|  |  |
| --- | --- |
| D:\Mohr\Arbeitsverzeichnis\Europäische Kommission\Mobile Communications & Technology Platform\EU Technology Platform\Logos NetWorld2020\NetWorld2020 large logos\NW 2020 LBB\NW2020.jpg | logo-5G-positive_ia-text |
|  | September 2018 |

**Strategic Research and Innovation Agenda 2021-27**

**European Technology Platform NetWorld2020**

**“Smart Networks in the context of NGI”**

**2018**

**Table of Contents**

## 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:

1. How networks developed and evolved (left side of the figure);
2. Past and current main trends behind such progresses (right side of the figure);
3. 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[[1]](#footnote-1), and also because a softwarised network will evolve more rapidly and incrementally with new software releases, rather than with major generational leaps.



1. 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 [162] – 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* [163], 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.

However, the high softwarization that ICT technologies will experience has a counterpart in terms of energy consumption. In this regard, data centres can increase their energy consumption from the actual 2% of worldwide electricity, to up the 8% by 2030.[192] Moreover, also autonomous cognitive strategies will further contribute to the energy budget, since AI based solutions are energy hungry. In fact, training a single AI is equivalent to 284 tonnes of carbon dioxide equivalent, i.e., five times the lifetime emissions of an average car [193].

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.

## The Physical Stratum: Communication and Computing Resources

### **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 [164]. 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 [165].

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 [166] and self-monitoring and adapting industrial materials [167].

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.

### **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.

### **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 [168] 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 [169] 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 [170], quantum public-key cryptography [171], quantum secure direct communication [172], quantum digital signatures [173].

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 [174] and in free-space links [175].

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 [176], 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 [177]. 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 [178]. 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 [179].

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.

### **AI/ML for the Physical Layer**

The ever increasing exigence of higher throughput, lower latency and extremely high density of connections is taking the Physical Layer to approach its ultimate performance limits. This implies dealing with channel and transceiver impairments that were usually discarded in the interest of understandable and tractable models. AI and ML can be very helpful tools, also for the design of new transmission and reception techniques, when we cannot rely on the classical models and optimization approaches.

In this regard, there are essentially two kinds of problems where AI/ML may help:

* Offline design of elements or algorithms that may be possible to generalize to a wide variety of scenarios and used afterwards as building blocks of the transmit-receive chain.
* Online optimization of some elements or even the whole transmit-receive approach, that may be able to adapt to the changing (and difficult to model) characteristics of the channel and interference.

Obviously the requirements for both are quite different. While offline optimization requires - but also withstands - training with a large amount of diverse-enough data, the convergence time and complexity are key requirements for the online optimization in real time. These may be achieved by reinforcement learning (RL), a method that interacts with a dynamic environment by producing a series of actions and receives rewards according to the performance of such action with respect to the environment situation.

An example of the offline approach is the use of modern evolutionary computation techniques as solvers of several complex optimization problems. Evolutionary computation [185] is a subfield of AI and soft computing, which is composed of global optimization techniques based on mimicking biological evolution. In [186] it is used to propose novel constellation designs for non-coherent massive MIMO. A different approach is used in [187] where the learned constellation with unsupervised ML mitigates nonlinear effects of the optical fibre channel.

End-to-end learning aims to learn transmitter and receiver implementations optimized for a specific performance metric and channel model. It was first presented in [188]. The idea is to interpret the whole communication chain as an “autoencoder”, an unsupervised learning technique. Here the communications system design is conceived as an end-to-end reconstruction task that seeks to jointly optimize transmitter and receiver components in a single process.

ML may also allow us to understand and model channels for new applications, such as molecular communications. In [189] the diffusion-based molecular MIMO channel is modeled with an artificial neural network (ANN).

Learning from the physical layer may also bring us a better knowledge and use of the channel state information, which may be helpful, for example, to locate people or detect their movement, gesture or activity.

