of 1 ms with an outage of 1 ms per day in the communication between the service and the IoT device. Although such latency can be only achieved by running the services on edge computing systems in proximity of physical systems, thus to reduce the communication latency, the near real-time requirements of the above-mentioned applications require explicit support from the network, beyond the capabilities of the current technologies [105]. Although 5G already aims at handling such challenge at the core of the design of its architecture and within the physical layer, supporting ultra-low latency on a large-scale still presents many open challenges [106].

The large-scale integration of very constrained devices, characterised by minimal computational and storage capabilities, will also represent a major challenge for critical IoT systems. This will demand for **novel QoS enforcement techniques to consider by design the minimal memory available for data buffering and data prioritisation**. A similar challenge is expected to ensure support for Multimedia Sensing applications. Considering that recent communication standards already introduced support for multimedia applications, introduction of large-scale multimedia traffic will require for novel quality of experience paradigms for IoT multimedia flows [102] and novel multimedia compression solutions to handle the reduced capabilities available on IoT devices [107].

The achievement of the 1 ms challenge will also require the optimisation of many aspects of the communication system, which received less attention so far. The analysis and process of packets within the network systems, for instance, will require novel approaches to minimise the processing latency. Moreover, **proper management of mobility to ensure seamless connectivity with stringent QoS requirements** will represent another critical issue that cannot be solved with current mobility management solutions. Future mobile system that rely

on communication to ensure basic functionality and safety will require novel handoff strategies to guarantee timed and reliable communication, regardless of high speed and large areas to cover.

5.3.2 Scalable management of massive deployments

Massive IoT deployments are anticipated in many different industry verticals such as smart city, building automation, e-health, smart energy, automotive, manufacturing, agriculture, etc. These areas have many diverse requirements in terms of, e.g., *mobility* (frequent handovers with guaranteed QoS for the automotive and manufacturing industries), *self-organization* and *autonomy* (to reduce the need for humans in the loop and at the same time to cope with situations when the connection to a central service is or becomes unavailable), *local connectivity* (to enable local routing of data to achieve ultra-low latency in case of alerts), and *security* (to arrange sophisticated and adaptable policies for data protection and privacy).

On the other hand, these verticals have also in common some key requirements. In particular, they all require **efficient and scalable operations management of connected devices**, as well as the management of their connectivity over a variety of network connections. Analysts predict that by 2021, 5G's broad enablement of IoT use cases will drive 70 % of G2000 companies to spend $1.2 billion on connectivity management solutions [108]. Connectivity management will involve many device-specific aspects at different layers of the protocol stack.

A recent analysis [98] reported a substantial growth in connectivity requirements for IoT deployments, and registered that typical large-scale deployments use up to four different technologies for connectivity, with mobile and WiFi the most prevalent, but also including newer ones, such as NB-IoT or LP-WANs. **Scalable connectivity management and operation over heterogeneous technologies in massive IoT deployments** is clearly one of the open challenges that need to be dealt with. In tomorrow's 5G networks, one key enabler to tackle this challenge will be the widespread use of Network Functions Virtualization (NFV) technologies [109]. NFV allows for easily composing highly customized network functions, including connectivity management, in a cost-effective and vendor-independent manner. NFV will therefore let network operators to implement custom network functions in virtual *slices* of their access networks, i.e., where IoT endpoints are attached to, and provide them "as-a-service" to vertical IoT industries. Current NFV paradigms and frameworks need therefore to be evolved in order to support and integrate multiple heterogeneous 5G vertical infrastructures and service platforms for massive IoT deployments in a scalable and highly automated manner [110].

Another common requirement to massive IoT deployments is that IoT devices require scalable remote configuration and control capabilities. Connected devices will need to be switched on and off adaptively, configured for a specific network bearer, provisioned for services, monitored and maintained, location-tracked, etc., and all this must be performed remotely and in a device-specific manner, for scalability of system operation [111]. As an example, in energy-efficient IoT infrastructures, sensor devices need to implement optimized policies to run important sensing when needed, e.g. during specific data periods (the peak traffic hours when pollution is more a threat) or based on measurement thresholds that are set remotely, and then go to sleep disconnecting from the network otherwise. In automotive, device's location has to be provided to a tracking centre with an adaptive frequency that needs to be managed remotely.

To **remotely convey device-specific capabilities and perform management of a wide range of constrained devices** is therefore another key challenge for future massive IoT-deployments. A key enabler for IoT tailored device management is the standardization of application-layer protocols and IoT-specific data models for remote configuration and control, which reduces the degree of fragmentation in heterogeneous deployments while providing for interoperability and extensibility. A notable example of these standardization efforts is OMA Lightweight M2M (LWM2M) [112], which leverages the IETF Constrained Application Protocol

(CoAP) for data transfer to and from low-cost and low-energy connected IoT devices. Although tailored to IoT environments, it is still an open issue to what extent such approaches will scale in massive IoT deployments, considering the impact that carrying device management messages will have on the underlying network infrastructure.

In addition to device management, and as an extension to it, both firmware upgrades and **automatic deployment/provisioning of the application logic running on IoT devices** will be a further requirement for massive IoT deployments [113]. This means that devices can be upgraded and provide new capabilities as a result of dynamic over-the-air service provisioning and enablement from their management servers. As an example, they could be updated by receiving software packages containing protocol and network functions dynamically adapted to the changing surrounding environment in which they are deployed. This needs again to be accomplished in a scalable manner, from the network infrastructure point of view, as the same upgrade/service provision might be needed for a large number of devices at the same time.

Advanced software development techniques, such as *Infrastructure-as-Code* (IaC) [114], could help supporting the quick development and delivery of advanced IoT applications across multiple vendors of IoT products: by focusing on the **automation of their provisioning, deployment and operation management in cloud/fog/edge infrastructures**, it might be a promising approach towards scalability.

Last, but not least, it is worth to recall that in many areas IoT endpoints are tiny devices with (sometimes very) limited capabilities in terms of processing, memory, and energy. Therefore, all adopted solutions across the protocol stack, from low-layer data-link, network and transport protocols, to application data encoding and encryption, need to be **carefully designed in order to be low-energy, with small processing and memory footprints, and highly-efficient over bandwidth-constrained communication links**. This is a fundamental requirement in order to prolong the device lifetime to its maximum extent in large-scale deployments where it is expected that no, or very little, human intervention will be possible [115].

5.3.3 Distributed/autonomous and cooperative computing

Self-organisation and autonomy will be crucial requirements in massive IoT. Such features allow on one hand to minimize the need for humans in the loop and, on the other hand, to loosen the dependency from cloud computing. According to Gartner's estimate, the number of IoT devices will reach by 2020 approximately 20 billion [116], thus requiring 1000 times cloud and networking resources consumed by IoT devices in 2016. For this reason, the current centralized model based on cloud computing will demand for novel distributed approaches to handle the exponential increase in terms of resources [117] and cope with situations, such as emergencies, in which the connection to the cloud is or becomes unavailable. To this aim, future IoT devices are expected to implement novel solutions to de-centralize decisions and data analysis.

Recent research activities have focused on the definition of distributed approaches to enable **secure cooperation of devices to reach distributed decisions**. Many approaches are based on the **blockchain** technology, which enables verifiable transactions and smart contracts through a distributed ledger stored on a network of trustless devices [118]. In these systems, the direct integration of IoT devices into the blockchain will allow IoT systems to make autonomous decisions, e.g. for identity and access management or for automating in a cryptographically and verifiable manner workflows/interactions. The blockchain technology will not only reduce constraints in terms of costs and capacity and increase resiliency from cloud failures; it will also eliminate system susceptibility to manipulation: its decentralized access and information immutability will ensure that malicious actions aimed at information manipulation will be easily detected and prevented. Although the direct integration of IoT devices into blockchain systems is taken for granted – the latest predictions foresee that by 2020 up to 10 % of production blockchain distributed ledgers will incorporate directly IoT

sensors [108] – practical implementations will have to cope with the limited resources available on IoT devices, e.g. by reducing memory footprint and computational overhead.

AI/ML techniques are currently applied in cloud systems for big data mining. Some areas, however, such as transportation, manufacturing, smart building and smart grid, would benefit dramatically from the implementation of data mining techniques directly at the edge on IoT devices. The implementation of data mining techniques directly in proximity to where data is collected and where the physical systems to be controlled are located would enable real-time data analysis, thus allowing to make real-time autonomous decisions without human intervention [119]. Such **edge intelligence** would guarantee minimal latency and a dramatic reduction of data transfer through the analysis of the collected data at the source and by forwarding only elaborated or aggregated data. However, moving the intelligence from the cloud to IoT devices on a large scale at the edge presents many challenges. For instance, current AI/ML techniques are designed to run on powerful environments while edge devices are characterized by limited resources: the implementation of an edge intelligence will require the development of new models to take into account the scarcity of resources at the edge. AI/ML techniques usually require large datasets for training, therefore their implementation on IoT devices will need the optimization of current training techniques or their optimization to minimize their configuration time. An example of research activities that currently cope with these issues is the BONSEYES project [120], which aims at implementing a framework to bring AI/ML techniques on IoT devices.

5.4 Data Analytics and Data Monetisation

5.4.1 Big Data

Modern society is generating data at an unprecedented rate. The progressive digitalisation of society and the world is causing that huge amounts of data are generated continuously (2.5 quintillion bytes of data per day [121]), and the generation rate is continuously increasing. It is expected that when IoT devices are widely deployed this rate will increase exponentially [122].

The availability of huge amounts of data (what is commonly known as Big Data) represents both a challenge and opportunities. It is an opportunity, because there is a huge potential in the possibility of extracting information and knowledge from the data. In fact, there is a growing trend asserting that “data is the new gold” [123] and [124]. However, like extracting oil or gold, extracting information from Big Data is not easy, because its volume does not allow to process or transfer it quickly.

In the last decade there has been two trends related with Big Data processing. On the one hand, many services have been moved to the cloud, which in practice is a collection of large data centres with thousands of servers. On the other hand, the availability of many servers in one single location connected with high-speed networks has allowed the deployment of new data processing and analysis tools and technologies, specifically designed for Big Data processing. Examples of these technologies are the Hadoop (https://hadoop.apache.org/), Spark (https://spark.apache.org/), and Flink (https://flink.apache.org/) systems.

In the near future, Big Data processing is facing a number of challenges, which will have to be faced in the next decade or so. The first and possibly most important challenge is scalability. While the abovementioned systems have been able to adapt and handle easily terabytes of data in data centre environments, as mentioned before the data generation rate keeps growing, and whole new paradigms may have to be conceived. For instance, because its huge volume, it may not be feasible to move the data to a single location to be processed. Hence, the implementation of clouds may have to be extended to consider edge/fog computing, and Big Data processing may have to be distributed.

The distribution of the Big Data processing by using technologies like edge or fog computing has several important advantages over data centres. First, it allows the data to be processed near its source, reducing the amount of data that should traverse the communication networks.

If properly designed, only summaries of much smaller size may have to be transferred. Second, if may potentially reduce latencies, if the users of a service and the data required to provide it are both close in the Internet. Third, it may allow for a reduction of the energy consumption with respect to data centres, if the computational equipment is appropriately distributed (e.g., maybe no cooling is required if severs are not placed together). Finally, this allows the investment in infrastructure if it would be possible to do the edge processing in underused user equipment.

However, the described distributed Big Data processing is still far from mature. New algorithm and technology must be devised and deployed for the summarisation and aggregation of data, for the implementation of distributed queries and databases, and for the distributed computation of functions on the whole data set. From a systems point of view, using virtualisation both at the network level (using SDN/NFV) and at the processing level (using virtual machines and containers) seems reasonable, but requires solving scheduling, optimisation, allocation, and orchestration problems that are not trivial.

5.4.2 Distributed Ledgers

In parallel with the intensive work on Big Data processing, the last decade has also seen the popularisation of Blockchains or distributed ledgers (DL). Although DL have been usually viewed only as a support for crypto currencies, like Bitcoin or Ethereum, they are potentially much more powerful and versatile, and can be a useful tool for Big Data analysis (among other applications). A distributed ledger provides a data structure to which only new transactions/records can be appended, guaranteeing a globally uniform view of its state. This can be a powerful communication primitive for a distributed data processing system.

Additionally, if the distributed data processing is deployed in external equipment, the flexible economic primitives that can be built on top of crypto-currencies and smart contracts may allow exploring multiple types of incentives for the infrastructure provider. For instance, final users may allow network operators to use their computers to run a process in a docker container in exchange of a few units of crypto-currency, where the price paid has been fixed by an auction system implemented with a smart contract.

Although there is a huge hype about distributed ledgers, these are still not practical. The leading Bitcoin and Ethereum blockchains are extremely expensive in resource consumption and have a very low throughput. Unfortunately, no competitor seems to have an optimal solution for the issues these distributed ledgers have. Hence, there is a huge challenge in making distributed ledgers scalable, so they can be widely applied.

5.4.3 Artificial Intelligence/Machine Learning (AI/ML)

One third research and technological area that has seen an impressive advance in the latest years has been AI/ML. This has been due to several factors, among which the availability of large amounts of data is one of the most important. (Others are the improvement of algorithms and hardware support.) With many examples to learn from, certain AI/ML like deep neural networks are able to generate models that achieve an impressive level of accuracy. The smart usage and combination of these features is driving AI/ML to a new golden era.

However, as mentioned above, large amounts of data come with the curse of need for scalability to be able to manage them. Distributed AI/ML algorithms will have to be devised to be able to generate models without having to move the data far from its source.

Another obvious challenge is to clean and properly label the data to be fed for supervised learning. Training data that is not properly classified leads to a bad classifier. In fact, this may go beyond being a technical problem, since it is possible to end up with models that reflect in their classification biases and prejudices, raising moral and ethical concerns. This connects with a recurring issue with most models generated by (ML), which is the impossibility of understanding them. When the model generates outputs for given inputs, it is not possible for

the ML expert to explain why and where in the model these decisions are made. This is a very important problem, because it prevents to know a priory what the reaction of the model under certain situations would be.

5.4.4 Lack of awareness and knowledge in personal data monetisation

The eruption of very popular online services whose business model builds up on the commercial exploitation of personal information, together with several data breaches exposing misuse of users information, has raised a very intense debate around questions like: where are the ethical and legal boundaries in the management of personal information, how to raise societal awareness among average Internet users with respect to their exposure to the exploitation of their online personal information, etc. The World Economic Forum [125] concluded that the lack of awareness of citizens regarding the management of personal information and their increasing concern regarding privacy and data protection was a serious risk for the sustainable economic growth of online services. The last Eurobarometer about data protection [126] reveals that 63 % of EU citizens do not trust online businesses (search engines, online social networks, e-mail services), which makes them the least trusted companies regarding the use of personal date. Also, more than half of the citizens do not like providing personal information in return for free services, and 53 % do not like that Internet companies use their personal information in tailored advertising. This intense debate about the management and exploitation of personal information led to an ambitious regulatory effort in Europe. The European Union has very recently enforced (May 2018) the General Data Protection Regulation (GDPR) [127] that defines a new regulatory framework in the management of personal information.

Although the regulatory effort is necessary, it should be accompanied by multidisciplinary solutions that raise societal awareness regarding the business models of Internet services exploiting personal information. This means, achieve that average Internet Users understand simple elements like: (i) every time they perform a search in Google they are generating money for Google, every minute they spend in Facebook they are generating money for Facebook, every time they look at the news at CNN.com they generate money for CNN, etc., (ii) that money is generated because there are third party companies that pay Facebook, Google and CNN.com to show Internet Users tailored advertising based on the personal information Google, Facebook and other online services know about the users. This pedagogic effort requires a multidisciplinary effort to design novel methodologies that let Internet Users understanding in a simple and practical way complex aspects associated to the way online services exploit their personal information and monetise it.

An efficient way to increase the interest of average Internet users regarding personal data exploitation is informing them of what is the monetary value they generate for online services out of their personal information. In this line, the OECD acknowledged the importance of having accurate methodologies that allow measuring the monetary value associated to personal data to understand its economic and social value [128], but also highlighted that: (i) it is an extremely complex task, (ii) the existing methodologies are still rudimentary, and (iii) research in this area is necessary since it is in a very preliminary stage. In addition, other initiatives such as the Data Transparency Lab [129], a recent community effort that promotes transparency in the management of personal information, has included in its research agenda the necessity of methodologies to measure the value of personal information in order to raise Internet users and societal awareness.

In addition to Internet users, there is an increasing necessity in some industrial players to better understand the monetary value associated to personal information and how that value can be enriched to propose novel sustainable services that monetise data following a transparent, ethical and privacy-preserving approach. For instance, there is a novel business model that aims to include Internet users in the revenue chain generated by the commercial trading of their personal information and improve the quality of the offered service to their customer using more reliable information from users [130].

The described context reflects a common request from different socio-economic forces that urge the definition of innovative, transparent and privacy-preserving methodologies that allow measuring in an accurate way the value associated to personal data transactions started in the edge of the network. A comprehensive methodology will require multidisciplinary knowledge that allows efficiently covering technical, ethical, legal and business aspects.

5.4.5 Fraud mitigation in data monetisation

Online businesses sustained on personal data monetisation are a major social and economic driver of the so-called *Information Society.* For instance, the most lucrative business monetising user data, *i.e.*, online advertising, sponsors free essential services to billions of users, such as Online Search Services, Map Services, and Social Media. The market volume of online advertising reached, only in the US, $72.5 billion in 2016 with an inter-annual growth rate of 22 % [131]. Furthermore, online advertising represents an important source of jobs. For instance, recent studies have estimated that 0.9 million (0.4 %) direct and 5.4 million (2.5 %) indirect jobs were associated to online advertising in the EU-28 workforce in 2014 [132]. Therefore, it is in the best interest of everyone (citizens, governments and the private sector) to guarantee the sustainable growth of online businesses that monetise personal data (subject to the strict application of the General Data Protection Regulation and ethical practices). However, this sustainability is in jeopardy due to the presence of FRAUD.

There is an increasing trend of attackers that exploit the ecosystem in fraudulent ways to obtain an economic benefit. For instance, in the area of online advertising, a fraudster may set up several websites and lease ad spaces from these websites to one or more ad networks. Once this infrastructure is set up, the fraudster configures bots (automatic software) to visit these websites and eventually clicking in some of the ads. These actions of the bots translate into monetised ads that produce a profit for the fraudster, but also for the intermediaries intervening in serving those ads. Some estimations quantify ad fraud only in US in more than $ 8 billion in 2015 [133]. If fraud is not mitigated it may eventually persuade advertises, who are directly sustaining Internet innovation by paying most popular online services, to dramatically reduce their investment in online advertising due to the large presence of fraud.

It is proven that different criminal activities are totally or partially funded through online advertising, which may be using fraud techniques to increase the profit obtained from organic traffic (i.e., visits from real users). Examples of this are:

- Intellectual Property: Violation websites, also known as pirate websites, offer free access to copyrighted protected content. This is considered a cyber-crime by the European and National legislation of the EU-28 countries. These websites obtain a profit through online advertising, by showing ads to their visitors [134]. Note that these sites may use ad fraud techniques to increase the profit obtained from organic traffic (i.e., visits from real users).
- Extremist, Hate of Speech and Terrorist-related websites, which are a critical channel for both propaganda and recruitment for criminal organisations in these contexts, obtain funding through online advertising (e.g. [135]).

This situation requires to promote methodologies that mitigate fraud in online advertising and other data monetisation services in order to guarantee the Internet innovation in the long-run, and in turn, all associated benefits such as millions of jobs in EU-28.

6. Network and Service Security

6.1 Security Transformation

The current and expected network models' evolution is a key rational for Security transformation. From system or even System of Systems point of view, we have simultaneous disruptive changes both in time and space leading to so-called metamorphic properties. One can easily conclude that applying static security solutions to a problem where dynamics are becoming predominant will fail. From data plane towards service (and their usages) plane, security functions deployment must adapt to the temporal properties. Another historical basic security paradigm is becoming inefficient with the dispersed capabilities across a hybrid computing/network model, which is a perimetric approach. Among the ICT challenges, security is a fundamental cornerstone, but security must in turn transform towards renewed paradigms.

By mimetism, an obvious fundamental dimension of the security transformation may be described as Software Defined Security. This means that flexibility, spatial/temporal distribution of functions, adaptive capabilities, models and abstractions, etc. apply to the security domain. It also induces that control and management technologies should remain in a converged framework, optimising re-use to reduce skills requirements, complexity and OPEX. This expected concomitant evolution disrupt traditional approaches where security concerns often came afterwards (or too late), it even goes beyond the "by-design" paradigm with functions and services integrated as intrinsic component of the concept. Last but not least, fundamental science advances whether physical or logical will equally impact traditional network functions and security functions. Security will then have in addition to provide solutions to secure the functions themselves. While the need of transformation is considered, it comes with a major challenge of definition of integrated renewed security framework, processes, engineering able to provide relevant answers to the ongoing disruption of architectures and technologies.

Our society is more and more depending on the communications and ICT infrastructure. As a result of the ever-growing demand from people, but also Business-to-Business or Machine-to-Machine, the complexity/manageability issues are becoming predominant while introducing any innovative or disruptive technologies. Cybersecurity that often requires a significant level of situation awareness is very sensitive to those dimensions. As the sustainability of ICT mandates technology introduction, it will translate in focused security challenges that should be considered from start. Considering addressing, routing area, although combined with ciphering spread everywhere, a major issue is the ability to authenticate packets (fields), flows, applications, services, users, etc. as it should be. Among other requirements to protect for instance slice integrity and isolation across multi-owned infrastructure segments. This is even more sensitive looking at control, management, service planes exchanges, leading to a potential specific dedicated approach such as a super secure slice. Finally, this should be done while protecting privacy and confidentiality. Solving the mismatch between historical approaches and security in the new picture is somehow mandatory but a difficult challenge.

Another major area advocating for security transformation is linked to old issues of distributed systems. The overall architecture will rely on the composition of modular systems and services, potentially mission-oriented. As a direct consequence, the scheduling of resources must be mission-aware but also security-policy aware. Security here is raising numerous challenges, the first set is related to policies themselves: how to declare those policies, how to compute compliant resources and services with the policy constraint, the second part is related to the security enforcement deployment: how to schedule (and guarantee) security functions and associated resources.

Beyond the required overall framework, security transformation is the mandatory companion to most of if not all foreseen disruption dimension. On radio side, where sophisticated mechanisms may increase the vulnerabilities, programmability leading to a new range of

potential attacks, slicing, operating systems, quantum and AI/ML as attacker tools, data at rest, sharing, computing, etc. transformed security must be embedded and considered together at initial stage with all network visions.

6.2 Network-wide Security

The recent trend in network attacks has demonstrated that the network as a whole is both a target and a vector of attacks. The Mirai botnet has demonstrated the capability to knock out even large organisations of the Internet. The Wannacry worm has demonstrated the reach and extend of networking connectivity, connecting every device to the Internet and exposing its vulnerabilities to the whole world. Malicious applications are increasingly present in application stores, providing computing power, storage and over all access to sensitive information to attackers, free of charge.

At the same time, attack detection and mitigation has progressed but remains one step behind. Anomaly detection has not resolved the false positive issue that is making it extremely complicated to handle outside of very structured environments.

Therefore, the first challenge of next generation networks is to include protection and resilience by default, deeply rooted in their architecture, so that attacks become harder to carry out, less effective impact-wise, easier to attribute. This implies that the management infrastructure becomes more resilient to attacks, against configuration protocols, routing and naming. This requires that trust anchors are put in place and resilient configuration patterns deployed, so that networks or network overlays can resist attacks, and that it becomes impossible for attackers to abuse the network to inject malicious management traffic. From a non-technical standpoint, it also means that it should become more expensive for attackers to attack, and more expensive for developers to create vulnerable applications than safe ones.

When protection is not feasible, because of practical or economic considerations, the next best thing is to ensure that detection of attacks occurs. Detection must be timely (in the sense that attacks should not have created damage by the time the alert is handled by the security operating centre) and accurate (provide limitations on false positives and false negatives). It should also be actionable, leading to either the elimination of the source of the attack, or to a limitation of its effects.

Another challenge is the capability to better understand attackers, to enhance detection. Beyond honeynets and honeypots, there is a definite need to improve the traceability of activity occurring in the network. This requires the ability both to generate better logs, to make it more difficult for the attacker to hide in the flow of ”normal” traffic, but also the ability to include tripping points in the network, that will signal anomalous activity. This will support the development of better detection and deterrence methods, reducing the attacker’s gain and increasing its risk.

The last challenge is the coupling of network and application security. While network security is handled independently of application and services security, there is a high likelihood that certain network attacks occur because of application or service vulnerabilities, or that network insecurity impact services. This challenge deals with the coupling of network and application modelling, to support more effectively anomaly detection than has been possible before. Only by getting more and better data as input for activity models will we be able to leverage AI/ML techniques to build efficient anomaly detectors, which are able to include contextual information in alerts to support threat mitigation.

In the end, the next expected result should be that networks are easier and cheaper to operate for network operators. Regular users should get access transparently and their services must be available with appropriate quality. Attackers should have a much higher cost of attack as well as a much higher probability of being caught.