Immediate problems to solve are the need for algorithms that have speed and complexity compatible with the real time data transmission. Also, the availability of large and representative data sets for training and benchmarks are less developed than in other current applications of AI/ML.

### **DSL**

The ideal of offering ubiquitous broadband for all requires a high capillarity of high speed wireline access, even to facilitate wireless connections. Today DSL is an alternative to the deployment of optical fibre. Will we be able to count on the already deployed copper wire infrastructure to support 5G and future networks? The possibility of reaching data rates in DSL comparable to fibre would be instrumental to providing universal broadband connectivity in a cost-effective and quick way. However, current DSL technologies seem to have achieved their maximum potentiality with downstream transmission rates of up to 100 Megabits per second at a range of 500 meters, and more than 1 Gigabit per second at shorter distances.

In [190] the channel properties of a 200 GHz signal transmitted through a waveguide structure that is designed to approximately emulate the type of paired phone cable typically used for DSL transmissions are investigated. They find that aggregate data rates on the order of Terabits per second are feasible over short distances.

Transmitting Terabits per second through a copper pair’s sub-millimetre waveguide modes is an emerging technique that would allow us to continue leveraging the copper infrastructure for wireline access and also contribute to accelerate the deployment of small cells with reduced infrastructure cost. Future research needs to bring this idea from the lab to practical applications, solving issues such as how to extend the system to a larger range by reducing the amount of energy lost due to resistance.

### **The Mobile Network**

The new Air Mobility Network, are the new network serving all the “things” between the ground and 20KM height. It provides control and communication services for the drone, urban air mobility (future urban transportation systems that move people by air), balloon, aircraft, etc.

Besides traditional Manned Vehicles, the number of flying things has (and will continue to) skyrocketed with the increasing adoption of Unmanned Aerial Vehicles (UAVs), , and it is envisioned that this space (20KM above the ground) will be filled with flying things, which required to be connected in multiple scenarios. The flying things can be categorized into multiple categories, using different classifications, according to the speed of the flying thing, the altitude where they fly, the inter-distance expected between them, or the type of requested services and communications, just to name some examples. However, here, we can categorize the flying things into two main classes: i) aircrafts (airplanes), and ii) UAVs.

Aircrafts have existed for multiple decades now. Airplanes fly at very high altitude being at the extreme limit of considered elevation, with high speed (around 900Km/h), and also at the highest end of the considered speeds of flying things. UAVs, resides at the other edge of the spectrum, when it comes to flying altitude and speeds, but with wider variation, when it comes to speeds and altitude, since UAVs vary widely in their speed and altitude capabilities. Note that popular expectations on urban mobility suggest private helicopters, or flying taxis/dispatching units. Nevertheless, these flying units will be essentially self-flying, and will be on the flight altitudes associated to these UAVs, so from the communications challenges perspective, they can be considered in the same class as “UAVs”.

The two classes differ much in their properties, when it comes to speed and altitude, but they still have common properties when it comes to their required communication capabilities. In both classes, the control service of each flying “thing” is critical and mandatory, while the data communication maybe more flexible, depending on the service requirements (Surveillance? Transportation? Logistics? Robotics?). Future communications and networking research should be able to provide the required reliable communications for both categories (control and data), in the most efficient way, while considering cost-effectiveness as a design criterion/requirement.

A) The control service (which can also be referred to as control and nonpayload communication - CNPC - services) of airplanes is critical and mandatory, which defines the basic rules, forms the virtual paths in the air, avoids collision, and reserves passage for airplane/UAVs take-off and landing. The control service may be based on satellite (low earth orbit in most cases), ground base stations, and even among flying “things”, with dynamic multi or single connection, to guarantee the “control” with very high reliability. Passive radio sensing detection is also necessary for the “control”, since it is very dangerous to rely on the positioning reported by the flying “things” themselves in congested airspaces.

B) Data communications (which are used for non-control necessities, including entertainment for passengers ) maybe rather flexible, depending on the service requirements, and it could be supported by somewhat traditional communication (for example, connect with a ground base station), or probabilistic communication (carry out data communication within a defined period of time, while each data transmission complies with a certain success probability) among the flying vehicles.

**Airplanes**: In recent years, we started seeing Internet services onboard of airplanes, and even live TV, with minimal quality. Future networking architectures are required to provide improved reliable broadband communications capabilities to aircrafts. The first type of communication services requires high-reliability, very low-latency, and mostly low data rates. On the other hand, the reliability of the second type may be a bit lower, but still depends on the type of service. We may have variable communication requirements, even ultra-low latency, if for instance gaming, video-conferencing, etc. services are to be offered onboard. However, it is important to differentiate between the criticality of both types of services. In other words, control services have to be guaranteed, and provide high-reliability, to secure the safety of the aircraft itself, while data communication may need the same requirements; however, losing such communications would not jeopardize the safety of the aircrafts.

Communications to aircrafts, flying at very high altitude with ultra-high speed, face three major challenges, to meet their requirements: i) continuous (reliable) communications, even at high altitude; ii) low latency; and iii) high bitrates.

For Air mobility new network, the wireless network architecture could be completely different from the “flat” ground cellular network. The network could be based in clusters of three- dimensional spheres, without interference between clusters if dedicated spectrum could be allocated. How to share spectrum between all these clusters, ground and satellite is a challenge.

One main direction will be how to design an integrated ecosystem, encompassing the multiple networking paradigms, including mobile cellular networks, satellite systems, in addition to even the very high frequency (VHF) spectrum. The future architecture will need to propose interference management and mitigation schemes, between the different systems. For this requirement, channel modelling of cellular networks, towards such high altitude users, needs to be achieved. How cellular networking can provide such seamless continuous connectivity will be another challenge, required to be answered by the future networking generation.

Finally, another dimension for communications will be ad hoc communications between multiple aircrafts, again for safety reasons, to avoid collisions between aircrafts. Ad hoc communications may require completely different networking protocol and architecture.

**Unmanned Aerial Vehicles**: UAVs are becoming critical these days, and is foreseen to be an essential part of our future ecosystem, especially in the era of IoT and smart city. A study by PricewaterhouseCoopers (PwC) estimates the business market of UAVs to be valued at over $127 billion.

This huge market has opened new challenges for the communications society trying to answer two main questions:

1. What can networks do for UAVs?
2. How can UAVs benefit the networks in general (mobile networks specifically)?

Future networks need to answer the two above questions.

The first challenge is to guarantee reliable communications for all the flying “Things”. Most UAVs require continuous reliable communications, for the safe travel (flying). The challenge is how to provide such communications, within the random, high variable, high mobile environment of UAVs. It is clear that we will be entering a new domain, different from the plane cellular on the ground, but here, maybe with a huge number of things coexisting in a close environment.

From this new perspective, new research challenges arise:

* How to share the spectrum between different networks, and avoid interference to maximize resource utilization and increase quality and reliability?
* What would be the real-life propagation models in the new 3D environment?
* Study of the effect of low/high mobility of drones on the UAVs communication and how to design communication networks to provide reliable communications in such challenging environment
* Energy efficiency will be another challenge. UAVs are power-limited; hence, minimizing of energy consumption is a priority, to allow extended periods of flying, hence service, specially on swarm scenarios.

The second challenge is to answer the question of how the UAVs can be integrated in the overall networking ecosystem. The use of UAVs within the mobile cellular ecosystem has attracted lots of attention from the academic and industrial entities: The use of UAVs as relays (in particular as small cells), to provide cellular coverage in certain situation has been seen as a high beneficial solution, in the future networks. UAVs can provide networking in emergency situation; to remote areas; temporarily during events with expected large number of devices, etc.