6.3 Slice-Specific and Convergence on Common Software Defined Patterns

Future ICT cybersecurity is the field of numerous challenges due to the change in technologies, architectures and paradigms but also will widely re-use the advances in network technologies [136] and [137]. The slicing aspect is here used to illustrate directions, combining Software Defined Security and Security as a Service challenges that apply also beyond slicing aspects.

The network-slicing concept implies de-facto some sort of sharing of the control/management with the underlying infrastructure. The immediate obvious security issue is related to the isolation between slices but already in [138] and [139] NGMN listed a set of key security issues beyond basic isolation concerns. Thus, as of today, the 5G slicing security is one security area among many other specific issues that are already identified. As a continuity of 5G, Next Generation Internet will encompass non-3GPP domains, extended roaming procedures, Service Based Architecture (SAB, including Security as a Service) and will certainly bring unprecedented architectures, services or business models. Considering the 5G architecture and procedures introduced in [91]and [92], security areas are more specifically addressed in [140] and [141]. Going far beyond 4G complexity, 5G and beyond is imposing to reconsider many security aspects such as Authentication and Authorisation, RAN (multi-access) security, User Equipment (including IoT), confidentiality and key management, etc. and finally Network Slicing security. An overall vision of this security landscape is given in [142] where one can find references to groups and bodies active on 5G security such as ETSI – in particular working groups dealing with Network Function Virtualisation (NFV) and management issues (MANO), IETF, IMT-2020, etc.

The nature of 5G and beyond components, systems and services lead to an unprecedented combination of specific software-based vulnerabilities, function distribution, boundaries variations in time and space. Moreover, the multiplicity of stakeholders and authorities in the case of network slicing raises serious challenges which are exacerbated by the so-called Mission-Critical support. Through technological but also architectural and business aspects, or even regulation, the slicing shows a novel attack surface but also great opportunities to deliver the relevant level of cybersecurity.

The slices are software-based and as such, inherit among others from SDN and NFV security issues and solutions. Many threat Intelligence aspects are already addressed in the literature including a comprehensive survey [143] or, for instance for SDN-NFV components, available as Common Vulnerabilities and Exposure (CVE) list. As common ICT pattern, the code life cycle, the virtualisation will continue to be the subject of research from the security angle.

As described in [91] 5G relies on a set of security functions that should be instantiated on a per slice basis. Security should be tuned as per tenant/vertical policies for a given slice that implies in turn to manage/duplicate the security functions on logical slices. A user may access different slices for different services, but the confidentiality, integrity and availability should be preserved for any slice. Most of the issues pointed here may be summarised as isolation issues. The potential attacks can affect tenant services by targeting user data but attention should be paid to the control exchanges (e.g. negotiations of Slice-as-a-Service, templates etc.).

When considering knowledge of slice assets and its consequences in terms of security issues, a specific problem occurs by nature as the slices are both:

- an abstraction/composition of the actual systems and services delivered by third parties and,
- flexible, dynamic or even adaptive to satisfy the varying needs through multi-party complex business environment.

Ensuring trust and consistent security policies/governance between tenants and providers point of views is a remaining challenge. This may be considered as a pre-requisite when applied to vertical sectors under stringent security and resilience requirements, beyond existing standards and certification schemes often applied to limited perimeters. There is thus a need to evaluate and expose security attributes of subnets, systems and services involved in the composition of the 5G slices.

Software Defined Security: Slice Protection Deployments

5G slices is a new way to virtualise an infrastructure for tenants with dynamic software defined provisioning technologies and service aggregation rather than a system integration approach. Deployment, orchestration, chaining of (virtualised) protection functions becomes by analogy a software defined security concept.

This opens an entire research field, allowing fine grain policies, AI/ML-based smart algorithms etc. The end-to-end security is thus resulting from the diversity of stakeholders involved in the provision and usage of the slices. As an immediate consequence, horizontal (end-to-end), vertical (infrastructure provider versus tenants), multilateral (Security-as-a-Service, vertical services) coordination/composition of security becomes a challenge.

Nevertheless, the control and flexibility of the software defined security is an enabler for innovative slice defence strategies. Among 5G slicing related security concepts, two categories are emerging, in both cases using the ability to automate the manipulation of the system morphology:

- Micro-segmentation [144] can provide fine grain isolation, specific access control and security policies.
- Deception, overcoming historical honeypots, is assuming that advanced (often unknown/zero days) attacks will be defeated by deceiving the attackers with enabled dynamic/smart proactive security.

Security-as-a-Service, the Future Slice Enabler

Not all tenants will have internal up-to-date expertise available to manage all security aspects. It is thus expected that a wide range of security will be delivered by managed security service providers. Beyond the slicing case, Security-as-a-Service may be considered as one of the very few directions to scale the security needs across the ICT infrastructure and services. One associated challenge is to define meaningful service attributes, and in particular for security services allowing relevant usage of services. Similar to QoS issues, the qualification, evaluation, exposure of those attributes is a difficult problem considering variations of the systems and diversity of risk-based policies.

As an example, Identity and Access Management (IAM) as a Service will be key to distribute and control respective authorities and access across multiple 5G systems stakeholders. Many other security aspects (Key Management, Intrusion detection ...) will be handled this way participating to the slice security.

Monitoring Challenge

As aforementioned, knowledge of slice security conditions is not straightforward considering the multiplicity of authority perimeters and the complexity of dependencies between sub-systems, services etc. Slice security mandates continuous monitoring tracking events or anomalies end-to-end. This dynamic assessment may require both specific advanced tools but will have to face the boundaries of respective stakeholder perimeters. Monitoring or reporting security data is an open field. A particular case is the response to incident which basically require sharing of information from detection towards tenants or adjacent party interconnected.

Beyond Protection

As for protection deployment, detection and remediation are the purpose of smart AI/ML-based strategies that may be applied on a per-tenant basis. Each tenant may for instance deploy its own probes for attack detection across his slice, in addition to some reporting delivered by the providers. Specific remediation strategies may then apply relying on specific critical asset knowledge on tenant side. One can obviously anticipate some remaining challenges with respect to the actual isolation (inter-dependencies) of the slices from other slice/underlying infrastructure or the need of signalling/exchanges between the actors. It should be also noticed that those exchanges are subject to potential attacks and should be secured (may be a specific dedicated ”control and management” slice).

6.4 Distributed Trust Systems

The evolution of the ICT infrastructure is coming after years of hyper-centralisation, it is now considering use cases using hyper-connectivity for distributed applications. By nature, the widening of scope with edge computing, including mobility or IoT is re-introducing issues of distribution of trust across Verticals and Transversals.

The role of the ICT infrastructure, beyond its own security, is to provide a trustworthy platform allowing development and innovation on top of it. Various verticals and applications will benefit from the critical infrastructure with inheritance of the security properties.

With both intrinsic distribution issues and usage by verticals, one of the main challenges will be to manage the component of trust and their distribution, deployment and usage across the highly diverse and dynamic infrastructure. As mentioned before, the application of the perimetric security paradigm is not well suited for an architecture with billions of objects, sporadic activities, mobility. Novel approaches will have to be developed matching the ”metamorphic” properties, thus delivering smart tools for trust distribution and sharing within the unprecedented system complexity.

Distributed ledgers, or blockchains are promising technologies expected to play an important role where the distribution aspects are predominant. Beyond the known basics of those technologies, issues related to the ”evolutivity”, the interoperability, the sharing between numerous private/public initiatives are specifics of the ICT application. Networks are intrinsically end-to-end and the NGI shall provide operational solutions under constraints coming from this transversal limit often more difficult than standalone over-the-top or vertical application.

Beside the distributed ledgers, mostly useful for trusted exchanges and forensic, the heterogeneity of the architecture in terms of security grade is fostering a smart distribution of rot of trust. Balancing some weaknesses of the IoT and also the extra-large attack surface generated by the objects, gateways or secured elements will have to distribute security policy enforcement points such as gateways, probes and secure elements to protect, detect remediate to attacks. A massive issue here is to transform hardware-based security towards flexible software-based solutions. In particular, key storage/management and authentication/rights management appears to be must-have solutions. All forecasted directions for future ICT infrastructure raise critical issues in this field. Among others the evolution of systems widely based historically on the SIM cards and their extensions towards systems encompassing IoT, softwarisation, virtualisation is mandatory considering critical applications (Industry, automotive, health …).

Finally, the fragmentation and the increased openness is requiring mechanisms to identify, evaluate, certify the level of security and thus trust to be given to this or that provider/service participating to the whole system.

6.5 Artificial Intelligence and Machine Learning Application

The increased application of AI/ML, and other data analytics technologies raise the question about their robustness and inherent vulnerabilities. These technologies are a source for new attack vectors. On the other side applying these technologies in the security domain enable more intelligent security solutions. Both aspects of AI/ML and security, the security of these new technologies itself (how to make them more robust and secure) and applying these technologies to create more intelligent security solutions should be addressed in security research with a focus on research areas like:

- Dependencies between physical and cyber world aspects in connected cyber physical systems and their impact on security: Today safety analysis methodologies need to be enhanced to also consider physical vulnerabilities. Faked reality attacks (manipulating how sensors see the physical world) will become a new main threat vector by injecting manipulated data that finally provokes malicious responses of intelligent autonomously acting systems. On the other side, security monitoring can be improved by integrating information and models about the physical world (e.g. ignoring anomalies caused by a known physical component fault).
- Security risks inherent to AI/ML systems: We need a better understanding of the robustness of AI/ML systems. These systems should be aware of the limitations of their models and know the assumptions underlying it. A main security threat of AI/ML systems are adversarial learning attacks, the manipulation of the learning sample set, which in the worst case allows to create hidden backdoors.
- AI/ML systems that can explain their response: In particular, deep learning technology needs to be enhanced by integrating and creating human understandable models and reasoning. This is relevant for human interaction with such systems and the acceptance of semi-autonomously acting systems.
- Security analytics when facing encrypted data: Today security monitoring and anomaly detection very much depend on data that is not available in case of encrypted traffic. This data needs to be substituted by sampling and analysing new types of data and traffic characteristics that are available also in case of encrypted data and allow similar effective security analytics.
- Automated threat Intelligence creation and management: This will be a differentiating factor in future security business and pervasive use of AI/ML is the means to achieve this. The final goal is to create even some predictive capabilities. Challenges are to automatically collect and analyse examples from deployed security appliances, create new learning examples, measuring the effectiveness and efficiency of currently deployed models, and incrementally optimise the model learning. Another challenge is utilising a diversity of other Threat Intelligence resources (structured and unstructured) and being able to exchange and combine models.

7. Communication Satellite Technologies

7.1 Overall Vision

In the timeframe 2020 – 2030, it is expected to testify to a larger demand for ubiquitous content access offered to users with unprecedented data rates. The network infrastructure will have therefore to enable a multifaceted set of applications, exhibiting very different characteristics and needing to meet various Quality of Experience (QoE) requirements. Matching such expectations from citizens, governments, and enterprises means deploying a network infrastructure able to everywhere offer always available and efficient points of attachment to the telecommunication facilities.

From this standpoint, the mission advocated to the satellite technology will be to complement the terrestrial network infrastructures for the use cases where the actual capacity availability will be limited in comparison to the traffic requirements because of the users' density or their specific geographic location. In particular, the ever-increasing amount of traffic transported by today's infrastructures and the forecast to exceed 70 ExaBytes already in 2022 will call for deploying technologies suitable for mobility context, since more than 80 % of the traffic will stem from mobile users, and the ability to provide complementary capacity when needed. In other words, anywhere-anytime concepts are certainly an added value that satellite technology can offer, not just in the specific scenarios where terrestrial connectivity is not available at all, but also as means to efficiently offload user contents without penalising the perceived QoE of users.

Moreover, the further development of very high throughput satellite systems offering several Tbps capacity as well as the operation of newly envisaged mega-constellations would not only improve the overall signal coverage in underserved areas (e.g., rural) or underserved/unserved scenarios (e.g., maritime and aeronautical) but also improve the service delay performance, hence broadening the market towards new sectors. It is therefore expected that satellite technology will play an instrumental role in decreasing digital divide and increasing the overall Internet content accessibility and overall network penetration, which is still limited in Europe to about 80 % (according to 2017 statistics from ITU-T) and even less in other areas of the world. As such, its contribution to let citizens benefit from the available services independently of the specific location or of the GDP of the country they belong to will be a driver to further cutting off the digital divide. The expected advantages are not limited to Internet access, but also should be considered in a broader sense to include emergency/warning services as well as all the services that require a high level of security that satellite systems can almost inherently provide.

The role of satellite technology will also receive an important boost to support the M2M/IoT market, which is in continuous growth and therefore even enables new services in this context. The many IoT services implemented in the maritime, energy (i.e., remote monitoring of smart grid and energy plants) and agriculture sectors (e.g., reconnaissance and monitoring of fields) will definitely see the application of satellite communication as a necessary means to convey a larger amount of data to service operators, control centres and other involved stakeholders.

Last but not least, the recent technology innovation in the field of tiny microprocessor development and overall satellite payload miniaturisation has paved the way towards the definition of new operational concepts parallel to those assumed with classical LEO, MEO and GEO systems, i.e., building on nano/picosatellites as well as cubesats. Although their current missions are mostly intended as verifying the actual capabilities of these systems and figuring out a little bit more their actual employment in real use cases, their application is expected to become quite important in the horizon 2020+, especially to support IoT services (e.g., reconnaissance, surveillance, and monitoring), which will have large applications in different industry sectors. From this perspective, suitable payload design will have to be carried out, hence requiring deep studies so as to offer the necessary data rate to support the aforementioned services.

In general, the overall ecosystem, where satellite is expected to play a role in the next years, is depicted in Figure 8, where the main use cases and related services are highlighted.

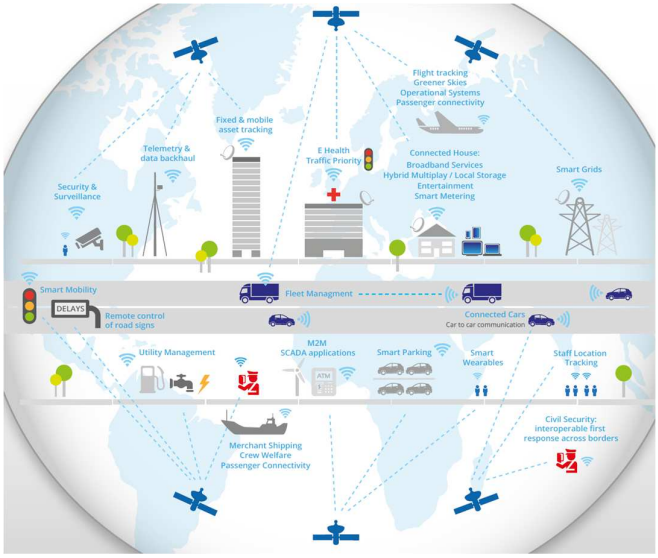

Figure 8 Overall positioning of satellite in current and future society

More specific details about the various sectors, for which satellite technologies expected to be further developed, are elaborated further in the next section.

7.2 Enabled Services

As outlined in the previous section, the role of satellite in the 2020+ horizon is perceived a pivotal to ensure data connectivity to fixed and mobile users for a large class of services, to extend or complement the coverage offered by wireless and broadly speaking terrestrial technology.

With many more people expecting to have the same coverage when travelling (on cruise liners, passenger aircraft, high speed trains and in holiday villas) it is key that satellite allows seamless extension of 5G+ services. Main services can be grouped according to the following classes:

- Multimedia Delivery
  - Classical broadcast to homes
  - Content delivery to the edge
- Broadband Access
  - Fixed broadband
  - Mobile broadband
  - Backhaul
- Machine Type Communication (M2M and IoT)
- Reliable and Critical Communications
  - Disaster and Emergency Communication
  - Air Traffic Management
  - Governmental Communication (resilience, security, availability)
- Connected Car
  - Traffic updates
  - Ecall
  - SOTA (Software update over the Air)

Some of the aforementioned applications are further developed in the next subsections.

7.2.1 Multimedia Delivery

Broadcasting to Homes

Direct-to-home services and the multimedia content distribution will stay as a key application for GEO satellites. Although there is a trend to also distribute multimedia and TV programmes over terrestrial lines or the mobile networks, the trend to produce programmes in ever higher resolutions, e.g. for live TV events including sports, will distinguish satellite services from terrestrial ones. In this respect, distribution of 3D/4D contents with 4k/8k resolution (or higher as expected on the long term) is certainly a challenge that satellite technology can fulfil.

Content Delivery to the Edge

Considering that more than 80 % of data traffic in 2020+ will be video-based, the support of satellite technology to deliver contents to mobile users will be of great importance. Data distribution is going to be implemented according to new *edge computing* paradigms to which the satellite ecosystem will have to adapt with respect to the caching capabilities of the ground segment as well as the satellite platform. The evolution of the latter will help developing a new satellite value chain, aimed at defining in-orbit, storage and processing as services capabilities. Those will be offered to the rest of the network infrastructure to enable new services and enhance the user's QoE of the existing ones. From this standpoint, it will be necessary to define proper caching policies embedded in a content oriented architecture, building on the concepts of Information Centric Networking extended to properly meeting security requirements.

Last but not the least, a proper infrastructure will have to be defined to make available the huge amount of data coming from Earth Observation (EO) missions, which is typically confined at the edge of the network and will require adequate scheduling and fetching mechanisms in order not to flood the rest of the network. This engineering task has obviously implications on the overall space segment design, which will have to duly considered.

7.2.2 Broadband Access

Satellites are well known for their ability to provide large coverage at low cost and thus complementing terrestrial networks to provide 5G+ global broadband services. Nowadays, broadband access is already a high-volume market for satellites. Main operators provide broadband access almost world-wide. They complement the terrestrial service provision in remote, unserved and underserved areas. The total throughput of next generation GEO HTS (VHTS, Very High Throughput Satellites) are expected to reach the Terabit/s allowing economically viable consumer services in the range of 50 – 100 Mbps, using multi-spot beam approaches and taking advantage of future optical feeder links. Furthermore, these solutions can provide high data rates at a comparable cost regarding terrestrial solutions. Additionally, many constellations are in discussion for future deployment to span worldwide satellite coverage also in lower orbits, especially in LEO. Moreover, MEO systems, such as O3B, are already in operation and serve e.g. cruise ships with high performance connectivity.

7.2.3 Mobile Broadband to Users and Vehicles

It is expected to offer mobile multi-play broadband interactive services (300 Mbps) in 2020+ for fulfilling the high user data rate demands for vessels and airplanes. This will be achieved by means of next generation HTS systems, either as GEO, MEO or LEO systems. Moreover, the availability of optical links will help to further boost the available capacity of future systems, hence accepting the challenge of distributing large amount of data with very high data rate to all users. In this context, the support of intelligent transport networks and connected vehicles (automotive, maritime, aero) will show satellite as one of the main players, complementing terrestrial networks where necessary or being the only possible candidate in the applications, where terrestrial infrastructure is not available (maritime and aeronautical communications).

More details about the aforementioned sectors are given in the following.

Ships

Far from the coast, only satellite can provide the required connectivity. Currently, low data rate solutions are widely used (more than 45,000 terminals) by all class of ships, whereas high data rate systems are bringing new services to ship operators.

For the crew, the level of connectivity provided on board is a determining factor in the choice of their employer. The acceptable minimum throughput is 512 kbps. Connectivity is also used for the operations related to the vessel, maintenance and to help navigation (meteo information, maps).

On the other hand, continuous connectivity is required by most of cruise ship passengers, pushing cruise companies to invest in broadband connectivity solutions. High speed Internet opens new business opportunities to cruise companies, like congress organisations. The addressable market (broadband connectivity), when considering ships of more than 1,000 tons, is about 52,000 cargo ships and 7,000 passenger ships. The global demand, which is around 9 Gbps of capacity today, is expected to growth to 90 Gbps in 2024. As such, the increasing fleet of ships populating seas and oceans in 2020 – 2030, will create the need for data distribution over such a maritime network, with new demands in terms of content caching and efficient security paradigms to control data access and confidentiality.

Aeroplanes

More than 5,000 aircraft are providing connectivity to passengers today. Growing data rate demand pushes the airlines to choose satellite solutions that provide better user experience (higher data rates and global coverage) through HTS solutions than existing air-to-ground solutions. It should be noted that the total fleet of aircraft with more than 100 passengers worldwide is valued at 18,000 aircraft by the manufacturer Airbus and is expected to double to reach more than 37,000 aircraft by 2034. Therefore, also in this case the fleet of aircraft will be so large that an actual moving network in the sky will require the necessary capacity to provide all users with the demanded content and corresponding data rates. As a matter of fact, the connection to the Internet can be provided either by GEO satellites, MEO satellites or by (a) constellation(s) of LEO satellites. The offered data rates are in the range of 10 to 100 Mbps by plane, shared between passengers, and expected to become even larger.

Trains

Globally, a need for high-speed railways to connect key cities is identified to boost economic development. These trains provide several advantages such as reliability, punctuality, security, avoidance of traffic congestion, and comfort. Changing consumer lifestyles and business needs require people to stay connected during their travel. Train operators and telecommunication service providers are actively looking at different technologies to address business opportunities.

Among the wireless technologies deployed by various train operators, satellite and LTE are the two most popular technologies for broadband Internet access in high-speed trains. As such, proper handover and network selection solutions must be put in place to cope with the expected trains' high speed and the increasing demand for flat data rate for the trains' passengers. The upcoming availability of (mega-)constellations consisting of several LEO satellites may be able to meet these demands, provided that proper ground segment engineering will be carried out.

Cars

Satellites could also be used to provide broadband services to vehicles. A global satellite network is the solution for those car manufacturers, which sell cars on every continent.

Satellites can reach rural areas that don't have cellular service. Combination of GEO satellites with LEO constellations may help to make the service also available in the cities, which require a high elevation angle. Without such solutions, manufacturers must contract with hundreds of terrestrial operators all around the world.

Services offered could be for example multicast of data to vehicles: the satellite sends software and firmware updates to computers on board vehicles. This may eliminate the need for an owner to bring his car into a garage for maintenance and would accelerate the response to manufacturer recalls. Moreover broadcast/multicast of multimedia contents like videos, music, newspapers is also in the agenda of car manufacturers for digitalising the journey experience of passengers.

Currently, the antenna is a critical point. Flat satellite antennas for cars are under advanced development by several manufacturers and will be soon available.

Backhaul Connectivity

Many applications, where satellites can play a role, are identified in industrial automation and utility, wireless health services and for smart cities. Moreover, the major vertical applications that would be mostly adopting 5G and evolution based applications includes automotive, energy and utility, healthcare, industrial automation, intelligent buildings and infrastructure, public safety and surveillance, retail, consumer electronics as well as home automation.

7.2.4 Machine Type Communication (M2M and IoT)

It is expected that the M2M/IoT market continues to grow globally, as more sensors and devices are connected to high-speed networks to provide valuable, large-scale data. By 2020, it is estimated by Frost & Sullivan that there will be more than 80 billion connected devices, globally, across various industries. At the same time, each human being may have an average of 5.1 connected devices, and these devices will permeate multiple facets of various industries. NSR even estimates that by 2023, there will be 5.8 million M2M/IoT connections via satellite around the globe. It is however claimed that a larger number will make this sector even more attractive for the growth of satellite interests, hence making the evolution of satellite systems a necessary task, especially with respect to antenna design and overall system capacity. The most attractive verticals for satellite M2M/IoT are displayed in Figure 9.

Figure 9 Most attractive verticals to be served by satellite

As it can be seen, the most attractive applications in the different verticals are asset tracking, fleet management and equipment monitoring. Most of these applications may be served by a significant portion of all low-bandwidth S- and L-band communications' bandwidth. Low-bandwidth satellite operators will continue to focus on M2M communication modules and sensor integration to facilitate growth in this market. Next to this, high-bandwidth Ku- and Ka-band satellite technologies will also be used in the IoT industry. However, unlike S- and L-

band technologies, high-bandwidth satellite services will not normally be adopted solely for IoT due to the high cost of service and equipment. These technologies will instead be used as backhaul for low bandwidth IoT networks in locations that are distant from landline and cellular services.

7.2.5 Reliable and Critical Communication

The support of satellite communication in disaster management is a clear requirement, since in these events the terrestrial infrastructure can be partly disrupted or totally unavailable. A similar necessity arises in the case of crowded events (such as Olympic Games or concerts), where many people are concentrated in small areas, whereas densification alone is not sufficient to guarantee high-data rate connectivity to all users. On the contrary, users' demands will have to be met by complementing the terrestrial infrastructure with high altitude and satellite platforms. Their availability is not just aimed at introducing more capacity into the overall network but also to provide ground for efficient offloading techniques against congestion events.

7.2.6 Other Applications

Agriculture, Mining and Forestry

The utilisation of modern techniques like precision agriculture, agricultural robots, equipment telematics etc. seeks to improve the productivity yield and enables sustainable agriculture practices. There is a compelling case for an urgent action in improving the current global agriculture practices to meet the unprecedented rise in demand for the increasing population and to mitigate climate changes. Incorporating new communication technologies in agricultural systems will help reducing the current stress on farmers and facilitates the economic improvement of the farmers. In this context, the availability of satellite technology is considered pivotal to help better controlling the different phases of agricultural processes and take timely corrective decisions when needed.