The concept of UAVs as an integral part of the mobile cellular network opens a plethora of challenges for the future networks:

* What frequency should UAVs use? For fronthaul towards the UAVs? For cellular service provisioning to end-users?
* How to coordinate between UAVs, legacy BSs, small cells (which are randomly setup) etc.? Dynamic interference management in such scenario will be highly challenging
* Management of attached, detachment and re-attachment of UAVs will also be challenging, especially with mobility (if allowed in case of UAVs as small cells or relays).
* Energy efficiency will definitely be a high priority, due to the limited power capacity of UAVs.

Similar to how 5G networks revolutionized how communications were handled, so will the next generations further push the envelope beyond just performance enhancements solely based on radio and core architecture evolutions. Whole system rethinking allied to the integration of new enablers (e.g., as SDN, NFV, SBA, Cloud-native were thought for 5G), will transform and create new opportunities for Air Mobility New Network. As the amount of devices with link capabilities increases, due to the decrease in electronics size and price, more connected flying entities will occupy our skies and space. This means that a greater amount of simultaneous connectivity opportunities will present themselves to each individual connected flying thing. By coupling them with multiple network interfaces, air multihoming scenarios will surface, allowing such devices to pick the best connectivity path according to the specific needs of the application (or network control) services traversing those air link channels. This mandates a closer integration with Machine Learning techniques in order to optimize link connectivity choice, as communications will not only scale laterally but also upwards, and become 3D. This will also give birth to new kinds of delay-tolerant / relaying operations and protocols, where the dynamic connectivity opportunities need to be anticipated by learning algorithms, in order to minimize losses or misguided overutilization of air links that do not successfully deliver the information towards the intended target (as needed for reliable communications mandated by verticals or response teams in serious situations). As autonomy is a serious matter in unmanned air vehicles, the added complexity of this type of per-flow optimization (where each flying node can have different applicational flows going through different links in the sky and/or ground) demands greater computational power, new offloading mechanisms towards the edge of the ground network (or flying nodes with greater flight autonomy and even remote battery charge sharing between nodes), will allow such processing to become fully distributed. Obviously, the operational capability for the abovementioned scenarios emphasizes the need for thorough secure operations, deployed both at the control and data paths.

## Protocols, Algorithms and Data

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

### **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. The intelligence will progressively migrate towards the edge of the network, requiring research that goes beyond traditional approaches and moves towards novel federated learning solutions, to enable multiple edge devices to build a common, robust ML model without sharing data; multi-task learning, to learn from multiple related tasks, simultaneously; deep transfer learning to solve a particular task using a pre-trained model on a different task; multi-agent reinforcement learning, to distributively learn decision making policies, facing the game theoretic dynamics that appear due to the non-stationarity of the environment.

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.

Nevertheless, the massive adoption of AI tools will exacerbate the problem of energy consumption of the ICT infrastructure. As introduced before, training a single AI is equivalent to 284 tonnes of carbon dioxide equivalent, i.e., five times the lifetime emissions of an average car[193]. Therefore, it will be crucial to devise energy efficient architectures and computation algorithms in order to have energetically sustainable communication and computing paradigms for future mobile networks. For doing so, networks may exploit distributed energy generation and energy harvesting/ storage hardware which have to be included in the design of the network solutions in order to efficiently operate.

***Changing Network Design***

From the today’s perspective AI/ML 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 significant increase in the amount of machine-to-machine communications and of correspondent data to be processed 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 understand 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 mobility, 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.

Mobile networks can be used as additional sensing platforms. Indeed, the extreme pervasiveness of the mobile telecommunication sector within the urban population together with its ubiquitous coverage may be exploited to monitor large metropolitan areas. The services that can be generated are manifold (housing, transportation, energy systems, education and health care). For instance, the detection of critical anomalies caused by unexpected crowd gathering in metropolitan areas (e.g., concerts, football matches) can be achieved through the collection of information that the different network elements (e.g., base stations, mobile terminals) are exchanging over time through the control channel [194].

***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).

### **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 heterogeneityalong several dimensions: access technology, identification/naming/addressing scheme, traffic patterns, deployment extension, device capabilities, etc. Such heterogeneity calls for a network which is highly flexiblein 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 slicesthat 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.

**Tactile and Industrial-Tactile IoT:** Research priorities in this area focus on real-time sensing/actuating using haptic interaction with visual feedback, and the integration of IoT systems supporting not only audio-visual interactions but also involving robotic systems to be controlled with a real-time response.

**Emerging industrial IoT applications,**

It is expected that new emerging applications based on Tactile IoT/IIoT, will be developed in the near future, see e.g., [191]. Examples of Emerging applications using tactile IoT/IIoT, described in [191] are:

* Holographic media applications: involve not only the local rendering of holograms but networking aspects, specifically the ability to transmit and stream holographic data from remote sites,
* Multi-Sense Networks: include emerging applications that involve not only optical (video, holograms) and acoustic (audio) senses, but as well smell and taste senses.
* Time Engineered Applications: uses a communication system that can coordinate between different sources of information such that all the parties involved have synchronized view of the application.
* Critical Infrastructure support applications: support of critical infrastructures that refer to those essential assets that are considered vital to the continued smooth functioning of the society as an integrated entity.

Emerging industrial IoT applications, Tactile Internet and autonomous/robotic systems solutions will require far faster reactivity at the edges of the networks as it becomes increasingly inefficient to extract insights from the cloud with growing numbers of IoT devices. Research priorities include the development of new open integrated horizontal platforms for mobile edge computing and edge analytics solutions.

**Extreme Automation and Real-Time Zero-Touch - Service Orchestration:** In a few years, social machines, smart contracts and other types of more advanced interaction will be a reality. Some machines will be indistinguishable from people from the perspective of business processes and interactions, with higher capacity of decisions, the ability to orchestrate common actions, make requests, etc. Future networks will therefore require higher demands on real-time network service management and a high degree of automation. The challenge is how, without breaking overall end user experience. Several topics are to be addressed, including:

* Enhanced policy management including huge data analytics
* Artificial intelligence driven orchestration
* Cloud-native management applied to Network Function Virtualization orchestration

**Service Injection Loop:** The creation of services should be reinvented for the new digital area. Architectural micro services provide modular, distributed software components that can be deployed in any environment with a standardized infrastructure, allowing distributed applications to be installed on a cloud infrastructure while maintaining maximum flexibility. Research is required into new ways of describing the entire platform in metamodels. This innovation should be driven not only in the network transformation but also in the creation of a catalogue of new services. These services interoperate with platform capabilities and can automatically adapt to the needs of the user, and will involve new business models in pay-for-what-you-use services.

**Digital Twins for IoT:** Digital twins are virtual representations of material assets. While current solutions for IoT platforms have mainly been for the representation of physical objects, features such as simulation, manipulation and optimization are missing. Digital twins can be used, for example, to trigger and simulate threat scenarios, and help to optimize the security strategy to handle such scenarios if they occurred in the real world. Research is needed to address the Integration of IoTsdigital twins Into IoT/5G industrial platforms.

### **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 range of potential use cases for future networks is huge: capacity sharing in distributed networks, spectrum, resource and infrastructure management and sharing, energy trading, resource and service federation in virtualized networks, etc. Facilitating all these scenarios Blockchain technologies would also further contribute to cut down the operational costs of running the network. Additional interdisciplinary interactions with distributed ML/AI solutions and the corresponding game theoretical effects are also to be explored to fulfil the view of Next generation self-organized networks.

***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 [180] in the 90ies), Blockchains – as we know them today – emerged only in 2008, as the technical foundation and enabler of Bitcoin [181], 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[[2]](#footnote-2), and has to date mainly used 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”* [182]. 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:

1. 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;
2. a way to resolve differences among different replicas of the blocks – this is accomplished by a suitable consensus protocol, and
3. 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 [183], 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 [184] 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.