Precision agriculture, also known as precision farming, is the use of a broad range of modern technologies including robotics and automation technology, imagery and sensors, data analysis, and bioengineering. The combination of all these technologies with agricultural sciences improves crop health and yield, and reduces weed generation and utilisation of fertilisers. In this use case, real-time data collection coupled with wireless/satellite communication technologies provides accurate data on field mapping, harvesting, fertiliser rate, and weather conditions thus improving the crop yield. As such, the cooperation between satellite and terrestrial technologies is a preliminary requirement to be fulfilled. Moreover, satellite links must be provided with the sufficient capacity necessary to accommodate large amount of data.

Agricultural robots otherwise known as Agribots are basically a digitised electromagnetic machine used during the harvesting stages of agriculture. Agribots are mainly designed for fruit picking, spraying, and operating tractors thereby reducing human labour. Unmanned aircraft and ground vehicles are operated to map, observe, and provide precise data about crop details. Also in this case, tight cooperation between all networked agents in place (satellite and non-satellite nodes) is necessary to achieve an efficient automation concept on the one hand and then implement a suitable data sharing and distribution mechanism.

Equipment telematics is a technology that deals with transmission of data through long distances. Telematics combines electrical engineering, telecommunications, and vehicular technologies, which can be incorporated in the agricultural sector to benefit farmers. The consequent real-time data access therefore requires technologies able to reach remote areas such as satellite systems.

Maritime Surveillance

Integrated maritime surveillance including existing Vessel Monitoring System (VMS) data, along with satellite enabled Automatic Identification System (Sat-AIS) and satellite based Synthetic Aperture Radar (Sat-SAR) data will gain prominence. Future applications will be utilising big data-based algorithms to include historical data as part of the detection/verification process. Specific solutions are moving toward a modular distributed architecture, especially in the military domain. However, in the civil market, web-based integrated surveillance solutions will gain prominence. As such, the capability of integrating the satellite ground system with the rest of the maritime network infrastructure, embedding analytics capabilities for the use of big data paradigms will be a necessity (Figure 10).

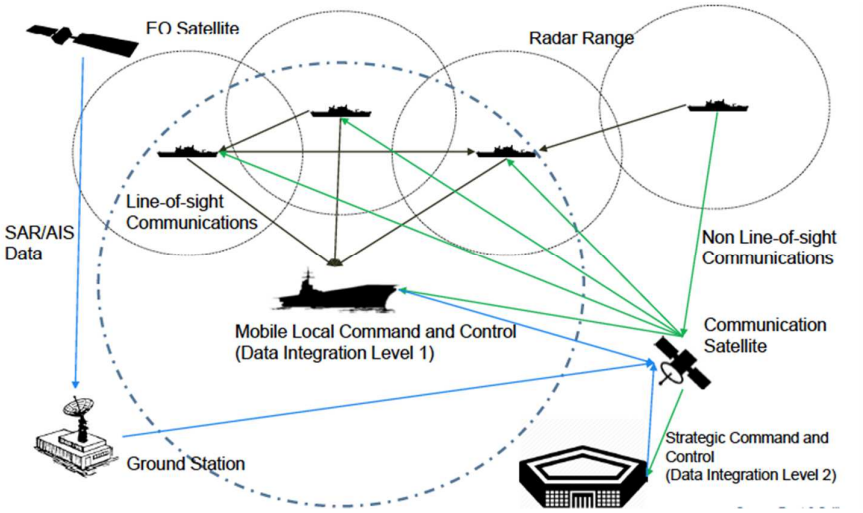

Figure 10 Maritime surveillance supported by satellite communications

7.3 Ground Segment

7.3.1 Physical layer

Antennas

Recent advances in antenna design will become even more relevant in the next years to make satellite systems more flexible to multi-orbit and mobility requirements. On the one hand, availability of arrays made of general purpose SDR along with flat plate antennas will be an important achievement to further boost the penetration of satellite communication. On the other hand, the deployment of multi-orbit systems or the opportunity to exploit connectivity with different satellite systems operating in various orbits will help the massive use of electronically steerable antenna able to adapt to different satellites in different orbits. Moreover, the overall modem design will have to be conceived to allow multiple connections towards different satellite platforms.

Radio Resource Management Solutions

Future satellite systems are expected to implement a larger number of beams to offer each user in a beam with higher data rate. Moreover, to further increase the available beam bandwidth, full frequency reuse will be implemented with the need for implementing the proper interference mitigation techniques, such as precoding and MUD (Multi-User Detection), where applicable. These options will also have to be properly conjugated with the countermeasures taken to contrast weather impairments introduced by operating the satellite systems at Extremely High Frequency (EHF) bands, i.e. Q/V and optical links.

Another important aspect to be taken into consideration is the implementation of distributed dynamic radio resource management techniques, not just limited to the ground segment but

also applied in the space segment, especially in the case of flexible payloads. In this case, obviously proper tuning of resource allocations schemes will have to consider also the impact of frequency/time and power allocation along with the traffic variations occurring on the satellite network.

User Terminal Solutions

The maturity of satellite user terminals will go through the unification of the physical layer specification (e.g., MF-TDMA, SCPC, Mx-DMA, etc.), which nowadays is very much specific to the various applications to be supported. The final objective would be on the contrary to have a universal satellite return link able to adapt transmission modes according to the different traffic flavours and the network changing conditions. Particular effort should be dedicated to the case of uncoordinated access to satellite capacity, which has been receiving more and more attention in the past years in the form of enhanced random-access schemes. Its consolidation is expected to play an important role in the next years to further support the increase of M2M/IoT applications over satellite systems. Yet in the context of IoT over satellite (but not just limited to this application), the evolution of current modulation schemes to non-linear ones will certainly represent a milestone, going through the adoption of the CPM format. The introduction of new modulation strategies will be pivotal in particular to satisfy the stringent requirements for energy efficiency coming from IoT application but will also have broader advantages for many other use cases.

7.3.2 Network Operations

Unlike the recent past, where satellite networks have been quite homogenous, the next generation of satellite systems will have to face important technological challenges because of the coexistence of different technologies under the same umbrella. In particular, the integration of Q/V frequency bands with optical feeder links to offer capacity higher than 1 Tbps will require important coordination functions, with relevant implications on the user, control, and network planes. On the one hand, the so-called cloudification of gateways will be a key task to make the overall system more efficient and flexible. On the other hand, such exercise will have important results from a network management point of view, stemming from the need of implementing SDN/NFV to achieve a proper orchestration of the network of satellite gateways. The coordination of the network operation will be tightly connected to resource allocation strategies, which will have to be implemented in software, with the ultimate goal of achieving the full softwarisation of the system in the time horizon 2025 – 2030. Moreover, given the large number of gateways and very diverse operative network conditions, dynamic network management will have to be supported by AI/ML solutions, to have proper and timely adaptation of network policies as result of traffic and channel conditions variations.

The design of gateway functionalities and the related coordination will also be triggered by the specific design of the space segment and the need for implementing mobile scenarios. As to the former, time-sliced gateways with time switched connectivity to users' beams will be implemented so as to reflect the use of beam-hopping satellite payloads. As to the latter, mobility will call for efficient and seamless handover at beam, gateway, and satellite level, hence requiring the deployment of electronically steerable antennas, which will eventually be more realistic than in the past.

7.3.3 Content Delivery Optimisation

Optimisation of content delivery over the satellite networks necessitates important enhancements at protocol and architectural level, especially for what concerns the management of quality of service and likely to achieve certain levels of QoE. In particular, transport protocols and overall networking functions must be optimised and tailored to the characteristics of future satellite systems. In that context, multi-orbit systems will be in place, exhibiting very diverse operating conditions in terms of offered capacity and introduced delay. In this respect, transport protocol acceleration coupled with data compression will be pivotal

to exploit the available network resource, also in the presence of hampering weather conditions. Moreover, proper dimensioning of the buffer will be a key-task to achieve quality of service optimisation, which will also require proper queue management, by introducing new scheduling policies. This will assume even further importance considering the solutions implemented at physical layer, with respect to precoding and ACM, which will require a tight coordination in the sense of coordination and configuration to meet the service requirements demanded by users. In the context of QoS/QoE, there will be the relevance of solutions aimed at facilitating the deployment of delay-sensitive services. In particular, the use of caching solutions in the ground segment will be important to implement MEC solutions and therefore reducing the overall medium and data access delay experienced by the final users.

Parallel to the optimisation of quality of service there are the security requirements, which must be considered while designing the satellite network. In particular, different verticals have different security requirements, whereby specific models applicable for the satellite ecosystems have to be worked out, by taking advantage the latest advances in cybersecurity. Moreover, the recent developments in the domain of quantum key distribution and in general of quantum communications are expected to bring important novelty as to how satellite networks are going to be re-engineered from a security viewpoint.

7.4 Space Segment

The growth to higher data rates is also reflected by current satellite developments. Up to 3 Gbps are possible using wideband transponders (recently shown 1.2 Gbps using 420 MHz transponders) and state-of-the art communication technologies achieving around 7 bps/Hz. The gain from statistical multiplexing along with the usage of transponders providing more bandwidth, satellite throughput can be considerably enlarged. Adopting new frequencies (Q/V band) with higher bandwidths for feeder-links increases the possibilities to operate at higher data rates. Multi-spot beams and ISL (Inter Satellite Link) additionally provide solutions to drive these developments. Moreover, it is expected that the further technology advances in software defined radio (SDR) will further help increasing performance in comparison to nowadays system and therefore resulting in the full replacement of custom hardware solutions. Such aspect will be actually not just limited to the case of space segments but also to the ground one, although the advantage it might bring will be certainly relevant for the space part, taking into account the case of mega-constellations, adaptive beam solutions for GEO systems, and in general the challenge of making satellite systems as full softwarised objects.

As such, it is expected that (r)evolution of the space segment in the next ten years will be a key driver to achieve very high data rate and therefore let satellite play an important role in overall data access for a large class of missions.

The main directions and challenges for the space segment (re-)engineering are highlighted in the next subsections.

7.4.1 HTS Broadband GEO

In the last two years, the next growth axis of the SatCom market has taken shape with the launch of high throughput satellites (HTS) and payloads worldwide enabling efficient high-speed internet-by-satellite. The EU now recognises that satellite broadband services have allowed to really deliver ”broadband for all” and to reach a major target of the digital agenda for Europe. However, while the European industry technical offer competes with the most advanced American ones, the low institutional support at demand level does not favour a rapid broadband market take-up in Europe and further concrete development in this field in the short term. These broadband satellites are indeed mainly serving the American and Australian market: Viasat, Echostar HNS and NBNCo. Additional satellites (like Echostar XVII) will be launched soon delivering even more total capacity (>100 Gbps).

Achieving even larger capacity is the next milestone for satellite industry in Europe, to offer enterprises and citizens with an even wider gamma of services provided with unprecedented

quality of service. To this end, implementing the concept of Tbps capacity can be sustainable by migrating the feeder links to Q/V or optical bands and delegating the Ka frequency band to the user links. To this end, suitable RF/optical conversion strategies implemented on-board the satellite (and on-ground) are desirable to implement microwave photonic on-board processing functionalities.

7.4.2 HTS Broadband MEO

Next to the above-mentioned GEO services, O3b with a constellation of MEO satellites is addressing emerging and insufficiently connected markets mostly in Latin America, Africa, the Middle East, Asia and the Pacific while offering a low latency approach. An example of business focus is maritime applications to serve large cruise ships with high data rate connections. One of the main advantages is to provide data access at a reasonable latency given the lower altitude of MEO satellites with respect to the GEO counterparts. More advanced design of antennas is however necessary to boost the performance.

7.4.3 LEO Constellations

LEO deployments will need complete constellations ranging from some hundreds satellites up to thousands of satellites. There is a bunch of constellations discussed operating in Ku- and Ka-bands. They mostly address Internet connection as a service but also secure point-to-point communications. Frequency coordination is of vital importance as they need to operate next to the already existing satellites in different orbits and in the same orbits. To provide their full service capabilities, inter-satellite links will be an enabling technology. On-board processing of the signals is essential for efficiently routing the signals to the right destinations.

The very next frontier will be represented by broadband mega-constellations equipped with optical inter-satellite links. On the one hand, the large number of satellites will help achieving high-granularity coverage, so that higher capacity will be made available from the entire satellite system. In simple words, the satellite constellations will work as a huge distributed network switch operating in the sky. On the other hand, the availability of optical ISLs will drastically reduce the latency of re-routing operations in space, taking also into account that speed of light in free space is higher than in terrestrial optic fibres. As such, the forward of large amounts of data between adjacent satellites will be performed in almost negligible time, hence letting users experience a new Internet experience from the sky owing to the limited delay, coming from the low altitude of LEO satellites. Extension of LEO constellation is also expected to be a top topic in the next years to achieve the so-called integrated space data highway, i.e., consisting in integrating different classes of satellite into different tiers of the same hierarchical satellite system. In other words, the integration of GEO, MEO, LEO, HAP and aircraft in a unique mesh network in the sky will help achieving large capacity to be offered to end users.

New concepts of satellite constellations will also include the cases of fractionated and cooperative constellations, to make different systems interworking provided that computing resources are available on board, whereby additional developments are expected and encouraged in the timeframe 2020+.

7.4.4 Highly Flexible Payloads

The success of new satellite systems will strictly go with the development of advanced and more flexible payloads. In particular, more digitalised payloads will be deployed with the aim of offering a large flexibility (e.g., beam hopping by 2020) in terms of on-board beamforming, modulation/demodulation function, power and frequency allocation of the different beams. To this extent, the use of active antennas for maximum coverage and power allocation flexibility will be one of the key points, as initially explored in the USA. Moreover, more sophisticated on-board signal processing functionalities are expected to support dynamic switching, linearization of transponders, filtering, as well as beam reconfiguration. In turn, beamforming

operations will be initially implemented on-ground (GBBF) and then complemented on-board (OBBF).

A key-point will be the dynamic reconfiguration of the satellite (beams, connectivity, power, etc.), automatically controlled by the communication network. To this end, adoption and extension of SDN concepts to implement SDN controllers on-board of satellites will be a fundamental step. This will be an important step forward toward the softwarisation of the whole satellite system, with particular advantages in the case of LEO constellations, where flying SDN controllers would help better automatizing data forwarding operation. Moreover, to make the entire space segment self-configurable and able to more efficiently track the variations of traffic and operative conditions, the use of AI/ML tools will be pivotal to achieve more sophisticated operational models. From this standpoint, data forwarding will be made more efficient by means of digital routing and channelization, combined with interference monitoring and mitigation in order to timely allocate data flows to the channels where the operative conditions are the most stable and suitable from a QoE perspective.

Finally, as instrumental to achieving a distributed-caching network and therefore enabling more efficient satellite-based MEC operations, satellite payloads will be equipped with large amounts of data storage, so that requested content could be immediately stored on-board rather than being requested to publisher located by teleports. This advantage is considered particularly relevant to mitigate the typical large latency introduced by GEO satellites or even to keep the data delivery delay below 10 ms in the case of LEO constellations.

Obviously accommodating all the aforementioned functions in a satellite payload requires an attentive design of all modules and in particular of the available power and the necessary mass to fly such an object.

7.4.5 Nano-Systems

Recent trends show the development of nano-satellite systems being mostly attractive because of the limited costs and the moderately high number of components required. Currently, the ratio between investment and return is not very appealing from a commercialisation viewpoint, in particular because current nano-satellites have been defined especially for experimental purposes. It is however expected that in the time frame 2020 - 2025, their exploitation will be more remunerative, taking especially advantage of inherent monitoring and reconnaissance capabilities. Moreover, their support to IoT services will be even more important, by means of improved on-board capabilities. Their function to extend the terrestrial network capacity will be carried out by means of nano- and pico-satellites, which together with cubesats, will make use of high data rate links built on Ka-frequency band or even optical links. To this end, special attention on miniaturisation of the payload as well design of highly efficient antennas will be necessary.

7.5 Communication Architectures

7.5.1 Virtualisation and Network Cloudification

As pointed out in the sections dedicated to the evolution of ground and space segments, full softwarisation of the SatCom ecosystem is expected, whereby classical satellite architectures will be modernised towards the deep adoption of NFV/SDN concepts. In particular, virtualisation of most functions will be carried out, with important implications also on the functions that are typically implemented in satellite modems. In particular, delegation of some of these functions to the cloud or data centres is of the approach already explored nowadays and to be further assessed in the next years in order to make SatCom systems ready to a full integration with the terrestrial counterpart. In particular, redesign of the radio access and physical layer will be necessary to properly virtualise the specific functions and expose them through the upper layer over given standardised interfaces, by taking properly into account the overall processing latency additional introduced by could computations.

The advantage and necessity of cloud-assisted functions inserted in the communication architecture is actually not just limited to the case of function virtualisation but in general to the entire system concept, where the implementation of multi-beam multi-gateways approaches will require tight operation coordination between all involved nodes. This aspect will be of quite some importance especially as future HTS systems building on feeder links operating in Q/V bands as well as free space optical links, whereby careful and optimised handover and distributed resource allocation strategies will be of paramount importance to boost performance. From this standpoint, it is immediate to see the need for sophisticated and advanced orchestration schemes, which could be either fully centralised or implemented in a distributed manner to better cope with the issue of increase delay processing.

According to the scenarios described above, the shape that satellite systems are going to take very much resembles that of automatized objects, where transmission must autonomously adapt to the specific service data being transported and the contingent network conditions hence representing time-varying constraints. In order to achieve this ambitious but necessary target, the implementation of deep-learning and overall AI/ML concept will be of paramount importance to efficiently automatize the operations across the entire satellite network. The task will become even more stringent, considering the larger volume of traffic and the higher demands of increased user data rates on the one hand and the non-negligible delay that will certainly impact on the design of proper AI/ML-inspired solutions on the other hand.

7.5.2 Enabling Networking for NGSO (Non-Geostationary Satellite Orbit) Systems

As highlighted in the section devoted to space systems, great expectations from the satellite industry are in the launch of mega-constellations. In this respect, the design of novel medium access protocol will be necessary to efficiently exploit the overall capacity without receiving much penalty in the overall access delay that could definitely penalise services requiring low latency delay. From this point of view, the last years have testified quite some effort from the satellite industry and research institutions in the direction of random-access, especially for IoT/M2M applications.

Moreover, the implementation of inter satellite links and in general the integration of RF and optical links will call for the design of a more efficient protocol architecture, especially for what regards data routing and forwarding. On the one hand, routing in the sky is a very attracting opportunity taking also advantage of the mobility features already offered by IPv6 protocol. On the other hand, routing functionalities in the satellite might be too power and resource demanding, whereby switching performed at the datalink layer might be more desirable. As such, however, important aspects related to address resolution and mapping as well as support for mobility will have to be taken in due consideration while designing the full systems. Another important aspect to be considered is the fact that the implementation of optical downlinks could be typically subject to link disruption or temporary unavailability because of clouds and scintillations, whereby suspend-resume networking concepts would be necessary. To this end, the support of the DTN (Delay Tolerant Networking) concept would be quite helpful, although the current specifications and implementation should be extended and tailored to the characteristics of the reference environment, especially with respect to content-based delivery and QoS management.

7.5.3 Optimised Content Delivery

A key-role played by communication architectures in the next generation of satellite systems will be to boost performance of content delivery. To this end, some importance will be given to implementing caching functionalities within the network nodes. In particular, satellite terminals are expected to be equipped with caching capabilities, implemented in the radio resource unit controller. The same is expected also in the satellite payload, to improve the user experience in terms of service delay. Certainly, the overall implementation will have to be supported by the selection of an efficient overarching architecture inspired to ICN

(Information Centric Networking) like principles, to achieve a more mature and advanced content delivery solutions for satellite systems.

Moreover, the implementation of satellite systems composed of links with different characteristics (e.g., use of Q/V bands and optical links, coordination of GEO and non-GEO systems) will call for proper network selection mechanisms as well as a multi-path oriented protocol. In this context, the consolidated MPTCP protocol will have to be further optimised towards its use in satellite networks, especially to take in consideration the various latency therein exhibited that might affect the multi-path operations for what concerns data scheduling and fetching functions.

7.6 Convergence with Heterogeneous Networks

7.6.1 Joint Radio Resource Management (RRM)

The integration of satellite and terrestrial networks obviously calls for adequate resource allocation schemes, considering traffic variations across the entire network as well as the characteristics of the various links over which data delivery should be carried out. Given the high heterogeneity of networks in play, an efficient resource allocation concept has to be implemented at the different layers of the protocol stack. In a first instance, the simultaneous use of different frequency bands as well as the opportunistic access of the same one will have to be properly coordinated by means of adequate cognitive systems. As such, efficient spectrum management will be of paramount importance to mitigate and where possible avoid cross-link interference. In turn, the capacity request from different vertical services will have to properly map on different network slices, to properly met QoS guarantees across the entire network. To this end, efficient joint resource allocation schemes will have to be developed, depending on the different traffic flavour and network condition variations over time. As a consequence, the use of AI/ML concepts will play an important role, especially in the short-term prediction of operative network conditions, as it will be instrumental to efficient network selection and therefore the consequent resource allocation.

7.6.2 End-to-End Content Delivery

The availability of multiple links originating from converged networks obviously opens the door to a more efficient distribution of data between data producers and consumers. On the other hand, the heterogeneity of the networks being integrated calls for optimised delivery strategies. In particular, the use cases benefitting from network diversity to relief congestion situations (i.e., traffic offloading from one network to another) will have to implement efficient routing and caching solutions in order not to penalise the users' experience in terms of throughput, delay and jitter (especially in the case of video applications). From this standpoint, the availability of the multi-delivery protocol is an added value for future systems to get a more reliable and efficient approach to transporting content over the available networks (satellite and terrestrial). Obviously, the correct selection and consequent implementation of these solutions will be very much dependent also on the characteristics of the reference satellite networks, where the availability of GEO or LEO constellations can lead to different network selection decisions. Moreover, proper integration of different networks will also call for efficient routing and forwarding mechanisms piloted at central level from a management viewpoint.

Finally, for dealing with efficient content delivery on an end-to-end basis, it will be necessary to implement a full ICN-like (Information-Centric Network) based architecture to have a well-consolidated architecture design spanning the different network segments, avoiding ad-hoc implementation (e.g., CDN-like – Content Delivery Network) which requires important workaround in the Internet evolution to allow proper interfacing.

7.6.3 Security

Different networks currently implement various security models depending on the specific vertical and the vulnerability issues exhibited by the underlying links. As such, the full convergence of terrestrial and satellite systems should carefully consider the implementation of a proper security concept, to avoid silos solutions that would essentially limit the actual convergence between networks.

In this perspective, it will be of paramount importance to properly define the vulnerability boundaries of each interested network along with the security requirements demanded by different services, to properly build security models able to reflect the actual needs of citizens and enterprises. This process has to be implemented at the different layers of the protocols stack, considering the different directions of policies enforcement, i.e., physical layer and application layer driven. The main expected outcome will be therefore to come up with solid proposals in the context of integrity, data access, confidentiality and encryption schemes (just to cite a few), building on the most recent advances on security engineering such as quantum key distributions and overall quantum communications.

7.6.4 Integrated Network Management

Properly operating a converged network is an absolute requirement, which translates into defining an efficient network management model. Usually satellite and terrestrial networks build on different network management concepts because of the intrinsic nature of the two network flavours. As such, it is necessary to unify them to have a flexible and consistent network management plane. One of the main objectives is to meet various QoS/Quek requirements on an end-to-end basis considering the fact that the overall network will be composed of multiple segments, each of them shared amongst multiple tenants. The overall network management plane will therefore have important implications on system orchestration, building on SDN concepts. Moreover, as mentioned in the previous section, an important function to be ascribed to network management tool is to enforce the proper security policies depending on different verticals and update them upon important application or network variations (especially in case of mission-critical services or disaster management situations, for example).

8. Human Centric and Vertical Services

8.1 Digital Service Transformation

Telecommunication networks have been regarded an essential pillar of any society's infrastructure to progress and sustain its economic growth. Millions of people rely on the diverse services offered by telecommunication networks, which is part of a real Digital Economy. These services require underlying network technologies to support higher workloads and increasing traffic volume while reducing the overall network operation costs. Each new generation of mobile networks (GSM, 3G, LTE, 5G) goes beyond a simple increase in network speed or reliability, and instead it brings unique new service capabilities for people and vertical economic sectors. Disruptions driven by emerging and maturing technologies are impacting businesses and society with increasing pace and depth, and this trend will accelerate in the next years. Future networks, evolution of cloud computing, any type of connected object and the strategic use of data and analytics are the foundations of the digital disruption.

The envisaged trend is the convergence of these foundations and their complete fusion in an ICT continuum platform. We already perceive this trend. Public clouds have become mainstream today and cloud computing evolves towards hybrid models, combining private clouds and being extended towards the micro edge clouds with lighter virtualisation techniques. Network softwarisation also reinforces this convergence of networks and IT systems. At the same time the number of connected devices, tablets, wearables and IoT elements is growing to at least 50 billion by 2020 in the most conservative forecast. As this range of compute capability in the cloud and at the edge becomes increasingly connected though flexible networks, we see the emergence of an ICT continuum.

In the same way that we observe insect swarms (like bees) coordinating their interactions, we expect the clouds, networks, IoT and data to enable multitudes of entities and devices to combine to form dynamic and intelligent collectives [145]. One feature of this will be localised and temporal interactions between compute nodes that combine resources to achieve a task greater than can be achieved with those nodes operating in isolation. The individual limited computing capacity of objects is complemented and supplemented by their connection to other objects in relevant communities. This new computing is emerging as a cooperative interaction between individual entities, each with their own autonomy, but working together for the benefit of the collective community. One example of this Swarm Computing would be autonomous vehicles, each capable of acting independently, but also interacting with connected objects around it and to centralised information and control centres to optimise traffic flow and improve safety. It will give rise to connected smart systems that are able to share information and autonomously regulate their performance in a concerted fashion, with the objective of optimising results, solving problems and mitigating detected risks. Depending on the specific business use cases, data will either be processed in a distributed manner by algorithms hosted in hybrid clouds, or be processed locally using algorithms hosted within or close to the connected objects and connected robots. Real-time data analytics and AI/ML will have a major impact in the automation, optimisation and flexibility of connected ecosystems and the collectively offered services. This will be especially so in industrial supply chains as they become increasingly collaborative and responsive to market demands. This will be part of next industrial revolution to create intelligent networked enterprises and service ecosystems.

As the world becomes ever more digitally and globally connected, industries of any sector (Healthcare, Energy, Manufacturing, Telecom, etc) are experiencing a digital transformation [146]. Communication Service Providers (CSPs) who evolve quickest and are most able to adapt to this new digital economy, will be the ones to thrive. The traditional methods of revenue (voice and data services) are slowing and the price pressures over these services continue. CSP are suffering in both consumer and enterprise domain and are viewed as simple pipe providers. CSPs are seeing their profits stagnate, while watching the over-the-top (OTT)

players erode their revenue [147]. The changing trend of their users moving to alternative OTT services was one of the reasons for their decision to start on the journey to digital transformation. CSPs enter in this digital economy selling new services that can be delivered and management over digital channels in areas such as consumer entertainment, mobile banking services, autonomous transportation, etc. The business opportunities are huge and to reach a market outside the consumer's area, the CSPs need to create strategic alliances with the vertical stakeholders to build and offer B2B2X (Business-to-Business-to-X) proposition in vertical sectors.

However, the telecom community still continues selling a lot of technology rather than solutions that enterprises want. Digital transformation is more than a technology drive, it encompasses end-to-end IT processes, automated operations, infrastructure management, human skills and corporate culture [148]. CSP need to digitally transform the internal organisation – we have seen actions on this direction with flexible DevOps, Zero Touch management – and also adapt the external language to properly communicate to vertical sectors.

An opportunity for CSP is to offer their new networks as self-service platforms by using a high degree of operational automation and complete customisation. Service innovation and customer experience are the main drivers for their digital transformation. Future services for connected enterprises and people will be context-aware, immersive, omnipresent, intelligent and autonomous for real end-user experiences and will bring new technical challenges to ICT infrastructure. In order to support new types of services, operators need to upgrade their IT stacks and service operations. Over decades, CSP's top priorities are network's reliability and performance. Telecom networks must be always-on, and guarantee continuous services regardless used technology, since society and any type of business depend on reliable communications.

CSPs have built up many sophisticated systems and strict operational processes around telecom networks following rigorous standards for stability and quality to the point where they guarantee their high reliability. The problem is that to adapt this legacy to be ready for the digital economy and smoothly manage the end-to-end ICT continuum is a real challenge. Legacy BSS and OSS need to evolve to real-time, automated open platforms that are fully virtualised and use microservices to deliver this agile approach to provide service assurance and the way to do business [149]. These end-to-end management platforms should be on one hand modular with a high level of resource abstraction so that can be based on multiple vendor combinations and on the other hand, also offer service capability exposure functions via open APIs to enable CSPs to partner with enterprises in vertical sectors. This will allow new digital services that combine CSPs offerings with partners' specialised domain capabilities (sophisticated content, IoT data, immersive technologies, etc.) to create innovative offerings. Bearing in mind the changes due to this ICT continuum and dynamic digital environment, there is a need for a new generation of service and network solutions for both operational and business management that enable the partnership of CSPs and vertical stakeholders while keeping an overall service assurance and carrier-grade reliability and performance.

8.2 From Software-Centric to Human-Centric Services

The continuous technology evolution, the habits of the digital consumer and the volatile market, poses big challenges for the enterprise to become more competitive and profitable; this will not be possible if organisations do not bet for innovation. The Human Centric innovation supports customers and communities in creating a prosperous and sustainable future where people are always connected. Today the concept of innovation only as the incorporation of a set of tools or management solutions is erroneous. The techno-social revolution should help customers to become more innovative and to help in the evolution of the digitalisation. As discussed in the previous section, we envisage a globally-connected continuum platform to enable new future digital services. This continuum must provide users with greater level of control, be more transparent in interactions with digital services (for instance search algorithms or data computes resulting in non-biased results presented to

users) and even introduce some ethical values and behave with certain social capabilities. New innovations are about not only digital transformation of industries and businesses but also achieve a better social inclusion.

These new paradigms will lead to a flexible and programmable architecture based to satisfy the large diversity of use cases and applications, the different vertical industries will demand. New and heterogeneous services as the autonomous cars, Industry 4.0, smart cities, remote and augmented reality, etc. will require networks beyond 5G to quickly adapt to new demands and provide more control of the network services. The transformation from hardware-centric to software-centric networks is already in place. Networks Functions (NFs) are moving from monolithic equipment's into programs running in virtualised computational pool of resources. This will also mean that networks beyond 5G will be easier to upgrade to, reducing the expensive current ones based on the replacement of physical infrastructure. Software-centric networks may become more democratic networks thanks to the open source developments enabling multi-vendor interoperability.

In the same way, the future generation of networks (beyond 5G) will go one step further from software-centric towards the concept of human-centric. Humans will be the central point and everything will revolve around humans. Human-centric business process management is an approach that considers human skills and activities first and uses automated functions to support them. Benefits can include reduced risks, higher rates of compliance, enhanced management support and improved interaction with users.

We envision that the future networks will be human in the sense that people will interact with them as they currently do with other humans. The network interfaces will recognise not only the order via commands or voice as nowadays but also gestures or moods and will be able to react based on them as it is done in a human-human communication. This extremely valuable information will be used to provide fully customised services as, for example, show a list of restaurants in the area the user is and based in his preferences and health status, thanks to the fact that the network interface can identify based in his vital signs, and the user behaviour while being stimulated by food smell.

And this is only an example. Another service that goes beyond would be the ”mind to mind communication” between human brains. In terms of services, right now the list of potential applications in beyond 5G networks fall in the field of science fiction, being one of the most present the ”mind to mind communication”.

Users will demand the future networks, beyond greater bandwidth or almost zero latency, global coverage and always-on connectivity. In this sense, we envision that humans will be part of the network, carrying microchips under the skin, so called e-skin, interfacing with multiple sensors applications as part of the service. The sensors will have the ability to compute data and perform cryptography in our bodies, or the ability to transmit and receive digital data and talk directly to machines in their digital language. There are already new companies [150] targeting ultra-high bandwidth brain-machine interfaces to connect humans and computers.

Moreover, the limitation of the connectivity in certain areas, could be enhanced with the concept ”follow me network” identifying the user location, and adapting the pool of resources infrastructure improving network service, in terms of network coverage, whereby services are following users by creating ad-hoc networks with fixed and unmanned infrastructure. On the other hand, it will be needed to adapt the networks in terms of adaptability of resources to assure the service, but also identifying the limitation that the coverage might have, be able to deprive some service capabilities to the most basic services, to assure the quality of the network and the best use of resources.

8.3 Services Everywhere, Infrastructure No Limits

In today's society, people are relying their entire daily living on the use of digital services and apps. Digital users are becoming increasingly dependent upon Internet services and moreover connectivity to those services – and this trend is clearly set to increase. Figure 11 depicts the global digital population (in millions) worldwide in January 2018.

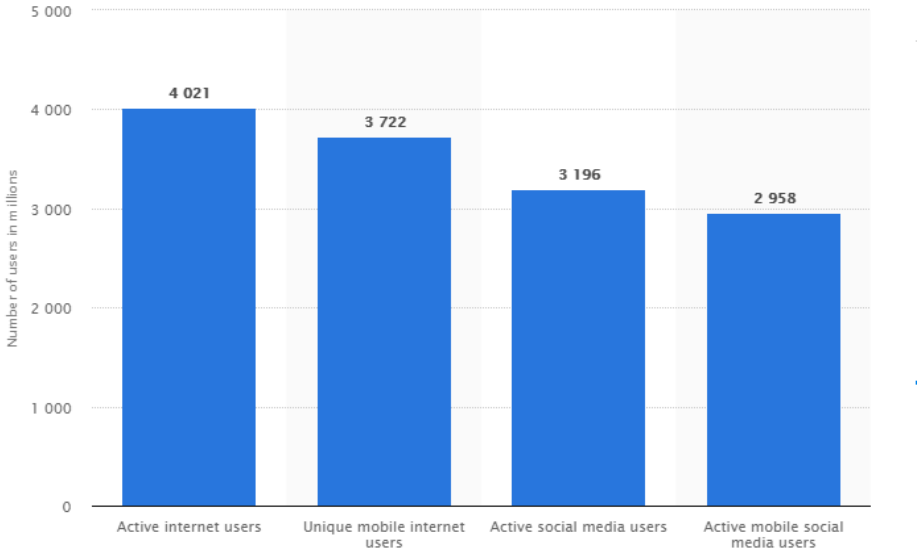

Figure 11 Global digital population [151]

Moreover, we have entered in an era, that we are not conscious of the necessity until the industry creates a new product or service that enters in our day-to-day life and rapidly becomes totally necessary. The benefits to the society have become obvious during the last years, even the users that were not born into the digital world, and has "survived" without the use of all these Internet services for their work and life, now we raised the question how did we previously survive without it? The main innovation is not just only to innovate with the youngest population, also to empower a more mature generation, and support them with easier use of the technology.

The potential and power of technology focus on the optimisation of the resources from time consuming, energy saving, reduce uncertainty, predict problems, generate experience or easier interaction, in order to provide a more productive environment. Around all these processes, technology is a must. This strategy has been possible with the exploitation of our personal data, and behaviour that we have assumed the loss of privacy and confidentiality when we accept the terms of privacy when selecting a new service, that will be used to create analytics to enhance decision and policy-making, and also to sell this information to their customers.

The digitalisation is making disruptive changes to traditional industries, products are offered on top of operators' networks. This necessity is increasing, with the digital connected society the market segment where any type of service is bringing a new level of technological sophistication to be able to access to all connected devices (robotics, immersive reality, IoT devices, autonomous vehicles), and the services acquired through the Internet: booking reservations (hotels, concerts, congress), financial, communication or social functions. The current IT infrastructure is not designed to support this intelligence everywhere; some architectural transformation and renewed focus on hardware and exploitation of communication regulations will empower companies to build the new future communication system.

At the same time, where connectivity brings convenience, it also brings challenges to cope with the increasing demand the infrastructure should cover. There are several emerging technologies that will become the backbone of the Internet, from the cloud to the edge, and everything in between. Currently there is a trend to get closer to the user, such edge architectures (MEC) will speed the maturity of technologies giving resources close to the user without relying on the core network, or personal data consolidator that will make easier the access to a user/client data. Customer hardware and hardware accelerators will meet the computing demands of intelligent environments. IT Infrastructures need to be developed to reach into the dynamic physical environments they want to serve – and it needs to happen now. Small cells (SM) are playing a vital role to better serve densely-populated areas, as part of the orchestration framework can reduce those costs by adding capacity to the network instead of new use of spectrum. The DAS can serve more than one mobile operator, so it can provide coverage and capacity for multiple carries, providing multi-tenancy. The new SC can increase the wireless backhaul, capacity and speed, locating the antennas close to the end-user and dynamically allocate the network capacity in the C-RAN (Cloud-RAN). Unlicensed SC will be at a lower power and can eliminate the need of wire connection to the equipment. The use of free unlicensed spectrum lowers the cost per bit significantly, although it has the lack of covering large distance, so to provide well coverage carries need to rely on license spectrum to provide the needed range, or to be coverage by macro cells in the wireless backhaul. The combination of both convergence of license will improve the network capacity and user experience, by relying in a SC controller-based architecture of those resources.

8.4 Network-Unaware Vertical Services

5G networks were conceived to support advanced use cases from different vertical sectors with different network requirements by means of slicing the physical network into several logical networks or slices, each slice tailored for a different vertical use case. This enables offering a flexible solution that allows the optimal configuration of necessary resources to serve a customised service, empowering different verticals, such as high-quality media on demand, health (remote surgery use cases), manufacturing (Industry 4.0) or automotive industry (autonomous driving or connected cars).

However, higher levels of abstraction are envisioned to be in place towards the verticals to make this process for them in a fully automatic and network unaware mode. A vertical stakeholder, which could be, even more in the coming years than now, not necessarily a business entity, but an individual, will not want to speak the language to request concrete network requirements for each of their requesting network services for his/her vertical applications.

Those higher levels of abstraction will be part of a network agnostic automation process that will be needed to lower the barriers to satisfy coming business and users necessities. That network agnostic process shall include also an automatic and transparent mapping to the network service consumer, who does not care about networking issues or requirements. The vertical applications (layer 7) will have to be totally network-unaware which turns out into a full automatic from human to network translation process. Future mobile intelligent applications will learn in the computing infrastructure and get balance deployments from the edge to core. The usage of AI/ML techniques will continuously improve application service delivery.

This will have to be accompanied consequently in the lower layers by higher degree of automation also in the orchestration processes that will need to combine vertical applications and network applications orchestration together as a whole. The future service architectures will have to comprise an intent-oriented service definition over abstracted infrastructure (advance models are needed), real-time telemetry of services and massive correlations, proactive adjustment of parameters to meet service intents. Literally, the network will be ‘always-on’ and automatically carry out ‘follow-me’ actions (as service motion concept, as pointed out in previous sections) to maintain QoE defined in composed SLAs. The whole architecture will embody a closed-loop structure for service life-cycle management. This

continuum will be a self-driven platform and will also perform proactive business actions such as fault isolation, prevention across multi-layer and multi-vendor environments, fraud detections, deal with trust areas, hybrid orchestrations for a global optimisation of services at scale. The combination of an intent-driven approach and AI/ML techniques for managing both network and services will bring enormous gains in service efficiency (doing things in an optimal way, faster or better), in service effectiveness (doing the right task and achieving goals) and in functionality (doing new things previously not possible, business edge).

In addition, the future ICT continuum platform (compute and network) will intelligently learn the network environment and historic data, and dynamically adapt to a changing situation and enhance their own intelligence by learning from new data. This will become a big data problem that will need to be solved together with the associated services. It will learn and complete complicated tasks, such as redistribute workload, intelligent placement, traffic load balancing associated with link utilities, autonomic network operation, keeping dynamic flows in large-scale networks, etc. The platform will be able to even predict the future network situation for proactive controlling services performance and delivery reliability across multiple networks. This is a step towards self-driven networks for advanced network-unaware services.

8.5 Extreme Automation and Real-Time Zero-Touch Service Orchestration

Future networks envisage going more and more into the need to support a ”hyper-connected world” where more challenging performance requirements are expected beyond 5G ones towards: an always-on and ultra-fast connectivity, with full world coverage everywhere and massive machine-to-machine communications from all kind of different devices. This is accompanied with a complete digital transformation in all the society in which the globalisation, trust, sustainability and automation are presented as main pillars as shown in Figure 12.

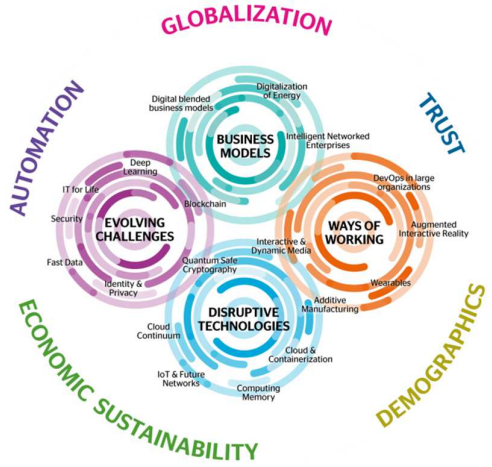

Figure 12 Digital shockwaves in Business [145]

There is a clear trend in the coming years towards the maximisation of automation in all processes and interactions (except those cases that could negatively affect customer experience). In a few years, social machines, smart contracts and other types of more advanced interaction will be a reality, in the way that some machines will be indistinguishable from people from the perspective of business process and interactions, with higher capacity of decisions, orchestrate common actions, make requests, etc.

This expected evolution should be taken inevitably also in the telco world. Future networks will require higher demands on real-time network service management and a higher degree of automation, which will be crucial to increase cost-efficient operation.

The relevance of this research work needed for automation beyond 5G is also illustrated that ETSI NFV has started a specific group on ”Zero touch network and Service Management” (ETSI ZSM ISG) that plans to focus on some of the aforementioned challenges in relation to network and service management, in order to allow all the operational processes to be executed automatically [152].

Several technical dimensions will contribute to increase the required level of automation towards extreme automation, which at the end will be driven by accomplishing real-time zero-touch orchestration including self-healing, this is, the capability of OSS remediation, and self-organised network services orchestration.

Future services of any vertical sector must interoperate with all management platform capabilities. Therefore, an open APIs should be offered to allow telecoms to partner with key enterprises in vertical sectors to properly address B2B2X scenarios. In addition, there is a need of a flexible and extensible universal template for service design, on-boarding and lifecycle management. The challenge is how to automate service behaviour updates, keep interoperability between diverse services without breaking overall end-user experience.

1. Enhanced policy management including huge data analytics

Greater policy-driven autonomic support will be required for network automatic self-healing and self-organisation [153] with no human intervention. Policy management will define operations that can be used autonomously by different network domain controllers providing activation and remediation functions, at the same time that cycle time and cost will be reduced. AI/ML is also needed to increase the degree of automation as an approach to enable the transformation from automatic functions (relying on handcrafted rules and hard coded logic) to autonomic functions that constantly and automatically learn from, and adapt to, rapidly evolving network behaviour. These data-driven functions eliminate the need for network tuning, detecting and resolving network incidents at much higher precision and significantly faster, in many cases even before they happen [154], which enables a proactive assurance, based on a performance analysis, root-cause analysis, troubleshooting and fast resolution functions.

2. Artificial Intelligent driven orchestration

Intelligent network management mechanisms using AI/ML combined with Big Data Analytics at the service orchestration platform will further optimise the network operations experience towards automation and zero-touch orchestration. It will enable a much more informed elastic management and orchestration of the network, often allowing proactive resource allocation decisions based on heuristics rather than utilising reactive approaches due to changes in the load. This approach would be in line with the goal of a recent ETSI ISG called Experiential Network Intelligence (ENI) [155], which proposes an engine that adds closed-loop AI/ML mechanisms based on context-aware and metadata-driven policies to more quickly recognise and incorporate new and changed knowledge, and hence, make automatically actionable decisions.

3. Cloud-native management applied to NFV orchestration

Cloud native refers to software built to change, scale, resilience and manageability. Cloud-native data is stored and structured in ways that encourage flexibility and gets comfortable with micro-databases that requires new levels of automation and self-service. Like cloud-native apps, cloud-native data platforms should scale up and scale out.

On the contrary, VNFs so far, have been implemented to be managed as monolithic applications by the current NFV MANO platform solutions. It is expected that the application of a cloud native orchestration approach to NFV MANO contributes greatly to increase that level of automation future networks is requiring towards real-time zero-touch orchestration or extreme automation [156].

However, this is not a trivial process. Most of the existing NFV MANO orchestration solutions now follow the principles of ETSI NFV specification, which is focused on virtual appliances based solutions, and lacks the information and guidelines to support the cloud-native architecture and environment. ETSI NFV ISG is currently proposing some modifications in their proposed architecture towards the support of a cloud native approach, last version is work in progress now on January 2018 [157].

The overall idea is that VNFs will have to be broken down into smaller microservices, and deployed as containers in both the public and private clouds. Leveraging Continuous Integration and Deployment (CI/CD), these microservices containers will be orchestrated and deployed with automation. The independent software vendors who used to produce full-fledged network functions now become the vendors of smaller microservices. In this way, it is expected to achieve:

- **Auto-provisioning:** This is the management of resources automatically facilitating on-demand, self-service, programmatic provisioning, and releasing of resources. This will enable network services to run smoothly with on-demand allocation of resources directly from the VNF packages, and automatically handling the task of data analytics, and releasing the resources back to the pool when the service is finished.
- **Auto-redundancy:** This will enable to minimise failure risks automatically. The network services will be expected to be inherently resilient to failures. They automatically handle the outages and enable corrective actions. In the event of failure, the process instantly moves from one data centre to another without interrupting the service. This is executed so quickly that the service consumer does not even know. In case of the occurrence of a partial outage in one data centre, the VNFs/Network Services will continue running seamlessly.

8.6 Service Injection Loop

The use of services in the NFV ecosystem provides existing network functions as on-demand services for Enterprise applications hosted within the cloud. Those services are created as a dynamic path to access virtual resources created on-demand and with the flexibility to be deployed in different locations. The provision and configuration of services are managed by the NFV orchestration layer in real time, to set up (and turn down as needed) suites or catalogues of connected services that enable the use of a single network connection for many services, with different characteristics, based on the available VNF to form the SFC.

Architectural micro services provide modular, distributed software components that can be deployed in any environment with a standardised infrastructure, allowing distributed applications to be installed on a cloud infrastructure while maintaining maximum flexibility. Digital transformation is requiring significant changes to the network to be reliable and provide a high level of service to comply with Quality of Services (QoS) policies.

The creation of services should be reinvented for the new digital area; the way that services are provided today, should be flexible enough and tailored to customer needs to deliver a delightful customer experience. Using NFV, SDN and cloud computing create a virtual, cloud-based network that has the flexibility to automatically meet the service requirements of a rapidly growing set of network-based services.

Now, the innovation should be driven not only in the network transformation but also in the creation of new services, this will unlock the potential of digitalisation. Like the current trend on the DevOps approach and the agile development philosophy and operation of network services should be transferred to deliver the next-generation of services. The use of predictive models is continuing to disrupt the way decision-making is occurring. This approach will create massive tailored service focused on the real need of the customer and specifically designed/customised for the user.

The main goal will be that services must interoperate with the platform capabilities, and be able to automatically adapt to the current needs and enhance the user experience. This approach can be followed by internally identifying and analysing the use of the service in order to recognise the limitations of the service. Human Centred Design (HCD) processes for interactive systems provides requirements and recommendations for design principles and activities throughout the life cycle of interactive systems.

Operators and infrastructure managers are already using AI/ML-based technologies to automate the use of the infrastructure. The requirements must be evaluated against existing operations, maintenance processes and tools. The new services should be enhanced with AI/ML also with highly specific requirements as mentioned before such as unifying strategies including carrier-grade, performance and operational capabilities.

Current work is focusing on the challenge of integrating new services in the NFV platforms, to automatize interoperability without breaking the components, and make those services available as a ”service store”. A potential approach could be achieved through a Metamodel [158] for services, to model, testing and onboarding in a runtime environment to become self-service platforms. To link them on demand, requires a new way of describing the entire platform and all its constituent endpoints at the right level of abstraction. The metamodel should be a single point of connection to support end-to-end automation of testing, onboarding and the lifecycle management process that can dynamically ingest and combine a broad range of microservices to compose a larger and complex one.

The customisation of the services based on the previous description will provide several benefits:

- Better user experience (UX), thanks to the Human Computer Interaction, based on consistency of solutions across multiple interfaces (mobile, traditional systems and human interfaces).
- Support inter-data connectivity between verticals slices to define new knowledge for defining automatic new services requirements, imposed by the service limitations and user interaction.
- Automatic test and validation of requirements and functionalities in several domains, (integration, deployment, on-boarding, lifecycle management).

For the onboarding phase, services will be complemented with testing files to be validated by the NFVO for a fully operational use, across multiple infrastructures. This approach will involve new business revenues to monetise the use of service by the user: pay-for-what-you-use services. It could also create models rather than sell a specific product: the most valuable part is going to be the output that the service offers.

9. Future and Emerging technologies

The Internet is arguably the most complex infrastructure created by mankind. It is constantly and rapidly evolving to satisfy increasingly important and diverse requirements. Its underlying network infrastructure is in the process of changing from a transport-only, data-less, dumb infrastructure to a multifaceted and distributed system mimicking a living being and consisting of a stratum of fluidified networking and computing resources, dynamically organised and managed by more and more intelligent and autonomous algorithms, which generate and exploit increasing quantities of data, and provide customised services and applications alike everywhere.

Physical connectivity supporting a transparent transfer of information is not anymore the only functionality required to the network. Intelligent algorithms are needed to make the network able to adapt and evolve to meet changing requirements and scenarios and to provide tailored services to users.

Data generated by the network and by the users need to be used by the network itself and to be exploited outside the network. Applications will be more deeply rooted within the network, to provide adaptive features, tailored to user needs, capable to better exploit network-generated data and functionality and to be (dynamically) instantiated (close to) where they are needed. Vertical industries stakeholders will be more and more involved in the communication network value chain, as integrated and distributed applications pose to the network diverse and specific demands.