### **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?

## Applications

### **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.

### **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.

### **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.

A recent trend, associated with aspects of tracing, is coming from the need to have simple ways of trace contacts. This recent pressure, associated with the current covid19 crisis, will most probably remain for the future, as a tool for managing future societal crisis, and hopefully without depending on user interaction, but instead on network capabilities. This is a completely different set of requirements to the network (or end devices – both Android and iOS are addressing this challenge) that need to be fulfilled while assuring adequate citizen protection in privacy and security aspects.

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.

## References

1. 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>
2. 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.
3. Azuma, K., Tamaki, K., & Lo, H. K.: (2015). All-photonic quantum repeaters. Nature communications, 6, 6787.

[185] A. N. Sloss and S. Gustafson, “Evolutionary Algorithms Review,” 2019.

[186] M.J. Lopez-Morales et al., “Artificial Intelligence-aided Constellation Design for Multi-User MDPSK Non-Coherent Massive SIMO”, submitted to IEEE Global Communications Conference (Globecom), Taipei, Dec. 2020.

[187] R. T. Jones, et al., “Deep learning of geometric constellation shaping including fiber nonlinearities,” in 2018 European Conference on Optical Communication (ECOC), Sep. 2018.

[188] T. O’Shea and J. Hoydis, “An Introduction to Deep Learning for the Physical Layer,” IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, Dec 2017.

[189] C. Lee, H. B. Yilmaz, C. Chae, N. Farsad and A. Goldsmith, "Machine learning based channel modeling for molecular MIMO communications," 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, 2017.

[190] R. Shrestha e al. “A wire waveguide channel for terabit-per-second links“, Applied Physics Letters, vol. 116, no. 13, Mar. 2020.

[191] "Network 2030: A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond", FG-NET-2030 ITU-T, May 2019

[192] “Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper

[193] Energy and Policy Considerations for Deep Learning in NLP. Emma Strubell, Ananya Ganesh and Andrew McCallum. Annual Meeting of the Association for Computational Linguistics (ACL short). Florence, Italy. July 2019.

[194] H. D. Trinh, E. Zeydan, L. Giupponi, P. Dini, Detecting Mobile Traffic Anomalies ThroughPhysical Control Channel Fingerprinting:A Deep Semi-Supervised Approach , IEEE Access, Vol. 7, pp. 152187-152201, October 2019.

|  |  |
| --- | --- |
| 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 |
| Daniel Corujo | Instituto de Telecomunicações, Univ. de Aveiro |
| 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 |
| Paolo Dini | CTTC |
| Herve Debar  | telecom SudParis |
| Emmanuel Dotaro | Thales |
| Omar Elloumi | Nokia  |
| 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 |
| Ana Garcia | University Carlos III of Madrid |
| Lorenza Giupponi | CTTC |
| 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 |
| Liu Ju | Huawei |
| Georgios Karagiannis | Huawei |
| 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 |
| Marco Miozzo | CTTC |
| 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 |
| Ayman Radman | Instituto de Telecomunicações |
| Parm Raewal | Vodafone |
| 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 |
| Juergen Sturm | Texas Instruments |
| Natalie Samovich | Enercoutim |
| Pablo Serrano | University Carlos III of Madrid |
| Dimitra Simeonidou | University of Bristol |
| Alena Sjarheyeva | Gradiant |
| 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 |
| Ovidiu Vermesan | Sintef Digital |
| Rainer Wansch | Fraunhofer Institute |
| Simon Watts | Avanti Communications |
| Elisabeth Weller | Eutelsat |
| Michael Welzl | University of Oslo Norway |
| Wen Xu | Huawei |

1. For the sake of having a neutral future-proof placeholder name. [↑](#footnote-ref-1)
2. 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. [↑](#footnote-ref-2)