The network will become even more pervasive and more integrated, further absorbing residual conceptual differences, e.g., between telephone/cellular and Internet/data networks, and imitating the structure of a living being, composed of a Physical stratum + Algorithms + Data + Applications.

Figure 13 cursorily summarises the evolution process of the telecommunications networks, showing:

i) How networks developed and evolved (left side of the figure);
ii) Past and current main trends behind such progresses (right side of the figure);
iii) Traditional and new requirements (upper side of the figure) that should be satisfied by the future instance of the network, which we call *x*G[3], and also because a softwarised network will evolve more rapidly and incrementally with new software releases, rather than with major generational leaps.

[3] For the sake of having a neutral future-proof placeholder name.

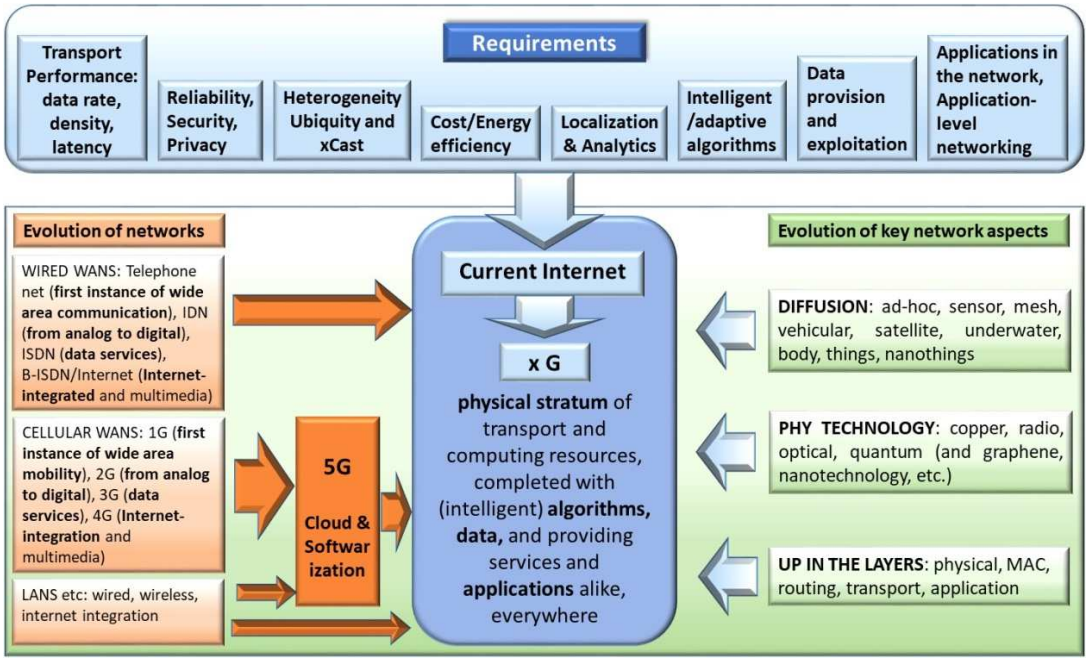

Figure 13 Overall picture of the evolution of networks

As shown in Figure 13 (left-hand side), networks have been traditionally designed and standardised as independent silos.

Wired Wide Area Networks (WANs) started as the plain old telephone system, evolved into the Integrated Digital Network (substituting analogue with digital technologies), tried to integrate services (ISDN) and improved performance (B-ISDN), but failed, due to the perhaps not fully expected deployment of the Internet in the early nineties; the aftermath is that technologies such as the Asynchronous Transfer Mode (ATM), meant to become ”universal”, became just another component of the growing Internet before disappearing.

Cellular WANs took a completely independent evolution path. The first analogue generation (1G) made a first giant leap when it was replaced by the digital GSM cellular network (2G) coming along also with the first data services (SMS, GPRS, WAP). Then 3G brought about new radio technologies, better security and full data services, and 4G made a further step towards the confluence within the Internet, which will be a major unifying feature of the emerging 5G.

In parallel, Local/Metropolitan Area Network technologies started from wired, proprietary, low capacity systems and later extended their realm in the transport network backbones. Meanwhile, the end of the 90ies gave the start to an impressive deployment of wireless LANs, which, starting from little more than a low-rate cable replacement, have now almost topped the 10 Gbps speed (e.g., the incoming 802.11ax technology), have become the most common Internet access technology in the home/office/campus, and are being integrated in the 5G network.

Meanwhile, also the networking research trends have evolved along different dimensions and key aspects (Figure 13 right-hand side). A first trend regards the diffusion of the networks both in space and in type, with specific infrastructures expanding both the reach and the functionality of networks (ad-hoc, sensor, mesh, vehicular, satellite, underwater, body, things, nano-things, etc.). A second, obvious, trend regards the improvement of physical systems, increasing the performance of copper, radio, optical, and now quantum communications.

The third trend highlighted in Figure 13 is probably subtler: networking research in the early days mostly focused on lower layers (MAC protocols, routing); more recently it shifted towards

higher layers. At the same time, also the functionality implemented in the network according to the ISO/OSI stack followed the same process: increasingly, higher layer functions are executed also within the network, rather than in end-systems only.

At this point, the reader may wonder why Figure 13 does not explicitly lists an ”Internet” box below the evolution of the networks column (and it is mentioned in the middle convergence box). The reason is obvious: The Internet was in fact not a network technology itself (in contrast with LANs, WANs, or cellular technologies or systems), but was focusing on the inter-networking among such underlay technologies, to provide end-to-end services.

However, is this approach still valid today? In the past Internet was one among the services and a marginal one if compared to the circuit-switched voice service, today the traditional differences between Telecom/communication and Internet/computer science communities is blurring (or has blurred), together with the integration of telephone/cellular and Internet/data networks. The TCP/IP paradigm started as an overlay of network technologies (the IP-over-anything frenzy), then with TCP/IP deployed in each technology, then today with TCP/IP having integrated all pre-existent network infrastructures and starting to include in the network also transport and application functionality.

In the meantime, 5G is threatening this model, being more ambitious than previous cellular generations and aspiring to play a bigger role. In fact, 5G is being designed by including also (part of) the wired/core section of the network, as well as LANs (as a matter of fact, important characteristics of 5G would be meaningless if confined to the cellular section proper, e.g., the slice concept), architecturally integrates cloud/fog systems, and is natively a software network, providing differentiated service support. Thus, the current Internet and 5G are almost on a par with each other and converging to an integrated fully inter-operable network.

In fact, 5G is not just an evolution of 4G in terms of performance, but creates a breaking point with respect to previous generations, with an end-to-end architecture which blurs and crosses many boundaries with respect to our historical definition of Internet and inter-networking. Moreover, the focus on simultaneously supporting (eventually via slicing) widely different services and ”vertical” applications will make of 5G a much larger ecosystem, including more stakeholders than in the past, with more complex relationships, more heterogeneity and more dynamicity. Last but not least, availability of (big) data, cloud integration and softwarisation will provide a viable technological infrastructure and an in-network ”knowledge plane” – quoting Clark et. al’s vision [159] – which, if today is ”just” used for cost saving and flexibility, may turn into an unprecedented and pragmatic ”arm” for realising the ultimate promise of a truly cognitive network.

To complete the picture, the upper part of Figure 13 reports traditional and new requirements that should be satisfied by *x*G networks. From left to right, we start with traditional needs expressed in terms of *performance of the network*, continuing with *ubiquity and xcast (broadcast is an important requirement for 5G)* and the more and more felt *reliability/security/privacy*, and *cost/energy efficiency*, which are cornerstones of the 5G network. Then we enter in more advanced and recent requests made to the network, which include *localisation and physical analytics*, *intelligent/adaptive functionality*, *data provision and exploitation* and finally the need of a tighter *coupling and interactions between telecommunication services provided by the network and applications* and the opportunity arising from directly providing applications as a network service.

This process leads quite naturally to the emergence of *Living and Fluid Networks* [160], namely network architectures holding the ability to autonomously change to best adapt to context. In *Living and Fluid networks*, algorithms and protocols will be capable of understanding what they are used for, and will tailor their algorithms and parameters to best suit the requirements of the different applications, appearing in different ways to different packet flows, behaving differently for each one of them, and providing the desired performance to each. Unlike the early cognitive networking insights, we believe that the time for such

networks has now come, due to the reached maturity of their two fundamental and complementary aspects:

- the unprecedented modularisation and ”fluidification” of computing and networking building blocks brought about by modern programmable networking trends (i.e. the ”fluid” aspect), and
- the ability to take autonomous (or semi-automated) cognitive orientation strategies and reconfiguration decisions, as provided by the extremely effective and increasingly pervasive modern AI/ML approaches (i.e. the ”living” aspect), and the intertwining of data and functions inside the network.

Key to this evolution is the availability of: i) better underlying technologies, drastically improving communication and computing performance; ii) new techniques for network softwarisation and related primitives and interfaces; iii) intelligent and autonomous algorithms; iv) data; v) applications integrated with the network, performing in part also networking functionality.

This chapter provides a view of the main future technologies and trends behind the evolution of the network as sketched above and is correspondingly structured in three subsections: the physical stratum, algorithms and data, applications.

9.1 The Physical Stratum: Communication and Computing Resources

9.1.1 Nano-Things Networking

The many ”Things” we are progressively interconnecting in the Internet are progressively extending to the micro-things, i.e. those computational and service elements that run on small/tiny and non-intrusive things. Nano-communications are emerging to extend the reach of smart control to the level of molecules and cells, with unprecedented impact in medicine and material manufacturing [161]. Combating diseases via autonomous nano-machines, ultra-fast degradation or toxic waste, self-healing and self-monitoring materials constitute a few of the most visionary applications. Materials with software-defined electromagnetic behaviour constitute applications presently under development, paving the way for programmable wireless environments [162].

Recent research on nanomaterials and nano-network architecture components (nodes, controllers, gateways) are opening new prospects of usage of nano-scale things. At the PHY Layer, graphene antennas enable nano-communication within the 0.1 − 10 THz spectral window, which promises unprecedented communication data rates despite the nano-scale. At the MAC Layer, pivotal protocols have been studied to target mainly Body Area Network (BAN) applications [163] and self-monitoring and adapting industrial materials [164].

Despite these initial promising results, there is a need to provide a more in-depth view and modelling of the network architecture and communication mechanisms in this field, which needs to address various challenges like channel modelling, information encoding and more efficient protocols which allow energy-optimised nano-networks communications.

Critical research challenges to be addressed in the area of nano-networks include:

- Classifying nano-communication paradigms per application scenario. A generalised and unified nano-node architecture is difficult to be obtained. Specific hardware and protocol designs must be produced for each envisioned application, to ensure that the limited nano-node size is optimally exploited in terms of the specifically required functionalities.
- Experimentally validated, application-specific communication channels. Pivotal studies in nano-networks have showcased the workflow to be followed for deriving communication channel models. Such models are important for crafting efficient, application-specific hardware and software for nano-nodes. Thus, it is important to

systematically repeat and expand these studies in a wide set of targeted application areas, ranging from solids to biological material.

- Solving the power supply problem. Studies must address the power supply problem in the case of all nano-network types incorporating electronic modules. To date, state-of-the-art autonomic power supplies are sizeable (approximately mm-scale) and of low-capacity. Wireless power transfer and carrier-powered approaches can abolish the need for batteries, under the condition that they are properly designed for their environment and application scenario.
- Solving the battery problem with nano-electronics, fast charging, explosion proof batteries are a must for the future.
- Cost-efficient, massive nano-node integration and production. Separate nano-node components have already been manufactured and tested. The next step is to fully assemble nano-node prototypes, including all separate components. In the case of electronic manufacturing, the related processes exist but yield a high cost. New approaches are required to produce massive numbers of nano-nodes at a low cost and in rapid design/prototyping cycles.
- Hardware-software co-design. Presently, different protocols and software stacks are being developed, mostly from an exploratory angle, given the immaturity of the underlying nano-node hardware. The general consensus is that the software and protocols for nano-networks will indeed face severe limitations in terms of complexity, varying per application scenario. Thus, nano-networks call for co-designed hardware and software right from the start. In other words, the usual workflow of creating general-purpose hardware first and then developing the required software is most likely not feasible in the case of nano-networks.
- Security and Safety. The design of nano-nodes must ensure compliance with authentication and privacy standards, whose severity and criticality vary per application. For instance, consider a biomedical scenario, with a swarm of nano-nodes within a patient's body. Finally, regarding safety, the presence of nano-nodes within an environment must be well-studied, to ensure that it does not upset its function in an undesired manner.

9.1.2 Bio-Nano-Things Networking

The Internet of Things (IoT) has evolved to Internet of Nano-Things (IoNT) inspired by the nanomaterials recently discovered such as graphene and metamaterials enabling the development of networks of nanoscale size embedded computing devices, called nano-things. IoNT will revolutionise various application areas such as military, healthcare, and manufacturing, due to the very tiny, concealable, implantable, and non-intrusive nano-things cooperatively sensing, actuating, processing and networking. IoNT can be the basis of many applications such as smart grids, intelligent transportation, environmental monitoring, healthcare systems, and home automation. However, the artificial nature of IoNT devices can be detrimental for some environments such as inside the body or in natural ecosystems, where the deployment of nano-things and the electromagnetic communication could result in undesirable effects in health or pollution.

The novel paradigm of Internet of Bio-Nano-Things (IoBNT) is introduced here by stemming from synthetic biology and nano-technology tools that allow the engineering of biological embedded computing devices. By stemming from an analogy between a biological cell and a typical IoT embedded computing device, a cell can be effectively utilised as a substrate to realise a so-called Bio-Nano-Thing, through the control, reuse, and reengineering of biological cells' functionalities, such as sensing, actuation, processing, and communication. Since cells and their communication are based on biological molecules and biochemical reactions rather than electromagnetic waves, IoBNT is expected to be paradigm shifting for communications and networking fields. The execution of DNA-based instructions, the biochemical processing of data, the transformation of chemical energy, and the exchange of information through the

transmission and reception of molecules, called molecular communication (MC), are at the basis of a plethora of applications that will be enabled by the IoBNT such as i) intra-body sensing and actuation where BNT's collect health data and release drugs inside the body ii) intra-body connectivity control where BNT's diagnose and/or repair communication failures between internal organs iii) environmental control and cleaning where BNT's collaboratively check for toxic agents and transform them through bioremediation, e.g. bacteria employed to clean oil spills.

Bio-Nano-Things are defined as uniquely identifiable basic structural and functional units that operate and interact within the biological environment. An analogy can be drawn between a biological cell, which is the basic unit of life, and a typical IoT device since both can perform tasks and functionalities such as sensing, processing, actuation, and interaction with each other. Thanks to the advancements in synthetic biology, it is possible to control, reuse, modify and reengineer the cells' structure and functions which enables effective use of biological cells as programmable substrates to realise BNT's as biological embedded computing devices. In particular, the engineering of biological circuits through genetic code manipulation allows specifically designed functions to be performed by the cells such as AND and OR logic gates, switches and counters. Furthermore, artificial cells assembled from bottom up with minimal structural components and functions compared to natural cells can be ideal substrates for synthetic biology with a more predictable behaviour. Although very promising, the aforementioned technologies should provide solutions to major challenges in biotechnology. Reliable mathematical models and computer simulations need to be developed to capture peculiarities of underlying biological processes with intrinsic non-linearity and noise. Also, reproduction and mutation pose extra challenges.

Due to the very small size, previously described nano-things can perform meaningful operations when they communicate and coordinate with each other. As the design of BNT's, the communication among BNT's is inspired by nature where the exchange of information between cells is based on the synthesis, emission, propagation, and reception of molecules, called molecular communication. In MC literature, many systems have been proposed such as calcium signalling based on Ca+2 exchange among neighbouring cells in muscle or heart tissues for short range, bacterial chemotaxis and conjugation where bacteria is loaded with information encoded in the genetic material by conjugation and sent to swim to the receiver by chemotaxis for medium range and endocrine communication inside the body among the cells of distant organs by the propagation of hormones through the circulatory systems. Main challenges in communication of BNT's lie in the mapping MC into the classical communication systems, and in the use of tools from systems and information theory with the final goals of modelling and analysing the main telecommunication characteristics and performance, such as range, delay (latency), capacity, and bit error rate.

Bio-Nano-Things are expected to not only communicate with each other, but also interact into networks, which will ultimately interface with the Internet. To this end, the definition of network architectures and protocols on top of the aforementioned MC systems is an essential step for IoBNT development. A further challenge for the IoBNT is the interconnection of heterogeneous networks, i.e. composed of different types of Bio-Nano-Things and based on different MC systems. A solution might come from the natural way our body manages and fuses several types of information to maintain a stable, healthy status, or homeostasis. Calcium signalling within a cell can trigger release of hormones to the circulatory systems which in turn control processes such as blood pressure, growth on the distant receiving cells. Biological circuits based on these processes could effectively provide a set of genetic instructions that mimic the classical gateways between different subnets on the Internet.

Finally, the realisation of interfaces between the electrical domain of the Internet and the biochemical domain of the IoBNT networks will be the ultimate frontier to create a seamless interconnection between today's cyber-world and the biological environment. A main challenge is to accurately read the molecular characteristics where information is encoded and translate them into the modulation of electromagnetic waves. This can be achieved by

novel nanoscale chemical and biological sensors composed of materials characterised by electrical or electromagnetic properties that can be altered by the presence of specific molecules. Electronic tattoos or artificial cells encapsulating electromagnetic antennas can also be considered as potential bio-cyber interfaces.

IoBNT can revolutionise biomedical technologies and improve human health and quality of life. A possible application scenario is using IoBNT for continuous monitoring and early detection of infections earlier than regular methods relying lab culture. To accomplish this system, a tiny implantable BNT composed of electronic circuits and genetically engineered cell based biosensors that eavesdrop the quorum sensing (QS) communication of bacteria inside the body can be designed. A gateway BNT will transfer the collected info about the infection from the other BNT's and relay it to a wearable hub for the disposal of healthcare professionals.

With the aim of realising minimally invasive, heterogeneous and externally accessible electrical/molecular communication channels between implantable or wearable electrical and biological IoBNT devices, microbiome-gutbrain axis (MGBA) can be exploited. The molecular information exchanged among bacteria inside the gut is translated into electrical signals by the enteric nervous system and transported to the brain and other IoBNT devices inside the body connected to the nervous system. Hence, MGBA can be considered as a IoBNT communication network infrastructure.

9.1.3 Quantum Networking

During the last 20 years a new generation of systems based on the intrinsic quantum properties of the physical stratum has been progressively made available. Such systems consist of a classical interface communicating with the standard equipment on one side, and a quantum sub-system sufficiently decoupled from the environment that is constituted by individual atoms, photons and charge carriers (either electrons or holes), controlled by the former. Encoding the information by quantum objects enables the implementation of quantum computing algorithms and communication protocols. Such algorithms rely on the principles of quantum mechanics. The features of the physical stratum at the quantum limit allow encoding information by novel degrees of freedom such as the spin, to exploit quantum superposition of states, entanglement and to rely on the impossibility of cloning states. Quantum systems for networking are generally subdivided between quantum communication systems and quantum computing systems.

Quantum communication will play a central role in the creation of the next generation of secure telecommunication networks. Quantum communication relies on the use of quantum resources to achieve tasks that cannot be reproduced with classical theory. Because quantum communication involves numerous technologies, platforms and application, recommendations on protocols, components and infrastructures require continuous update.

Quantum key distribution (QKD) based on either single or entangled photons relies on the capability of detecting an intruder because of its effect on the quantum states used to encode the secure one-time-pad key. The physical stratum for quantum communication consists of single photon and entangled photon sources, and single photon detectors. Several physical implementations of various protocols have been developed by disparate materials, operating at different wavelengths and consequently capable of peak performances at some temperature, ranging from cryogenic to quasi-room temperature. Nevertheless, true single photon sources operating at room temperature still lack, so low cost solutions can be deployed by using weak coherent lasers for which the communication protocols (such as BB84) are adapted by more sophisticated ones (such as decoy-states protocol). Single-photon detectors that combine high performance (high detection efficiency and low dark counts rate), with low cost possibly by integration in the silicon photonics platform still lack. In order to spread the development and the employment of quantum communication, the implementation of quantum sources and detectors by integrated photonics operating at room temperature is mandatory.

Such hardware is complemented by quantum random number generators, currently targeting 1 Gbps, based on intrinsic randomness of quantum processes, ranging from single-charge related electronic noise to photon arrival time. The most successful system is currently based on the conversion into a stream of bits of the arrival time of photons emitted by a weak source as detected by an array of 1024 SPADs in silicon in 45/65 nm technology node.

Quantum Key Distribution protocols are currently the most advanced among the secure quantum communication protocols. However, to be competitive with existing security technologies two main areas of development are identified.

The first area concerns security. Various demonstrations of quantum key distribution systems have been made over the last years. Mainly, the security proofs of all these demonstrations still rely on that the communicating parties have full control of their local ideal devices and therefore there is no information leakage except that needed for the protocol. For quantum encryption to fully guarantee security, the protocols used must be independent of the type of preparation or measurement performed, or the internal workings of the devices used to make the measurement. Hence realistic device independent protocols [165] should be developed and implemented. Short term (3 years) and medium term (6 years) targets are > 10 Mbps and > 100 Mbps at metropolitan distance.

The second area concerns the performance and the application. Device independent protocols address security loopholes issues introduced by imperfect or untrusted systems. Parallel to the development of these protocols, proposition or implementation of protocols for new applications or improvement of secure transmission performances have to be proposed. Among the potential method allowing the improvement of the secure transmission, high dimensional protocols [166] is a way to increase the capacity and to enhance the robustness against eavesdropping. New challenges are to: i) propose different implementations, in particular different from those using orbital angular momentum of light to make the method compliant with existing single mode fibre networks, ii) Extend the transmission up to 10 km in fibre links.

Apart from the Quantum Key Distribution, other quantum secure communication protocols should be extended beyond laboratory proof-of-principle demonstrations. Here a non-exhaustive list is presented: quantum multiparty communication protocols [167], quantum public-key cryptography [168], quantum secure direct communication [169], quantum digital signatures [170].

Free-space and optical fibre are the most commonly used transmission channels of QKD systems. Numerous proof of concept experiments has been performed demonstrating long distance transmission both in fibre [171] and in free-space links [172].

Free-space QKD can address the increasing demand for security in handled devices for short distance applications, e.g. secure transmission to ATM terminals, or can allow the development of global scale QKD network using satellite communications. The development of on-chip and relatively low-cost devices operating at 810 nm for short distances, or 1550 nm for satellite communication (immunity against the daylight) would be a key enabler for the realistic implementation of QKD for free space applications.

In fibre networks, extend the transmission distance beyond 400 km remains a challenge due to the intrinsic fibre loss. Two main strategies can circumvent this limitation. The first strategy would be the development of trusted backbones, which allow meshing the long distances. This strategy entails also the development of efficient, low cost, on chip sources, detectors and manipulating components operating at 1550 nm. The second strategy consists in implementing quantum repeaters, which is still challenging.

Alternatively, quantum communication based on continuous variables relies on the encoding by modulating the signal below the quantum noise level. The recipient of the communication splits a beam into idler and signal, so to be the only one capable to cancel the noise on the returning beam modulated by the transmitter.

Quantum networking for quantum key distribution deals with the emission of photons and their detection, so the communication ends by destroying the quantum state by the measurement action. Another option consists of enabling communication of quantum states between quantum processors distributed at the nodes of the network, so to create the equivalent of a computer cluster connected through a quantum Internet. Such distributed quantum computing is also called networked quantum computing.

A distributed quantum computer based on quantum networking requires quantum processors including one or more qubits, communication lines for photon transmission, optical switches and quantum repeaters for transport of qubits along distances, because of lack of amplification of quantum states at a fundamental level. The quantum processors may consist of arrays of quantum logic gates involving either defect centres in semiconductors from cryogenic to room temperature depending on the defect and the host material [173], or cryogenic ion traps or cavity quantum electrodynamics. The communication lines by either optical fibres or free space have been already discussed. Quantum switches may rely on the matter-radiation interaction such as single atom-photon based switches, capable for instance to switch the phase of the quantum state [174]. In order to extend the communication range, which is limited by both decoherence and losses, quantum repeaters based on cryogenic rare earths-based components are required. Alternatively, a suitable protocol scheme applicable at room temperature has also been proposed, with the advantage of not involving cryogenic equipment [175]. Another direction towards quantum simulation is the Ising-Hamiltonian model (a mathematical model of ferromagnetism in statistical mechanics) that is represented by room temperature networks of optical parametric oscillators as coherent Ising-machines [176].

Quantum systems have been demonstrated by several materials at cryogenic temperature. In order to spread their employment in standard telecommunication systems for a novel level of quantum based security, an effort is required to integrate such quantum systems into silicon photonics and integrated photonic platforms from quasi-room to room temperature.

9.2 Algorithms and Data

Softwarisation/cloud and security concepts are assumed as already included in the current network architecture.

9.2.1 Impact of AI/ML on the Network

During the last years, the use of AI/ML solutions has reached a great popularity, attracting several innovation activities and growing investments. As a matter of fact, since a few years we have been witnessing that AI/ML is one of the key enabling technologies capable of paving the way to the Digital Transformation of Telecommunications. In fact, AI/ML is impacting the three major techno-economic challenges that Operators are facing: simplifying the networks architectures (to provide any sort of digital services, with shorter time to market and better QoS); cloudifying/edgefying the virtual network functions and services; optimising and automating OSS/BSS processes to mitigate the increasing ”complexity”, dynamisms and pervasivity of the infrastructures.

On the service side, from autonomous driving to speech recognition, a plethora of functional applications have appeared in completely different business areas e.g., Internet of Things, Tactile Internet, Immersive Communications, Automotive, Industry 4.0, Smart Agriculture, Omics and E-Health, etc. The use of the huge data lake generated by the infrastructure will allow automating processes by introducing cognitive capabilities at various levels. Examples of cognitive capabilities include: understanding application needs and automating the dynamic provisioning of services; monitoring and maintaining network state; dynamically allocating virtual network resources and services; ensuring network reliability and enforcing security policies.

For example, services such as autonomous cars have latencies requirements that are so strict (e.g., order of ms) that it is not possible ”to close the loop” executing them with cloud computing

solutions. The deployment of local processing and MEC (Multi-Access Edge) solutions can help mitigating this problem, but it requires management/control and orchestration capabilities capable of integrating on-device, edge-based and cloud-based AI/ML-systems.

On the other hand, the management complexity of such future infrastructures (e.g., for FCAPS (Fault, Configuration, Accounting, Performance, Security) and orchestration of virtual resources and services will overwhelm human-made operations, thus posing urgent needs of designing and deploying OSS/BSS with AI/ML features. At the same time, the use of AI for the network will reduce the amount of person-power needed to deploy and operate the infrastructure, thus reducing the operational costs. Moreover, AI/ML can fuel the generation of new services that may lead to improved sustainability models for the network operators.

Changing Network Design

From the today's perspective AI/ML will be, will enable innovative features when provisioning future digital cognitive services for homes, businesses, transportation, manufacturing, and other industry verticals, including the smart cities. In future scenarios, the increasing usage of End-Users' devices (i.e., smart-phones or tablets) together with the centralised and distributed computational resources will encourage the move of the computational and memory/storage resources from huge data centres towards the edge of the network (e.g., MEC).

Moreover, the huge amount of data sent by new AI/ML applications will lead to hybrid architectures where the data may be partially analysed/compressed in the edge of the network to speed up the whole process and save network resources. In particular, we will see applications able to execute the first layers of a deep neural network locally or in the edge to finish the execution in powerful data centres.

Furthermore, we expect a significantly increase in the amount of machine-to-machine communications with an increasing number of sensors and other IoT devices continuously monitoring smart cities, Industry 4.0, smart energy, etc.

Automated Operations and Network Intelligence

Today, many Telecom Operators are still relying on manual management processes, but there is a clear awareness of the potential for using AI-powered solutions for automation thus reducing costs, increasing productivity, and driving more value. The rationale is to use AI for automating the operations processes based on collection and elaboration in (almost) real time of data about states and level of performances of nodes/systems and logical/virtualised resources etc. For example, AI/ML can automate the management, control and orchestration (e.g. MCO) processes of physical pieces of equipment, which today are mostly carried out by humans, introducing control loops acting on virtual/logical entities (e.g., Virtual Machines, Containers, appliances etc.). In this direction, AI/ML promises to deliver scalable OSS/BSS functions based on AI/ML models capable of seeing and interpreting the state of millions of network entities via the analysis of huge data streams. Moreover, network and service computational intelligence (e.g., in the Radio Access Networks and in the Core), based on data about Customers' service patterns and traffic would allow improving the quality of the customer experience whilst optimising the use of resources.

Reducing Network Costs and Smart OPEX

In general, the use of AI/ML methods and algorithms will decrease both the costs of deployment and the costs of operations of the network in the following years. This technology will learn the correct network behaviour, being able, in a first step, to help understanding possible problems and anomalies, and finally, autonomously acting over the network to correct those problems. For example, to reduce the cost of deployment, the AI/ML will be able to offer zero-touch network configuration for the most common network deployments. This will reduce both the time needed for new deployments and the manpower needed to achieve a proper configuration.

Moreover, different learning techniques will be used to predict the behaviour of the network. This will lead to better provisioning of resources in the network, avoiding the nowadays-typical situation where the networks are over-dimensioned. For example, AI/ML will also enable the adoption of ”QoE” models and indicators to support investment and design processes based on a data-drive approach (e.g., selection of deployment regions, strategic priorities, etc).

Eventually, regarding OPEX optimisation, it is well known that energy consumption is one of the major cost items for Network Operators: AI/ML methods and systems will allow using the data lake for implementing performance analysis and optimisation methods for energy consumption versus quality of service.

Creating new Services Using Network Data

It is likely that the appearance of new services powered by AI/ML will bring significant socio-economic impacts, together with improved sustainability models for Network Operators. Among these services, those ones able to improve both privacy and security levels will be of great importance.

The appearance of personal data platform (tightly connected with the network service) is also expected that will allow Internet users the control their data. To this end, solutions will appear to analyse the network traffic in a privacy preserving and controlled way.

Cybersecurity

Future Networks and 5G will have to face all the security challenges typical of today’s telecommunication infrastructures, but with a new and IT-oriented perspective brought by SDN and NFV. Nevertheless, these same enabling technologies, integrated with AI/ML will provide new instruments to mitigate such risks. To mention some examples: inferring proactive actions (even based on early-warning signals of attacks) allowed by AI/ML; adoption of flexible and automatic features for fast traffic steering (e.g., quarantine, honey pots, slicing segregations); automatic configuration of security virtual appliances to be added into the service chains.

Conclusions

The effective applications of AI/ML methods and systems in future 5G scenarios are likely to require multi-domain orchestration of distributed processing in the terminals/devices (could be e.g., Fog Computing), at the edge of the network (e.g., MEC) and in the cloud computing facilities.

In this direction, the end-to-end interoperability is a must and it requires more standardisation efforts and further achievements. First of all, it is necessary to consider the impact of current, and future, AI/ML systems and methods in the functional architecture of 5G. This means understanding which and how architectural functional blocks will be impacted, and what will be the related standardised interfaces. In this direction a global effort is still required from both hardware and software vendors to participate in standardisation bodies, including collaborations with Open Source communities (e.g., Linux Foundation, ONF, OCP-TIP).

9.2.2 Impact of IoT on the Network

The realisation of the Internet of Things vision has already gone through several profound transformations in recent years. In this context, we can clearly identify:

- a first generation of communication and network architectures and protocols, such as those proposed in the context of EPCglobal (https://www.gs1.org/epcglobal), mostly aimed at supporting the exchange of data produced by RFID systems;
- a second generation of networking solutions (6LoWPAN – for „IPv6 over low power Wireless Personal Area Network and CoAP), mostly aimed at making *things* equipped with low capability devices reachable through the Internet and enabling web programming in the resulting environment; and

- a third (the current) generation of solutions (e.g., NB-IoT, LoRaWAN, virtualisation technologies) aimed at supporting the interactions between *things* and some service running in the cloud within silo-ed *platforms*.

We can easily foresee that most of the effort in the next decade will be devoted to the development of solutions aimed at supporting the seamless integration of the above platforms and then at going beyond the *Internet of Platforms* model. Therefore, we will analyse here the impact of IoT on the network based on what we expect it is happening and is going to happen shortly.

Seamless integration of existing platforms requires applications to access IoT resources through some identifier, independently of their native platform, their hardware characteristics, and the protocols executed to interact with them. The corresponding services are today demanded to the application layer but we can expect that they will become major components of the network itself. Such services should be distributed, should not be under the control of a single (or a few) player(s), should support resource discovery enriched with means for reputation management.

We expect that such solutions will start from the work carried out within EPCglobal in the context of the so-called *Object Name Service* and within the IRTF group ”ICNRG” for what concerns the application of Information Centric Networking techniques to the IoT. Also, concepts will be exploited introduced in the context of peer-to-peer systems, for what concerns the creation of a distributed catalogue of existing IoT resources, and in the context of the Social Internet of Things, for what concerns the creation service discovery and reputation management.

Also, the above solutions will be designed so that they are ready to support the next expected leap forward in IoT evolution, which envisions that individual IoT resources are not bounded to a specific, isolated platform. In other terms it is necessary that they support the case in which IoT resources owned by individual users are used by third party applications. In this way users become *prosumers* of IoT services which requires appropriate new authentication and accounting solutions. Starting point in this context will be the ongoing activities within the IETF Authentication and Authorisation for Constrained Environments (ACE) WG which is working on authenticated authorisation mechanisms for accessing resources hosted on servers in constrained environments and has completed a comprehensive use case document (RFC 7744).

In any case, the major feature of such context will be heterogeneity along several dimensions: access technology, identification/naming/addressing scheme, traffic patterns, deployment extension, device capabilities, etc. Such heterogeneity calls for a network which is highly flexible in all its segments**,** well beyond what is possible to achieve with current software defined networking and network function virtualisation technologies. In fact, while SDN/NFV mostly focus on the programmability of the behaviour of the network infrastructure, in IoT it is crucial to make the protocol stack of end devices programmable as well, so that they can react promptly to changes in the working environment. Also, slicing, which is one of the major concepts exploited to support several logical networks with heterogeneous behaviours on top of the same physical infrastructure, needs to be profoundly revised in the IoT contexts. In fact, in several IoT scenarios the same piece of information transported by the same packet can be of interest of several applications with very different QoS requirements. Such a frequent case cannot be supported by the current implementation of slices that are partitions of the packet space.

Also, the amount of data generated by the IoT is expected to increase with the number of devices at a pace that is orders of magnitude higher than the available data rates. This trend is not sustainable unless radical changes in the Internet infrastructure are introduced. This means that the Internet, which is mostly a communication infrastructure today, must turn into a *computing and communication* infrastructure capable of executing data processing and fusion in any of its components.

The raise of interest towards edge cloud and edge computing goes in this direction; however, the process must go well further and should impact the Internet architecture in its fundamentals. In fact, by turning all network switches/routers into computing nodes the Internet will become a huge and pervasive network of *middleboxes* and several assumptions that are at the very basis of the TCP/IP protocol stack will not be valid anymore.

Finally, it is clear that the true IoT revolution will happen only if a reasonable level of security can be guaranteed. In this context, work is needed to go beyond the work carried out by the "*DTLS In Constrained Environment"* (DICE) WG that has produced a TLS/DTLS profile that is suitable for constrained devices. In fact, a recent Internet Draft has been produced by the IRTF "Thing to Thing Research Group" (T2TRG) which provides an overview of open security issues in the IoT domain.

9.2.3 Impact of Blockchain Technologies on the Network

We posit that the integration of Blockchain technologies in the Internet infrastructure itself, opposed to application-specific add-ons, will emerge as one of the major and most impactful innovation trends in the Future Internet. As discussed in the following, our belief is that *permissioned* blockchains will gradually extend beyond the very specific single-application realm of most of the today's use cases, and will hold the promise to emerge as an open large-scale trust infrastructure, duly controlled and regulated unlike the current massively deployed permission less technologies laying at the foundation of today's crypto currencies. Such a trust infrastructure, while complementing the Internet's connectivity and data distribution services, will likely shape as a federation of independent (and mutually untrusting) providers. Current distributed ledgers' anarchy will most likely be replaced by a form of control and coordination loosely mimicking the way in which multi-domain/multi-country Internet regulation bodies and authorities are today governing and steering the operation of competing Internet Service Providers and autonomous systems.

The Dawn of Blockchains: The Era of the "Wild"

Even if the three underlying baseline technology dimensions inside Blockchains root back to works carried out many years before (hash chains and Merkle Trees in the 70ies, consensus protocols in the 80ies, and smart contracts [177] in the 90ies), Blockchains – as we know them today – emerged only in 2008, as the technical foundation and enabler of Bitcoin [178], the first fully decentralised (peer-to-peer) digital/virtual currency. The massive interest in Bitcoin emerged because of its ability to permit transactions without any trusted financial institution intermediaries managing them. Indeed, the Bitcoin's Blockchain, as well as any other Blockchain technology behind the subsequent crypto currencies (Ethereum, Ripple, Litecoin, Cardano, Iota, etc – 1583 at the time of writing) is *completely wild*. With this terminology, we mean that anyone willing to deploy time and resources (e.g. computing power in the case of Bitcoin's Proof-of-Work), not only can participate in building – mining – the relevant blocks, but might in principle even try to bias or change its operation. In fact, in Bitcoin, a new crypto currency can be deployed by "just" convincing a critical mass of block miners to adopt different rules – see the many "hard forks" popped up just in 2017 (Bitcoin Cash, Gold, Diamond, Private, etc. – we leave the reader to judge which of these initiatives were really necessary in solving real problems).

Even more interesting is the case of Ethereum: the utter flexibility of the relevant scripting language (a Turing-complete language called Solidity) permits anyone to easily create new applications on top of its blockchain by simply programming a "smart contract". Despite the hype, and the huge perceived potential in fields also outside crypto currencies, it is fair to say that such flexibility does not nearly come along without concerns[4], and has to date mainly used

[4] Among the many disasters, see for instance the catastrophic Ethereum's DAO hack in June 2016 or the case of the Parity wallets, severely hacked as much as twice in one single year, 2017.

to launch new coins, often of questionable value – see https://uetoken.com for a very ironic Initial Coin Offer (ICO) which explicitly names itself ”Useless Ethereum Token” and self-describes it as (verbatim quote): ”*the world's first 100 % honest Ethereum ICO: you're going to give some random person on the Internet money, and they're going to take it and go buy stuff with it”*. Perhaps not so unsurprisingly, given the current level of hype, even such a clearly fake ICO (Initial Coin Offering) ended up in being traded for real, gathering as much as 310.445,00 ETH (Ethereum)!

The Emergence of Permissioned Blockchains

Even if emerged in the above discussed *wild* context of crypto currencies, most of the industry is nowadays understanding that blockchains may bring a significant value also in many concrete application domains, as a shared ”database” replacement. In this direction, great business attention is currently posed on the so-called ”permissioned” blockchain technologies (e.g., Multichain, Hyperledger, etc.), whose somewhat controlled/federated trust model permits them to circumvent the scalability issues and resource consumption (e.g. energy) which affects their public counterparts.

But what is a ”permissioned” blockchain? Quoting a crystal-clear explanation by Gideon Greenspan, leader and developer of a permissioned Blockchain technology called Multichain, ”*the core value of a blockchain is to enable a database or ledger to be directly shared across boundaries of trust, without putting any single party in charge. A blockchain lets a group of actors achieve real-time reconciliation of validated, authenticated and timestamped transactions, without the cost, hassle and risk of relying on a trusted intermediary”* [179]. In other words, blockchains are clearly pointless in contexts where there is a trusted intermediary which guarantees that what you read from its database is ”true”. But they do unleash their full value when you need a shared (append-only) database, with multiple writers which do NOT trust each other, and without any trusted intermediary which may validate (and hence guarantee) that what writers are registering in the shared database is truthful.

This latter point – explicit and upfront validation of every transaction prior to storing it in the ledger – is what makes blockchains very different from ordinary databases. Indeed, the trustworthiness of the information contained in a blockchain is accomplished by the *joint* involvement of three complementary techniques:

i) a way to make sure that a transaction recorded at a given time cannot be modified in the future – i.e. the hash-pointer block structure which guarantees storage integrity;
ii) a way to resolve differences among different replicas of the blocks – this is accomplished by a suitable consensus protocol, and
iii) a way to explicitly verify that a transaction being stored is valid, via a suitable formal script associated to the transaction and ”executed” prior to adding a transaction to the ledger.

The key advantage of permissioned blockchains with respect to their ”wild” public counterparts is that not everybody can create blocks and add them to the chain, but only the subset of parties that have been granted an explicit *permission* to do. This fact completely changes many underlying technical requirements, and permits to significantly widen (and make explicit) the consensus protocols employed [180], improve scalability, guarantee fork-less operation (e.g. with signature-based consensus), improve timestamping and time necessary until a transaction is guaranteed to be registered of orders of magnitudes with respect to the today’s Bitcoin hours. Most notably, an upfront fixed number of ”miners” permits to get rid of the need to defend against Sybil attacks, the primary reason which mandates the impressive waste of energy in the Bitcoin’s Proof-of-Work.

Blockchains as Internet Infrastructure Extension?

A further advantage of a properly implemented permissioned blockchain also resides in the possibility to further control *who, specifically,* can create a smart contract (in other words, an application on top of the blockchain), and how.

Many readers might of course strongly complain that the presence of *controlled* parties which manage the chain, along with the possible restrictions set forth in terms of smart contracts' deployment (or even permissions to transact on the chain) are in sheer contrast with the original decentralisation reasons that have led to the invention and emergence of the Bitcoin's blockchain. While in principle we highly value full decentralisation and freedom, it is a matter of fact that lack of any form of control may easily yield abuses, scams, and fakes. The previously mentioned Useless Ethereum Token is a blatant example of how users, lacking the ability and the instruments to thoroughly vet ICOs, may fall into false ones. And, arguably well beyond the discussion carried out in this section, but still related to trust, the problem of data quality and fake information circulating over the Internet is arguably one of the most challenging and widely open Internet threats.

The point we wish to make here is that public, large-scale, infrastructure variants of permissioned blockchains, extending beyond the realm of a specific application, and rather providing a platform, managed by a controlled multiplicity of non-mutually-trusting "trust providers", may permit to share explicitly validated information across boundaries of trust. To remark that a large scale permissioned blockchain governed by agreements between independent countries and relevant authorities might not be impossible, it is worth to note that a controlled set of multiple competing providers is exactly the model at the basis of the today's Internet! Such a large-scale trust infrastructure may come along with a Copernican revolution, and turn the burden of verifying the validity of a claim from the end user to the infrastructure itself. In other words, a data or transaction is valid when it is recorded in the chain, thus relieving the verification burden from the end layman's user. And validity is clearly specified by a validation script (a smart contract) deployed following a clearly specified governance model enforced by the permissioned blockchain infrastructure itself.

Before concluding, to bring evidence that our thesis might not be too far to come, we remark that initial steps in this direction have been already made. For instance, the "Certificate Transparency" initiative [181] launched in 2014 by Google gave end users and domain owners the possibility to transparently verify that a formally valid certificate (i.e. correctly signed by a certification authority) was really issued to the domain owner, thus mitigating the problem of fake TLS (Transport Layer Security) certificates. While Google's massive-scale block-based data structure leverages Merkle Trees and closely reminds a ledger, it is still purpose-specific (tailored to the very special case of TLS Certificates), is not meant to support decentralisation and shared ownership (via consensus), and – most notably – lacks any formal validation of the data inserted, which only a scripting language may provide.

With the growing understanding and maturity of permissioned blockchain technologies, with the support of policy makers for identifying the appropriate governance models loosely mimicking the way in which the multi-domain Internet is today controlled by multi-country Internet regulation bodies and authorities, and with the help of technicians for identifying the necessary extensions in the technological platform, such future is not too far to come.

9.2.4 Evolution of Protocols

Several technological trends will affect protocol development in the following years. These include:

1. Achieving ultra-low latency end-to-end communication is now recognised as the most important goal for many applications that will become ubiquitous in the years to come (e.g., networked virtual and augmented reality, automation, etc.). Moreover, for some of these applications (e.g., Augmented Reality (AR) / Virtual Reality (VR)) both ultra-

low latency and very high data rates are required, so the traditional latency-throughput trade-off will not be longer applicable.

2. The capacity of access links rapidly increases, especially in the wireless domain. Additionally, hosts can now efficiently use multiple interfaces as if they were a single resource. Users make good use of the higher total capacity, as consumption and production of high-bandwidth video have also been rapidly increasing. The net effect is that, after more than a decade of almost-certainty, it is today much less clear that congestion always appears in access links. Measurements have shown that core peering interconnections can also be throughput bottlenecks for traffic on end-to-end Internet paths.
3. At the same time, the infrastructure underlying 5G networks exposes increasingly diverse characteristics with e.g. Visible Light Communication links, millimetre-wave links and modern WiFi standards. All these access technologies have in common that they no longer emulate the behaviour of a static-capacity bottleneck. The increasing dynamicity of the exposed behaviour is further intensified by a shift towards greater mobility of both humans and machines (increasing usage of cellular networks, and intrinsically mobile usage scenarios such as Vehicular Networking or Unmanned Aerial Vehicle (UAV) networks).
4. Internet communication patterns have changed, in the sense that connecting to nearby Content Distribution Network (CDN) servers has become the most common way of consuming / using popular Internet services and content, instead of connecting to servers that are far away. Trends such as fog computing will increase the ”locality” of communication for many users (both humans and machines) and applications.
5. Increased flexibility in both in-network devices and networking software in end-hosts is becoming the new norm. The former has become malleable as they are changing from a static hardware design to software-based designs (Software Defined Networking (SDN), Network Function Virtualisation (NFV)). For the latter, developments such as the Internet Engineering Task Force (IETF)’s work on Transport Services (TAPS) and user-level protocol stacks are paving the way for avoiding ossification and making networking stacks more adaptive and future-proof.
6. Finally, security, privacy and trust have moved from being an afterthought in the design of new communication protocols, to an absolute necessity in the face of a growing and ever-evolving threat landscape.

Several of these trends conflict with the traditional layering in the Internet, where TCP/IP protocols interconnect applications across any underlying link layer technologies, and transport-layer congestion control optimises the sending rate. For example, TCP cannot handle quickly changing bottlenecks well and assumes a static bottleneck capacity (conflicting with trend #3) and causes delay by filling buffers (the ”bufferbloat” phenomenon, conflicting with trend #1). TCP is also ”blind” to the underlying technologies, even when there may only be a few hops across one or two types of link layers between a CDN server and an end user (trend #4 is an unused opportunity), and such a short path might be swiftly adapted and controlled in software (trend #5 is an unused opportunity).

Some developments that partially address these trends have surfaced: for example, Information Centric Networking (ICN) focuses almost exclusively on content distribution. The (mostly US-American) industry has been developing methods to improve the performance of the Internet’s transport layer, as well as making it more secure and more flexible; examples of such developments include Multipath TCP (MPTCP), new Active Queue Management (AQM) algorithms, the QUIC transport protocol, novelties in Explicit Congestion Notification (ECN) usage, and new types of congestion control. However, it is unclear whether these point solutions will be flexible and robust enough to both satisfy the needs of upcoming and future applications, and be suitable for 5G network technologies and beyond.

The increasing heterogeneity and dynamics of the underlying infrastructure will necessitate greater flexibility, both in end systems and inside the network. Internet transport protocols will

have to be exchangeable at run time. Also, better interplay between applications and the underlying network will be necessary. This will enable dynamically mapping the service needs of applications to the current network infrastructure. Inside the network, long-term traffic engineering by deploying hardware will of course prevail, but traffic engineering controlled by humans using software will be replaced by automation and new protocols that learn both from historical data and traffic conditions in real time. AI/ML techniques and data analytics will be key drivers of self-adaptation and self-management, both in network nodes and in the protocol stacks of end hosts. However, all these solutions are still in their infancy – at best – and will require important research efforts before they can be widely used and deployed.

Even more drastic solutions could try to address all the problems related to the trends above by challenging the traditional role of protocol layering. A possible step in this direction are recursive network architectures, as they allow to react faster by tightening the control loop, thus solving problems closer to where they occur. Could we envision a future where TCP/IP would only be used as a rendezvous protocol in the common case, and all communication would switch over to an entirely different technology when this different technology is found to work for a (typically short) end-to-end path?

9.3 Applications

9.3.1 Application Level Networking

The continued growth in the video space and the push for increased quality, interactivity, and personalisation of media content, coupled with the widespread introduction of augmented and virtual reality, and increasingly heterogeneous and mobile platforms, challenges network performance, and will require new approaches and solutions. Surveillance and monitoring, whether fixed feeds or drone-based on-demand monitoring for disaster management, event security, etc., will further complicate the space; as will the growth in real-time sensor data distributed via machine-to-machine networks for control and management of industrial facilities and smart cities. Virtualisation of applications and their supporting services, enabling ubiquitous deployment via cloud and fog computing services, poses novel infrastructure management challenges. And the need to support innovation and overcome the performance limitations of edge devices requires new APIs and programming models.

The ongoing shift of TV distribution from broadcast onto the Internet will accelerate, driven by the need to transition spectrum to interactive services, cost constraints for all but the most popular content, and the desire to personalise and customise content to suit user interests, to support targeted advertising, and to match device capabilities. The complete transition of such content onto the Internet will involve at least a 10x increase in video traffic volume, yet video already comprises > 75 % of Internet traffic (Cisco Visual Networking Index, 2017). The ongoing transition towards 4k and 8k video, high dynamic range colour, and higher frame rates will further drive the traffic load. The desire for interactivity (e.g., dynamic viewpoint selection, augmented- and virtual-reality) further impacts load and introduces strict latency bounds for an effective user experience. The implications on application level networking are tremendous: the existing protocol stack cannot meet the needs of such applications and must evolve. It is necessary to move away from video as a specialist service, and rather integrate video content, live or pre-recorded, within the web infrastructure and content model. This does not mean abandoning quality of experience or quality of service guarantees – such will become ever more critical – but rather integrating those with the web content framework, delivery model, and APIs, to make video becomes addressable, accessible, and embeddable, and a fundamental part of the web experience.

The initial steps in this process are visible in the WebRTC standards, developed by W3C and IETF, that began the integration of real-time content into browsers – exposing novel APIs for capture, playback, and processing of real-time audio-visual content in web applications. The process is set to continue, with deployment of HTTP/2, QUIC, and future versions of the MPEG DASH and CMAF standards enabling convergence of real-time media and the web. But this

is only the start – deep integration of multipath, to make effective use of ultra-dense and diverse wireless networks, is essential, as is effective multicast and multiparty distribution. Both require transport protocols and web infrastructure evolution, since they change the delivery and security models, and require effective trust delegation and novel security mechanisms.

Increasing network capacity and quality of service, deployments of ultra-dense wireless, and related 5G technologies, will make live upload increasingly possible and relevant for breaking news, live sporting and entertainment events, to augment and replace traditional broadcast coverage. It will require and enable live contribution feeds, editing, and content composition. Raw video content will increasingly be available to augment and supplement professional content, and will be edited and processed live or near-live. This will push requirements for contribution bandwidth and quality of service, edge storage and compute, and edge processing for high quality and capacity video content. As with professional video production, user video contribution will increasingly transition from being a pseudo-isochronous feed to be a contribution of tagged-frames, with rich metadata including and geolocation and social context, integrated instantly into the web infrastructure for viewing, processing, and redistribution.

Video provenance will become a key issue, to combat ”fake news” and the effects of AI/ML-generated video that attempts to subvert legitimate content; these pose strong risks to the integrity of political and societal discourse, news, and the reputations of public figures, organisation, and events. While solving this is primarily a societal, political, and legal problem, the wide deployment of data provenance and signing infrastructure, to ensure the veracity of content before reputable organisations will distribute it, can support solutions in this space. Strong, vendor and government neutral, approaches are needed here, that must be multinational and verifiably outside the control of any single operator, vendor, or government to limit accusations of censorship and bias. This plays into the security and integrity of applications, network transport, and in-network processing.

To support these developments, the network must evolve to support highly distributed content, stored, processed, and delivered from a pervasive fog computing infrastructure, with effective quality of experience management. The security challenges are immense: how to ensure integrity and provenance of data, through multiple layers of caching, processing, and distribution, while maintaining privacy. Similarly, for meta-data management, quality of experience, and quality of service for media delivery, transport, and processing.

User device performance is strongly limited by thermal and battery constraints, despite the impressive growth in mobile compute performance. Edge compute, in the form of fog- and cloud-computing, virtualised infrastructure, offers an impressively scalable and flexible platform for off-loading processing, provided such processing scales in a parallel form. We need new APIs, replacing the venerable Berkeley Sockets, to address edge compute limitations, and user-space and kernel bypass networking protocols are increasingly needed to match application performance to the performance of the network. There are challenges in supporting the range of applications and protocols: more flexible APIs are needed, to align with application uses of the network, to support the increasing range of transport services (security, reliability, timeliness, quality of service, quality of experience, application intents, offload) needed by modern applications, and to support innovation by democratising network application development – excessive specialist knowledge is required to develop effective applications in this space.

Precursor projects such as NEAT and Post Sockets set the direction for novel APIs, and are beginning to set the direction for future API standards, but are just the starting point for the evolution in this space. Applications must express their needs for (partial) reliability, timeliness, robustness, security, quality of service, etc., in an abstract manner – requirements are mapped to underlying network capabilities, taking account network control information, load, wireless infrastructure capacity, and the need to co-exist with other virtualised applications. The policy

framework below such an abstract API will increasing take account not just user, application, and system policy, but also the state of the entire network, interacting with the SDN control place, NFV services, and virtualised fog compute resources to determine what network paths, features, protocols, and resources are available for the application to use. The traditional network API provides no support – the network of the future must do so in a manner that doesn't overwhelm applications, developers, operators, or network managers. We must abstract the complexity and enable intelligent control, while supporting policy choice and application needs.

9.3.2 Applications (Components) in the Network

One of the key developments in the network architecture is the deep integration of application and service functionality pervasively within the network. This is not just data centre and cloud computing resources, but the integration of programmable processing resources throughout the network: in the core, the edges, and pervasively. The concepts of fog computing apply, but also software-defined networks, network function chaining, virtualisation, and container provision. There are numerous challenges are developing this vision.

Service discovery is essential. Existing mechanisms rely on a combination of anycast routing, domain name system (DNS)-based identifier-to-location mapping, and application-specific directories to locate services. Anycast routing abuses the Internet routing system to route to the nearest replica of a service, but scaling it to large numbers of services bloats the Internet routing tables and is not sustainable. Parallel trends, including the transition to IPv6 and the subsequent use of IPv6 addresses as content identifiers – e.g., the Glass-to-Glass Internet Ecosystem proposal to give each frame of video content a unique IPv6 address, and similar approaches to service and container identification – already push routing scalability to its limits. Alternative routing algorithms, e.g., practical Compact Routing algorithms, or clean slate content centric networking architectures may help scale the routing infrastructure, but there are many open questions on how these will work. Directory services also have limitations around ensuring consistency, update performance, and scaling. The architecture will have to become much more dynamic, since the network of the future will no longer be addressing O(1000s) of proxies for a small number of centralised sites, but O(billions) of sophisticated data management and processing services within the network.

Service provisioning, management, and security are critical. A pervasive service platform is essential to supporting the applications of the future, but if implemented and architected incorrectly has the potential to be a significant platform for malware, surveillance, and denial of service. We must learn how to effectively manage billions of devices, ensuring that they are suitably configured, running appropriate software, kept up-to-date with security updates and patches, and run only properly authenticated and authorised applications. We have some solutions that approximate this in the cloud computing world, for managing large scale data centres, but these are much smaller scope than will be needed to support the services in the future – by several orders of magnitude – and make assumptions about the homogeneity of the user base, services, applications, platforms, and infrastructure that will not hold as the network scales. The network will become increasingly heterogeneous, devices will not be directly accessible from a centralised management node, will be subject to differing access and security policies, and will have strong and varying restrictions on usage. Management tools will have to adapt to support this heterogeneity, and differing policies, usages, and requirements. Existing cloud computing infrastructure, homogeneous and managed by, and in support of, a single organisation is not suitable, nor is it a good model going forward.

Security models must evolve. Some infrastructure will continue to be owned and operated by enterprises, network operators, and application/service providers, with controlled access to the data centres where it is hosted, but there will be increasing use of untrusted or partially trusted physically insecure infrastructure located in residential properties, public locations, or end-user devices. Tools for secure boot, code signing, and cryptographic verification of the execution environment will become critical. As will tools to manage and control data access,

management, and provenance. Techniques such as homomorphic encryption, that allow devices to process data without having access to the data they are processing, have potential in this space, but are currently too slow and limited to be realistic – development is needed.

Authentication of services and service providers, while accounting for resource usage, is an essential part of the economics of the network of the future. Micropayments will become a key part of the system. The infrastructure to support in-network services and applications is not free – the CAPEX to deploy the underlying network and computational platforms is great, as are the OPEX to manage, power, provision, and support these services. Billing and accounting models for centralised services are already complex – how can we support, manage, and control costs for services scaling to billions of nodes, within a heterogeneous infrastructure into which the service provider has little-or-no visibility?

Privacy and data management, location of processing and data to match legal and moral restrictions on data distribution, access, and processing become increasingly important. Many of the services and applications envisaged operate on, process, and deal with personal data, that is increasingly – and rightly – subject to strict regulation, control, and limitation. We do not have good tools to reason about, describe, and discuss, how data can be processed, where it is to be located, and how it can be distributed – not in human language, legal language, or code. Policy descriptions, rules, and constraints will need to be specified in a form that can be enforced by the infrastructure on the services, since direct human oversight is not feasible at the scales considered. We have no good models, languages, or tools in this space, yet they must urgently be developed if the service-based model of applications is to scale while retaining any user trust.

Finally, novel programming models and languages will be needed to support these services, applications, and deployments. We accept that initial deployments will be based on existing infrastructure – Linux containers, virtual machines, and traditional network programming models and protocols – but these are clearly insufficient for the network envisaged for the future. The Linux container model is not secure, is not portable, and offers a programming interface that is too broad an interface to be made secure, and insufficiently expressive to meet future needs. Many existing programming languages and APIs are not type or memory safe, making services difficult to reason about, monitor, and control. Virtual machines with well-defined semantics offer one approach, as do novel APIs and programming models such as CloudABI and Capicsum, that constrain and control what operations a service can perform. Languages like Erlang offer compelling concurrency and fault tolerance mechanisms, but are weak at service orchestration, management, discovery, control, and security. The application of programming research via languages like Rust or Idris offers potential to go further, as do platforms such as Singularity, but there is still a long way to do before we can have confidence that services and applications perform as intended, and respect data protection regulations and user privacy.

9.3.3 Applications Making Specific Demands to the Network

The final interface to consider is that between applications and the intelligent, pervasive, and service oriented network of the future. It is clear this interface must broaden, and offer more compelling, easier to use, services and features to meet the needs of future applications. The traditional networking API – the Berkeley Sockets API – is not fit for purpose. It is too low-level, too limited, and does not expose the dynamic, changing, nature of the network, nor the high-level services and features needed to support modern applications.

At the transport layer, the network must evolve to allow applications to make effective use of the resources offered by the network. The increasingly heterogeneous and dynamic nature of the network is not exposed by traditional APIs. We must make applications aware that the network is not constant, and that the services and features offered depend on the location in which the application terminal is located. Available services, and their accessibility, vary from location to location, as does name resolution, routing, transport performance, and policy and

privacy constraints. Network functionality varies – yet this is not exposed to the applications or higher layer protocols. At minimum, we must expose this heterogeneity to the applications: making the very different network functions and features visible, supporting multi-path and multicast services where offered, and allowing applications to understand their network environment, probe and use appropriate transport protocols, and make best use of the features and services offered by the network in which they find themselves. The Transport Services and Post Sockets work in the IETF Internet standards community is a first step in this direction, moving beyond the limitations of Berkeley Sockets and starting to embed a policy-compliant intelligence into the stack to help applications make best use of the network, but it is only a start and more needs to be done.

Higher-layer protocols must be enabled. The network API of the future will not be merely a transport API. Rather, the goal is to instantiate local support services and network functions to support the application, and move appropriate data to the user location – subject to privacy and policy constraints – such that processing can happen nearby, meeting quality of experience latency bounds. The move is away from networking APIs, towards pervasive distributed systems that can run irrespective of the underlying physical infrastructure, and that are location, policy, and regulatory environment aware.

There are numerous systems that require such support. Simplest are perhaps video content distribution applications, delivering increasingly interactive audio-visual content from service and content providers to users, but also uploading user-generated content to viewers. More latency and performance sensitive applications will follow: augmented and virtual reality, gaming, business support, etc., that require local data and processing to meet latency bounds and service the needs of the application, tracking user behaviour and supporting interactive applications that must predict, and respond to, user needs in real time. More demanding still are applications such as teleoperation of remote devices, surgery, healthcare, autonomous systems, conversational interfaces, and other interactive and pervasive services. These have strict latency and quality-to-service bounds to meet user expectations, and will increasingly rely on sophisticated, multipath, transport services, novel and secure transport protocols, and the ability to spawn local data stores and computation in support of the applications.

At present, we have no effective APIs, protocols, or interaction models to instantiate such services. The cloud infrastructure we offer is low-level, starting virtual machines or containers that run on specific operating systems and hardware infrastructure, with little in the way of support services – the Linux/Unix model is not suitable for the applications of the 21st century, but we have yet to settle of a type-safe, memory-safe, secure programming environment for the future, nor have we begun to develop the APIs and services that applications can use to understand the network, service, and regulatory environment in which they find themselves, or the data that they must manage, have available, or can distribute. The challenges in offering this are immense, and rely on the effective integration of network transport services, pervasive computing infrastructure, policy, and data management.

10. References

[1] United Nations: Sustainable Development Goals, August 12, 2015, http://www.un.org/sustainabledevelopment/sustainable-development-goals/.

[2] United Nations – Broadband Commission for Sustainable Development 2025 Targets: ”Connecting the Other Half”. http://www.broadbandcommission.org/Documents/publications/wef2018.pdf.

[3] EU Commission: Digital Agenda Scoreboard – The EU ICT Sector and its R&D Performance. 2017, http://ec.europa.eu/newsroom/document.cfm?doc_id=44503.

[4] World Bank: Exploring the Relationship Between Broadband and Economic Growth. Michael Minges, World Development Report, 2016, http://documents.worldbank.org/curated/en/178701467988875888/pdf/102955-WP-Box394845B-PUBLIC-WDR16-BP-Exploring-the-Relationship-between-Broadband-and-Economic-Growth-Minges.pdf.

[5] McKinsey & Company: The Internet of Things: Mapping the value beyond the hype. McKinsey Global Institute, June 2015, https://www.mckinsey.de/files/unlocking_the_potential_of_the_internet_of_things_full_report.pdf.

[6] EU Commission: Digital Single Market. Making the most of the digital opportunities in Europe. 2017, https://ec.europa.eu/digital-single-market/en/policies/shaping-digital-single-market.

[7] EU Commission: Europe's Digital Progress Report 2017. Commission Staff Working Document, SWD (2017) 160 final, https://ec.europa.eu/transparency/regdoc/rep/10102/2017/EN/SWD-2017-160-F1-EN-MAIN-PART-27.PDF.

[8] ETNO: Accenture Study ”Lead or Lose – A Vision for Europe’s digital future”. https://etno.eu/digital2030/people-planet-prosperity.

[9] K. Sakaguch et al: “Where, when, and how mmWave is used in 5G and beyond,” IEICE Transactions on Electronics, vol. E100-C, no.10, pp.790-808, 2017.

[10] Ericsson: Traffic Exploration. https://www.ericsson.com/TET/trafficView/loadBasicEditor.ericsson.

[11] WorldoMeters: Western Europe Population. http://www.worldometers.info/world-population/western-europe-population/.

[12] Cisco: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016 - 2021 (White Paper), [Online]. Available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[13] 3GPP: 3GPP TS 22.261, Service Requirements for the 5G System, (Release 15), 2017.

[14] M.S. Islim, R.X. Ferreira, X. He, E. Xie, S. Videv, S. Viola, S. Watson, N. Bamiedakis, R. V. Penty, I.H. White, A.E. Kelly, E. Gu, H. Haas and M. D. Dawson: “Towards 10 Gb/s OFDM-based visible light communication using a GaN violet micro-LED”, Photonics Research, vol. 2, no. 5, pp.: A35-A48, 2017.

[15] T. Cogalan and H. Haas: “Why would 5G need optical wireless communications?” in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1-6, 2017.

[16] Y. F. Huang et al: “17.6-Gbps universal filtered multi-carrier encoding of GaN blue LD for visible light communication,” in Proc. of the *Conference on Lasers and Electro-Optics (CLEO)*, pp. 1-2, 2017.

[17] J. Fakidis, S. Videv, H. Helmers and H. Haas: “0.5-Gb/s OFDM-based laser data and power transfer using a GaAs photovoltaic cell,” *IEEE Photonics Technology Letters*, 2018.

[18] H. Haas, L. Yin, Y. Wang and C. Chen: “What is LiFi?,” *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1533-1544, March, 2016.

[19] D. Tsonev, S. Videv and H. Haas: “Unlocking spectral efficiency in intensity modulation and direct detection systems,” IEEE Journal on Selected Areas in Communications, vol. 33, no. 9, pp. 1758-1770, Sept. 2015.

[20] I. F. Akyildiz, J. M. Jornet and C. Han: “Terahertz band: Next frontier for wireless communications,” Physical Communication (Elsevier) Journal, vol. 12, pp. 16–32, 2014.

[21] A. Nikpaik, A. H. M. Shirazi, A. Nabavi, S. Mirabbasi and S. Shekhar: “A 219-to-231 GHz frequency-multiplier-based VCO with 3 % peak DC-to-RF efficiency in 65-nm CMOS,” IEEE Journal of Solid-State Circuits, vol. 53, no. 2, pp. 389–403, 2018.

[22] H. Aghasi, A. Cathelin and E. Afshari: “A 0.92-THz SiGe power radiator based on a nonlinear theory for harmonic generation,” IEEE Journal of Solid-State Circuits, vol. 52, no. 2, pp. 406–422, 2017.

[23] W. R. Deal, K. Leong, A. Zamora, B. Gorospe, K. Nguyen and X. B. Mei: “A 660 GHz up-converter for THz communications,” in Proc. of the IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS), pp. 1–4, 2017.

[24] H.-J. Song, K. Ajito, Y. Muramoto, A. Wakatsuki, T. Nagatsuma and N. Kukutsu: “Uni-travelling-carrier photodiode module generating 300 GHz power greater than 1 mW,” IEEE Microwave and Wireless Components Letters, vol. 22, no. 7, pp. 363–365, 2012.

[25] S.-W. Huang, J. Yang, S.-H. Yang, M. Yu, D.-L. Kwong, T. Zelevinsky, M. Jarrahi and C. W. Wong: “Globally stable microresonator turing pattern formation for coherent high-power THz radiation on-chip,” Physical Review X, vol. 7, no. 4, p. 041002, 2017.

[26] T. Nagatsuma, G. Ducournau and C. C. Renaud: “Advances in terahertz communications accelerated by photonics,” Nature Photonics, vol. 10, no. 6, p. 371, 2016.

[27] Q. Lu, D. Wu, S. Sengupta, S. Slivken and M. Razeghi: “Room temperature continuous wave, monolithic tunable THz sources based on highly efficient mid-infrared quantum cascade lasers,” Scientific reports, vol. 6, 2016.

[28] A. C. Ferrari, et al: “Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems,” Nanoscale, vol. 7, no. 11, pp. 4598–4810, 2015.

[29] I. F. Akyildiz and J. M. Jornet: “Graphene-based plasmonic nano-transceiver employing HEMT for terahertz band communication,” U.S. Patent No. 9,397,758 issued on July 19, 2016 (Priority Date: December 6, 2013).

[30] J. M. Jornet and I. F. Akyildiz: “Graphene-based plasmonic nano-transceiver for terahertz band communication,” in Proc. of the 8th European Conference on Antennas and Propagation (EuCAP), The Hague, The Netherlands, April 2014.

[31] I. F. Akyildiz and J. M. Jornet: “Graphene-based plasmonic nano-antenna for terahertz band communication,” U.S. Patent No. 9,643,841, issued on May 9, 2017 (Priority Date: April 17, 2013).

[32] J. M. Jornet and I. F. Akyildiz: “Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks,” IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Emerging Technologies in Communications, vol. 31, no. 12, pp. 685-694, Dec. 2013.

[33] J. M. Jornet and I. F. Akyildiz: “Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band,” IEEE Transactions on Wireless Communications, vol. 10, no. 10, pp. 3211-3221, October 2011.

[34] C. Han, A.O. Bicen and I. Akyildiz: “Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band,” IEEE Transactions on Wireless Communications, vol. 14, no. 5, pp. 2402–2412, 2015.

[35] J. M. Jornet and I. F. Akyildiz: “Femtosecond-long pulse-based modulation for terahertz band communication in nanonetworks,” IEEE Transactions on Communications, vol. 62, no. 5, pp. 1742-1754, May 2014.

[36] C. Han and I. F. Akyildiz: “Distance-aware bandwidth-adaptive resource allocation for wireless systems in the THz band,” IEEE Transactions on Terahertz Science and Technology, vol. 64, no. 5, pp. 2130-2142, May 2016.

[37] T. Kürner and S. Priebe: “Towards THz communications - Status in research, standardization and regulation,” Journal of Infrared, Millimeter, and Terahertz Waves, vol. 35, no. 1, pp. 53–62, 2014.

[38] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath: “Spatially sparse precoding in millimeter wave MIMO systems,” IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499–1513, 2014.

[39] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors and F. Tufvesson: “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40–60, 2013.

[40] E. Torkildson, U. Madhow and M. Rodwell: “Indoor millimeter wave MIMO: Feasibility and performance,” IEEE Transactions on Wireless Communications, vol. 10, no. 12, pp. 4150–4160, 2011.

[41] E. Bjornson, E. G. Larsson and T. L. Marzetta: “Massive MIMO: Ten myths and one critical question,” IEEE Communications Magazine, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[42] A. L. Swindlehurst, E. Ayanoglu, P. Heydari and F. Capolino: “Millimeter-wave massive MIMO: the next wireless revolution?” IEEE Communications Magazine, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[43] I. F. Akyildiz and J. M. Jornet: “Realizing ultra-massive MIMO communication in the (0.06-10) terahertz band,” Nano Communication Networks (Elsevier) Journal, vol. 8, pp. 46-54, March 2016.

[44] I. F. Akyildiz and J. M. Jornet: “Ultra massive MIMO communication in the terahertz band,” U.S. Patent No. 9,825,712, issued on November 21, 2017 (Priority Date: December 6, 2013).

[45] O. El Ayach, R. W. Heath, S. Rajagopal and Z. Pi: “Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays,” in IEEE Globecom, pp. 3476–3480, 2013.

[46] X. Huang, Y. J. Guo and J. D. Bunton: “A hybrid adaptive antenna array,” IEEE Transactions on Wireless Communications, vol. 9, no. 5, pp. 1770-1779, 2010.

[47] S.-H. Wu, L.-K. Chiu, K.-Y. Lin and T.-H. Chang: “Robust hybrid beamforming with phased antenna arrays for downlink SDMA in indoor 60 GHz channels,” IEEE Transactions on Wireless Communications, vol. 12, no. 9, pp. 4542–4557, 2013.

[48] C. Lin and G. Li: “Indoor terahertz communications: How many antenna arrays are needed?”, IEEE Transactions on Wireless Communications, vol. 14, no. 6, pp. 3097–3107, 2015.

[49] C. Lin and G. Li: “Adaptive beamforming with resource allocation for distance-aware multi-user indoor terahertz communications,” IEEE Transactions on Communications, vol. 63, no. 8, pp. 2985–2995, Aug. 2015.

[50] 3GPP: 3GPP TR 38.801, Study on New Radio Access Technology, (Release 14), 2017.

[51] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago and F. Wiedmann: “5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications,” IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, Feb. 2014.

[52] Y. Medjahdi, S. Traverso, R. Gerzaguet, H. Shaiek, M. B. Mabrouk, D. L. Ruyet and a. D. R. Y. Louet: “On the road to 5G: Comparative study of physical layer in MTC context,” IEEE Access, vol. 5, pp. 26556-26581, 2017.

[53] Y. Tao, L. Liu, S. Liu and Z. Zhang: “A survey: Several technologies of non-orthogonal transmission for 5G,” China Communications, vol. 12, no. 10, pp. 1-15, Oct. 2015.

[54] R. Gerzaguet, N. Bartzoudis, L. G. Baltar, V. Berg, J.-B. Doré, D. Kténas, O. Font-Bach, X. Mestre, M. Payaró, M. Färber and K. Roth: “The 5G candidate waveform race: A comparison of complexity and performance,” EURASIP Journal on Wireless Communications and Networking, no. 13, 2017.

[55] M. van Eeckhaute, A. Bourdoux, P. de Doncker and F. Horlin: “Performance of emerging multi-carrier waveforms for 5G asynchronous communications,” EURASIP Journal on Wireless Communications and Networking, 2017.

[56] F. R. Kschischang and S. Pasupathy: “Optimal nonuniform signaling for Gaussian channels,” IEEE Transactions on Information Theory, 39(3), pp. 913-929, 1993.

[57] N. S. Loghin, J. Zöllner, B. Mouhouche, D. Ansorregui, J. Kim and S.I. Park: “Non-uniform constellations for ATSC 3.0,” IEEE Transactions on Broadcasting, 62(1), 197-203, 2016.

[58] G. Boecherer, F. Steiner and P. Schulte: “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” IEEE Transactions on Communications, 63(12), pp. 4651-4665, 2015.

[59] M. Pikus and W. Xu: “Bit-level probabilistically shaped coded modulation,” IEEE Communications Letters, vol. 21, no. 9, pp. 1929–1932, Sept. 2017.

[60] T. Prinz, P. Yuan, G. Boecherer, F. Steiner, O. Iscan, R. Boehnke and W. Xu: “Polar coded probabilistic amplitude shaping for short packets,” in Proc. IEEE SPAWC, 2017.

[61] P. Shulte and F. Steiner: “Shell mapping for distribution matching,” arXiv:1803.03614, 2018.

[62] W. Xu, M. Huang, C. Zhu and A. Dammann: “Maximum likelihood TOA and OTDOA estimation with first arriving path detection for 3GPP LTE system,” Transactions on Emerging Telecommunications Technologies (ETT), 27 (3), pp. 339-356, 2016.

[63] H. Wymeersch, G. Seco-Granados, G. Destino, et al: “5GmmWave positioning for vehicular networks,” IEEE Wireless Communications, pp. 80–86, Dec. 2017.

[64] A. Dammann, R. Raulefs, S. Zhang: “On prospects of positioning in 5G,” in Proc. IEEE International Conference on Communication Workshop (ICCW), 2015.

[65] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich and M. Coulon: “Direct localization for massive MIMO,” IEEE Transactions on Signal Processing, vol. 65, no. 10, pp. 2475-2487, May 2017.

[66] R. di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson and H. Wymeersch: “Location-Aware Communications for 5G Networks: How location information can improve scalability, latency, and robustness of 5G,” IEEE Signal Processing Magazine, vol. 31, no. 6, pp. 102-112, Nov. 2014.

[67] Y. Polyanskiy: “A perspective on massive random-access,” in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2523-2527, 2017.

[68] H.A. Inan, P. Kairouz and A. Ozgur: “Sparse combinatorial group testing for low-energy massive random access,” arXiv preprint arXiv:1711.05403, 2017.

[69] J. Luo and D. Guo: “Neighbor discovery in wireless ad hoc networks based on group testing,” In 46th Annual Allerton Conference on Communication, Control, and Computing, pp. 791-797, Sep. 2008.

[70] E. Paolini, C. Stefanovic, G. Liva and P. Popovski: “Coded random access: applying codes on graphs to design random access protocols,” IEEE Communications Magazine 53, no. 6, pp. 144-150, 2015.

[71] S. Rangan: “Generalized approximate message passing for estimation with random linear mixing,” in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2168-2172, 2011.

[72] Z. Chen, F. Sohrabi and W. Yu: “Sparse activity detection for massive connectivity,” arXiv preprint arXiv:1801.05873, 2018.

[73] S. Haghighatshoar, P. Jung and G. Caire: “Improved scaling law for activity detection in massive MIMO systems,” arXiv preprint arXiv:1803.02288, 2018.

[74] G. Fettweis, H. Boche, et al: “The tactile internet,” ITU-T Technology Watch Report, August 2014.

[75] H. Lasi, P. Fettke, HG. Kemper, et al: “Industry 4.0”, Bus Inf Syst Eng, 6: 239, 2014. https://doi.org/10.1007/s12599-014-0334-4

[76] Ericsson Expert Analytics. [Online]. Available: https://www.ericsson.com/ourportfolio/products/expert-analytics.

[77] M. Agiwal, A. Roy and N. Saxena: “Next generation 5G wireless networks: A comprehensive survey,” IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617-1655, 2016.

[78] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang and L. Zhong: “Argos: Practical many-antenna base stations,” in Proc. of the 18th ACM annual international conference on Mobile computing and networking. ACM, pp. 53-64, 2012.

[79] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors and F. Tufvesson: “A exible 100-antenna testbed for massive MIMO,” in Proc. IEEE Globecom Workshops, pp. 287-293, 2014.

[80] S.-W. Jeon, S.-N. Hong, M. Ji, G. Caire and A. F. Molisch: “Wireless multihop device-to-device caching networks,” IEEE Transactions on Information Theory, vol. 63, no. 3, pp. 1662-1676, 2017.

[81] M. Ji, G. Caire and A. F. Molisch: “Fundamental limits of caching in wireless D2D networks,” IEEE Transactions on Information Theory, vol. 62, no. 2, pp. 849-869, 2016.

[82] M. Maddah-Ali and U. Niesen: “Fundamental limits of caching,” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856-2867, May 2014.

[83] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch and G. Caire: “Femtocaching: Wireless content delivery through distributed caching helpers,” IEEE Transactions on Information Theory, vol. 59, no. 12, pp. 8402-8413, 2013.

[84] M. Bayat, R. K. Mungara and G. Caire: “Achieving Spatial Scalability for Coded Caching over Wireless Networks,” ArXiv preprint: arXiv:1803.05702, 2018.

[85] D. Bethanabhotla, G. Caire and M. J. Neely: “Wiflix: Adaptive video streaming in massive MU-MIMO wireless networks,” IEEE Transactions on Wireless Communications, vol. 15, no. 6, pp. 4088--4103, 2016.

[86] Photonics21: Vision Paper “Europe’s age of light! How photonics will power growth and innovation”, Nov 2017, online.

[87] P.J. Winzer, D.T. Neilson: “From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade,” J. Lightw. Technol., 35(2017)5, Feb. 2017, pp. 1099-1115.

[88] S. Wolf et al.: “DAC-Less Amplifier-Less Generation and Transmission of QAM Signals Using Sub-Volt Silicon-Organic Hybrid Modulators“, J. Lightw. Technol., 33(2015)7, April 2015, pp. 1425-1432.

[89] ETSI: Mobile Edge Computing - A key technology towards 5G. ETSI White Paper No. 11, First Edition September 2015, http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf.

[90] ETSI: ETSI GS MEC 003: Mobile Edge Computing (MEC); Framework and Reference Architecture. V 1.1.1, March 2016, http://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.pdf.

[91] 3GPP: System Architecture for the 5G System. 3GPP-23501, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144.

[92] 3GPP: Procedures for the 5G System. 3GPP-23502, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3145.

[93] Zhang, Ke, Mao, Yuming, Leng, Supeng, Zhao, Quanxin, Li, Longjiang, Peng, Xin, Pan, Li, Maharjan, Sabita, Zhang, Yan: Energy-Efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks. IEEE Access, 20164. 1-1. 10.1109/ACCESS.2016.2597169.

[94] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang: A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications. IEEE Access, 2017, Digital Object Identifier 10.1109/ACCESS.2017.2685434.

[95] ETSI: ETSI GS MEC 002: Mobile Edge Computing (MEC); Technical Requirements. V 1.1.1, March 2016, http://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/01.01.01_60/gs_MEC002v010101p.pdf.

[96] ETSI: ETSI GS Mobile Edge Computing specifications. http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[97] Ericsson: Massive IoT. Ericsson White Paper: https://www.ericsson.com/ourportfolio/networks-solutions/massive-iot.

[98] Vodafone: Vodafone IoT Barometer 2017/18. http://vodafone.com/iot.

[99] The Boston Consulting Group: Making Autonomous Vehicles a Reality: Lessons from Boston and Beyond. October 2017, https://www.bcg.com/it-it/publications/2017/automotive-making-autonomous-vehicles-a-reality.aspx.

[100] J. Q. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming and Q. Yan: "Industrial Internet: A Survey on the Enabling Technologies, Applications, and Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1504-1526, 2017.

[101] S. Verma, Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama and N. Kato: "A Survey on Network Methodologies for Real-Time Analytics of Massive IoT Data and Open Research Issues," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1457-1477, third quarter 2017.

[102] W. Wang, Q. Wang and K. Sohraby: "Multimedia Sensing as a Service (MSaaS): Exploring Resource Saving Potentials of at Cloud-Edge IoT and Fogs," in *IEEE Internet of Things Journal*, Vol. 4, No. 2, pp. 487-495, April 2017.

[103] N. Hossein Motlagh, T. Taleb and O. Arouk: "Low-Altitude Unmanned Aerial Vehicles-Based Internet of Things Services: Comprehensive Survey and Future Perspectives," in *IEEE Internet of Things Journal*, Vol. 3, No. 6, pp. 899-922, December 2016.

[104] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos and M. Frodigh: "Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks," in *IEEE Wireless Communications*, Vol. 24, No. 2, pp. 82-89, April 2017.

[105] G. P. Fettweis: "The Tactile Internet: Applications and Challenges," in *IEEE Vehicular Technology Magazine*, Vol. 9, No. 1, pp. 64-70, March 2014.

[106] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner and Z. Li: "Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements," in *IEEE Network*, Vol. 32, No. 2, pp. 8-15, March-April 2018.

[107] Z. Li, H. Huang and S. Misra: "Compressed Sensing via Dictionary Learning and Approximate Message Passing for Multimedia Internet of Things," in *IEEE Internet of Things Journal*, Vol. 4, No. 2, pp. 505-512, April 2017.

[108] International Data Corporation: "IDC FutureScape: Worldwide Internet of Things 2018 Predictions." October 2017. https://www.idc.com/getdoc.jsp?containerId=US43193617.

[109] S. Abdelwahab, B. Hamdaoui, M. Guizani, T. Znati: "Network function virtualization in 5G." IEEE Communications Magazine, Vol. 54, No. 4, Apr. 2016.

[110] B. R. Al-Kaseem, H. S. Al-Raweshidyhamed: "SD-NFV as an Energy Efficient Approach for M2M Networks Using Cloud-Based 6LoWPAN Testbed." IEEE Internet of Things Journal, Vol. 4, No. 5, Oct. 2017.

[111] Z. Sheng et al.: "Lightweight Management of Resource-Constrained Sensor Devices in Internet of Things." IEEE Internet of Things Journal, Vol. 2, No. 5, Oct. 2015.

[112] Open Mobile Alliance: Reposit of LightweightM2M releases. http://openmobilealliance.org/release/LightweightM2M/.

[113] A. Taivalsaari, T. Mikkonen: "A roadmap to the programmable world: Software challenges in the IoT era." IEEE Software, Vol. 34, No. 1, 2017.

[114] K. Morris: "Infrastructure as Code: Managing Servers in the Cloud". O'Reilly & Associates Incorporated, 2016.

[115] H. Yetgin, K. T. K. Cheung, M. El-Hajjar, L. H. Hanzo: "A survey of network lifetime maximization techniques in wireless sensor networks." IEEE Communications Surveys & Tutorials, 19(2), 828-854.

[116] Gartner: IoT Gartner estimates. https://www.gartner.com/newsroom/id/3598917.

[117] N. Kshetri: “Can Blockchain Strengthen the Internet of Things?” in IT Professional, vol. 19, no. 4, pp. 68-72, 2017.

[118] K. Christidis and M. Devetsikiotis: “Blockchains and Smart Contracts for the Internet of Things.” in IEEE Access, vol. 4, pp. 2292-2303, 2016.

[119] K. A. Alam, R. Ahmad and K. Ko: “Enabling Far-Edge Analytics: Performance Profiling of Frequent Pattern Mining Algorithms.” in IEEE Access, vol. 5, pp. 8236-8249, 2017.

[120] T. Llewellynn et al.: “BONSEYES: Platform for Open Development of Systems of Artificial Intelligence.” in Proceedings of the Computing Frontiers Conference (CF'17).

[121] T. Hale: How Much Data Does the World Generate Every Minute? 26 July 2017, http://www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/.

[122] D. Hernandez: How much Data will The Internet of Things (IoT) Generate by 2020? October 2017, https://www.versatek.com/blog/how-much-data-will-the-internet-of-things-iot-generate-by-2020/.

[123] B. Adams and Karen Judd: Data is the new gold – development players mine a new seam. Global Policy Watch, https://www.globalpolicywatch.org/blog/2017/11/27/data-is-the-new-gold/.

[124] The Economist: The world’s most valuable resource is no longer oil, but data. May 6th 2017, https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

[125] World Economic Forum: “Unlocking the Economic Value of Personal Data. Balancing Growth and Protection.” Brussels, 8 October 2012. Available at http://www3.weforum.org/docs/WEF_IT_UnlockingValueData_BalancingGrowthProtection_SessionSummary.pdf.

[126] TNS: TNS Opinion & Social, “Special Eurobarometer 431 Data Protection.” June 2015. Available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf.

[127] European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

[128] OECD: “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value”, OECD Digital Economy Papers, No. 220, OECD Publishing. April 2013. Available at http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en.

[129] Data Transparency Lab: http://datatransparencylab.org/.

[130] WIbson: Don't give out your data for free. Make a profit. https://www.wibson.org/.

[131] PwC: 2017. IAB internet advertising revenue report 2016 full year results. Technical Report. https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf.

[132] IHS Technology: Paving the way: how online advertising enables the digital economy of the future. Technical Report. November 2015, https://www.iabeurope.eu/files/9614/4844/3542/IAB_IHS_Euro_Ad_Macro_FINALpdf.pdf.

[133] Ernst and Young: What is an untrustworthy supply chain costing the US digital advertising industry. IAB US benchmarking study, November 2015, https://www.iab.com/wp-content/uploads/2015/11/IAB_EY_Report.pdf.

[134] TorrentFreak: Pirate Sites Generate 227 Million in Ad Revenue a Year. February 2014, https://torrentfreak.com/torrent-sites-6-million-ad-revenue-140219/.

[135] The Guardian: Head of Google Europe apologises over ads on extremist content. March 2017,https://www.theguardian.com/technology/2017/mar/20/google-ads-extremist-content-matt-brittin.

[136] GENI: Global Environment for Network Innovation, Key GENI concepts, http://groups.geni.net/geni/wiki/GENIConcepts.

[137] IETF: RFC 5212, Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN).

[138] NGMN Alliance: ”Description of Network Slicing Concept”, Version 1.0, January 2016.

[139] NGMN Alliance: ”5G Security Recommendations Package #2: Network Slicing” Version 1.0, April 2016.

[140] 3GPP::”Study on the security aspects of the next generation system”. TR 33.899, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3045.

[141] 3GPP: ”Security architecture and procedures for 5G System”. TS 33.501, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169.

[142] 5G PPP: Phase 1 Security Landscape, https://5g-ppp.eu/white-papers/.

[143] Scott-Hayward, S., Natarajan, S., & Sezer, S.: (2016). A Survey of Security in Software Defined Networks, IEEE Communications Surveys and Tutorials.

[144] Olli Mämmelä, Jouni Hiltunen, Jani Suomalainen, Kimmo Ahola, Petteri Mannersalo, Janne Vehkaperä: ”Towards Micro-Segmentation in 5G Network Security”. EuCNC 2016.

[145] ATOS: Journey2020: 2017, https://atos.net/content/mini-sites/journey-2020/?utm_source=ascent.atos.net/journey-2020/&utm_medium=301.

[146] Ericsson: 5G Business Potential. 2017, https://www.ericsson.com/assets/local/news-and-events/events/2017/mwcs-2017/topic-3_ericsson_5g_business_potential.pdf.

[147] Huawei: Next generation assurance. November 2017, https://www.sdxcentral.com/articles/featured/huawei-next-generation-assurance/2017/11/.

[148] ABIResearch: Mobile World Congress Highlights the Challenges Facing Mobile Service Providers. https://www.abiresearch.com/pages/future-mobile-industry/.

[149] OPENET: Digital Evolution: Cut BSS/ OSS Costs, Reduce Delivery Timescales. 2018, https://www.openet.com/doc-redirect/index_form2.php?docid=724.

[150] Neuralink: Ultra-high bandwidth brain-machine interfaces to connect humans and computers. 2018, https://www.neuralink.com/.

[151] statista – The Statistics Portal: Global digital population as of January 2018 (in millions). 2018, https://www.statista.com/statistics/617136/digital-population-worldwide/.

[152] ETSI: ETSI NFV ”Zero touch network and Service Management” (ETSI ZSM ISG). December 2017, http://www.etsi.org/technologies-clusters/technologies/zero-touch-network-service-management.

[153] TMForum whitepaper: ”OSS of the Future”. 2017, https://www.tmforum.org/resources/whitepapers/oss-of-the-future/.

[154] Ericsson: Zero-touch networks with cloud-optimized Network Applications. Ericsson whitepaper, 2017, https://archive.ericsson.net/service/internet/picov/get?DocNo=4/28701-FGB1010325&Lang=EN&HighestFree=Y.

[155] ETSI: ETSI ISG Experiential Network Intelligence Whitepaper. October 2017, http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp22_ENI_FINAL.pdf.

[156] YuLing Chen, Alon Bernstein: Bridging the Gap Between ETSI-NFV and Cloud Native Architecture. Cisco Systems Inc., October 2017, https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwiss4DOr63aAhUojK0KHQYuAsEQFgguMAE&url=https%3A%2F%2Fwww.nctatechnicalpapers.com%2FPaper%2F2017%2F2017-bridging-the-gap-between-etsi-nfv-and-cloud-native-architecture%2Fdownload&usg=AOvVaw3f517ofWFJMSssNLiEALDX.

[157] ETSI: Network Functions Virtualisation (NFV); Architecture; Report on the Enhancements of the NFV architecture towards ”Cloud-native” and ”PaaS” Release 3. Draft version. January 2018, https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjLosm2sa3aAhVNmK0KHZGzDIgQFggnMAA&url=https%3A%2F%2Fdocbox.etsi.org%2FISG%2FNFV%2FOpen%2FDrafts%2FIFA029_Arch_enhancement_for_Cloud-native_%2526_PaaS%2FNFV-IFA029v001.docx&usg=AOvVaw3eJ1RrMe6CbOHH7tEQFLFD.

[158] Analysis Mason: Flexible, extensible, adaptable: towards a universal template for VNFs onboarding and lifecycle management. Analysis Mason Limited, 2017, http://www.analysysmason.com/Research/Content/Reports/flexible-extensible-adaptable-towards-a-universal-template-for-vnf-onboarding-and-lifecycle-management-white-paper/.

[159] D. D. Clark, C. Partridge, J. C. Ramming, J. T. Wroclawski: “A knowledge plane for the internet”. ACM SIGCOMM 2003 conference.

[160] Marco Ajmone Marsan, Giuseppe Bianchi, Nicola Blefari Melazzi: Living and Fluid Networks: the way ahead? Computer Communications, to be published.

[161] I. Akyildiz and J. Jornet: “The Internet of nano-things”. IEEE Wireless Communications, vol. 17, no. 6, pp. 58–63, 2010.

[162] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz: “Using any Surface to Realize a New Paradigm for Wireless Communications”. Communications of the ACM (to appear), 2018.

[163] G. Piro, G. Boggia, and L. Grieco: “On the design of an energy-harvesting protocol stack for Body Area Nano-NETworks”. Nano Communication Networks, vol. 6, no. 2, pp. 74–84, 2015.

[164] C. Liaskos et al: “Design and Development of Software Defined Metamaterials for Nanonetworks”. EEE Circuits and Systems Magazine, vol. 15, no. 4, pp. 12–25, 2015.

[165] Edgar A. Aguilar, Ravishankar Ramanathan, Johannes Kofler, and Marcin Pawłowski: Physical Review A 94, 022305, 2016.

[166] Manuel Erhard, Robert Fickler, Mario Krenn, and Anton Zeilinger: Light: Science & Applications, 7, 17146, 2018.

[167] Massimiliano Smania, Ashraf M Elhassan, Armin Tavakoli, and Mohamed Bourennane: Nature Photonic Journal Quantum Information, 2, 16010, 2016.

[168] Charles H.Bennett, GillesBrassard: Theoretical Computer Science, 560,7, 2014.

[169] Wei Zhang, Dong-Sheng Ding, Yu-Bo Sheng, Lan Zhou, Bao-Sen Shi, and Guang-Can Guo: Physical. Review Letters, 118, 220501, 2017.

[170] Donaldson, R. J. et al.: Physical Review A, 93, 012329, 2016.

[171] Hiroyuki Shibata, Toshimori Honjo, and Kaoru Shimizu: Optics Letters 39, 17, 2014.

[172] Sheng-Kai Liao et al.: Nature Photonics, 11, 509, 2017.

[173] T. Shinada, et al.: Deterministic doping to silicon and diamond materials for quantum processing. In: Nanotechnology (IEEE-NANO), 2016 IEEE 16th International Conference on. IEEE, 2016. p. 888-890.

[174] Tiecke, T. G., Thompson, J. D., de Leon, N. P., Liu, L. R., Vuletić, V., & Lukin, M. D.: (2014). Nanophotonic quantum phase switch with a single atom. Nature, 508(7495), 241.

[175] Azuma, K., Tamaki, K., & Lo, H. K.: (2015). All-photonic quantum repeaters. Nature communications, 6, 6787.

[176] Marandi, A., Wang, Z., Takata, K., Byer, R. L., & Yamamoto, Y.: (2014). Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nature Photonics*, *8*(12), 937.

[177] Nick Szabo: “Smart Contracts: Building Blocks for Digital Markets”. (1996).

[178] Satoshi Nakamoto: “Bitcoin: A peer-to-peer electronic cash system”. (2008), https://bitcoin.org/bitcoin.pdf.

[179] Multichain: Three (non-pointless) permissioned blockchains in production. 2017, Quite taken from https://www.multichain.com/blog/2017/11/three-non-pointless-blockchains-production/.

[180] L. S. Sankar, M. Sindhu, M. Sethumadhavan: “Survey of consensus protocols on blockchain applications.” 4th IEEE Int. Conf. on Advanced Computing and Communication Systems (ICACCS), 2017.

[181] Certificate Transparency: General Transparency. https://www.certificate-transparency.org/general-transparency.

List of Contributors

Rui L. Aguiar	Instituto de Telecomunicações, Univ. de Aveiro
Ian F. Akyldiz	Georgia Institute of Technology
Achim Autenrieth	ADVA Optical Networking
Arturo Azcorra	University Carlos III of Madrid, Imdea Networks
Giuseppe Bianchi	University of Rome, Tor Vergata
Sebastién Bigo	Nokia Bell Labs
Nicola Blefari-Melazzi	CNIT
Harald Bock	Coriant
Andre Bourdoux	*IMEC*
Giuseppe Caire	TU Berlin
Gino Carrozzo	Nextworks
Fabio Cavaliere	Ericsson
Sonia Cazalens	CNES
Symeon Chatzinotas	Univ. of Luxembourg
Nicolas Chuberre	Thales Alenia Space
Nicola Ciulli	Nextworks
Tomaso de Cola	DLR
Angelo Corsaro	ADLINK Technology Inc
Ángel Cuevas	University Carlos III of Madrid
Rubén Cuevas	University Carlos III of Madrid
Filippo Cugini	Scuoala Superiore Sant'Ana / CNIT
Franco Davoli	U. Genoa / CNIT
Herve Debar	telecom SudParis
Emmanuel Dotaro	Thales
Michael Eiselt	ADVA Optical Networking
Joerg-Peter Elbers	ADVA Optical Networking
Barry Evans	Univ. of Surrey
Antonio Fernández-Anta	Imdea Networks
Gerhard Fettweis	TU Dresden
Miltiadis Filippou	Intel Germany
Ronald Freund	Fraunhofer HHI
Stein Gjessing	University of Oslo Norway
Jose Enrique Gonzalez	Atos
Roberto Gonzalez	NEC
Helmut Grießer	ADVA Optical Networking
Maria Guta	ESA
Harald Haas	University of Edinburgh
Artur Hecker	Huawei Technologies
Elisa Jimeno	Atos
Josep M. Jornet	State University of New York at Buffalo
Raymond Knopp	EURECOM
Christos Liaskos	Foundation for Research and Technology – Hellas
Diego Lopez	Telefonica
Antonio Manzalini	TIM
Marco Ajmone Marsan	IMDEA
Josep Martrat	Atos
Jean Marc Merolla	University of Franche-Comte
Enzo Mingozzi	University of Pisa
Werner Mohr	Nokia
Michael Montag	Nokia

Giacomo Morabito	University of Catania
Markus Mueck	Intel Germany
Raul Muñoz	CTTC
Antonio de la Oliva	University Carlos III of Madrid
Colin Perkins	University of Glasgow
Enrico Prati	CNR
Aurora Ramos	Atos
David Ros	Simula Research Laboratory, Norway
Marco Ruffini,	Trinity College Dublin
Dario Sabella	Intel Germany
Agnes Salvatori	Airbus DS
Colja Schubert	Fraunhofer Heinrich-Hertz-Institute
Detlef Schulz	SES
Egon Schulz	Huawei
Pablo Serrano	University Carlos III of Madrid
Dimitra Simeonidou	University of Bristol
Alexandros Stavdas	University of Peloponnese
Dirk Trossen	InterDigital Europe, Ltd.
Carlo Vallati	University of Pisa
Miguel Ángel Vázquez	CTTC
Luís Velasco	Universitat Politecnica de Catalunya
Rainer Wansch	Fraunhofer Institute
Simon Watts	Avanti Communications
Elisabeth Weller	Eutelsat
Michael Welzl	University of Oslo Norway
Wen Xu	Huawei

About NetWorld2020

NetWorld2020 is the European Technology Platform for communications networks and services. Communications networks enable interaction between users of various types of equipment, either mobile or fixed. They are the foundation of the Internet. The NetWorld2020 European Technology Platform gathers 1000 players of the communications networks sector: industry leaders, innovative SMEs, and leading academic institutions. NetWorld2020 has as mission developing position papers on technological, research-oriented and societal issues, in order to strengthen Europe's leadership in networking technology and services so that it best serves Europe's citizens and the European economy.

5G Infrastructure Association (5G-IA)

The 5G Public Private Partnership (5G PPP) is the 5G collaborative research program that is organised as part of the European Commission's Horizon 2020 program – The European Union Program for Research and Innovation. It is aimed at fostering industry-driven research, monitored by business-related, technological performance and societal KPIs. The 5G-PPP will deliver solutions, architectures, technologies and standards for ubiquitous next-generation communication infrastructure over the coming decade.

In the 5G PPP, the 5G Infrastructure Association (5G-IA) represents the private side and the European Commission the public side. The 5G-IA is committed to the advancement of 5G in Europe and to building global consensus on 5G. To this aim, the Association brings together a global industry community of telecoms & digital actors, such as operators, manufacturers, research institutes, universities, verticals and SMEs. The 5G-IA carries out a wide range of activities in strategic areas including standardisation, frequency spectrum, R&D projects, technology skills, collaboration with key vertical industry sectors, notably for the development of trials, and international cooperation.