Towards 5G Software-Defined Ecosystems

Technical Challenges, Business Sustainability and Policy Issues

Antonio Manzalini, *Telecom Italia Mobile, Italy*

Cagatay Buyukkoc, *AT&T Labs, USA*

Prosper Chemouil, Sławomir Kuklinski *Orange, France*

Franco Callegati, *University of Bologna, Italy*

Alex Galis, *University College London, UK*

Marie-Paule Odini, *HP, France*

Chih-Lin I, Jinri Huang, *China Mobile, China*

Mike Bursell, *Intel, UK*

Noel Crespi, *Telecom Sud Paris, France*

Eileen Healy, pdv Wireless, USA

Stuart Sharrock, Telemates, UK

Abstract — Techno-economic drivers are creating the conditions for a radical change of paradigm in the design and operation of future telecommunications infrastructures. In fact, SDN, NFV, Cloud and Edge-Fog Computing are converging together into a single systemic transformation termed "Softwarization" that will find concrete exploitations in 5G systems. The IEEE SDN Initiative[1] has elaborated a vision, an evolutionary path and some techno-economic scenarios of this transformation: specifically, the major technical challenges, business sustainability and policy issues have been investigated. This white paper presents: 1) an overview on the main techno-economic drivers steering the "Softwarization" of telecommunications; 2) an introduction to the Open Mobile Edge Cloud vision (covered in a companion white paper); 3) the main technical challenges in terms of operations, security and policy; 4) an analysis of the potential role of open source software; 5) some use case proposals for proof-of-concepts; and 6) a short description of the main socio-economic impacts being produced by "Softwarization". Along these directions, IEEE SDN is also developing of an open catalogue of software platforms, toolkits, and functionalities aiming at a step-by-step development and aggregation of test-beds/field-trials on SDN-NFV-5G. This will prepare the ground for developing new ICT ecosystems, thereby improving the quality of life and facilitating the development of the new digital economy.

Keywords — SDN, NFV, 5G, Open Mobile Edge Cloud

I. INTRODUCTION

A number of techno-economic drivers are converging to create the conditions for a paradigm change in the design and operations of future telecommunications networks and services. These drivers include progress in Information Technologies (IT), pervasive diffusion of ultra-broadband (fixed and radio) access, the falling costs of hardware, the maturity of virtualization techniques, a wider and wider availability of open source software and, eventually, ever more powerful terminals.

Software-Defined Networks (SDNs) [1], Network Function Virtualization (NFV) [2], Cloud [13] and Edge-Fog computing [12] can be seen as different dimensions of an overall trend that has been named by the IEEE SDN Initiative as the "Softwarization" of telecommunications (Figure 1).

"Softwarization" is an overall techno-economic transformation impacting the design, implementation, deployment and operations of infrastructures, deeply integrating network nodes and IT systems. It fully exploits the nature of software, such as

[1] Manzalini, A., et all, IEEE SDN Initiative SDN4FNS white paper "Software-Defined Networks for Future Networks and Services - Main Technical Challenges and Business Implications", January 2014, http://sdn.ieee.org/publications

flexibility and rapidity, for both network functions and services. This transformation will enable new architectural models, in turn implementing automated operations processes (e.g., self-management) while opening innovative Information and Communications Technology (ICT) service paradigms 0.

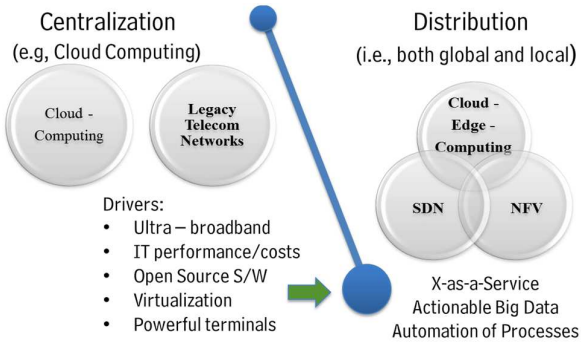

Figure 1: Softwarization of Telecommunications

It is very likely that this transformation will find first concrete expression in the 5G (Fifth Generation) of network and services infrastructure, which will be much more than a direct evolution of current LTE-4G networks. It is expected that 5G will handle 1000 times the current wireless area capacity, radically reduce the average service provisioning time, and meet significantly more stringent performance targets for reliability (packet error probability) and latency.

The 5G infrastructure will also enable a wide variety of new applications and ICT services. In fact, the huge amount of data collected by sensors – embedded in all sorts of terminals, machines, and things – will be networked with low latency fixed-radio connections, elaborated in the Cloud and Edge-Fog Computing facilities, to eventually be actioned into a variety of ICT services.

This enormous amount of data will create new service scenarios such as the Internet of Things, Tactile Internet, Immersive Communications, and, in general, X-as-a-Service. For example, 5G will enable remote the radio control and programmability (via (via Application Program Interfaces) of advanced robotic systems, with various applications for both industry (e.g., Industry 4.0) and agriculture (e.g., Precision Agriculture).

5G systems will therefore assume the characteristics of a powerful networking-computing-storage infrastructure. Its functions will be partly distributed and partly centralized, supporting pervasive connections (both wired and mobile) characterized by both high capacity and very low latency (of only a few milliseconds).

This techno-economic transformation of telecommunications is currently under the spotlight, not only in academic research communities but also in several industrial initiatives. This is clearly evident in forums and bodies like ONF, IETF, ITU-T, and ETSI that are developing reference architectures for SDN, NFV, Cloud and Edge-Fog Computing. At the same time, there is still a fragmentation in these efforts, which is delaying, if not jeopardizing, an effective exploitation of this techno-economic transformation worldwide. The IEEE SDN Initiative, established in 2013 by the IEEE Future Directions Committee, is a cross-society IEEE program with the goal of contributing to overcoming such fragmentation by proposing a leading effort and vision for Softwarization which includes not only technological aspects but also business sustainability and policy issues.

This is the second white paper of the IEEE SDN Initiative. Specifically it reports: 1) an overview of the main techno-economic drivers steering the "Softwarization" of telecommunications; 2) an introduction to the Open Mobile Edge Cloud vision (which will be the objective of a companion white paper); 3) the main technical challenges concerning operations, security and policy; 4) an analysis of the potential role of open source software; 5) some use case proposals for proof-of-concepts; and 6) a short description of the main socio-economic impacts being produced by "Softwarization".

II. TOWARDS THE 5G ERA

A. C-RAN: Softwarization of the RAN

The Radio Access Network (RAN) is the most important asset for operators: base stations outnumber the nodes in the core networks and are closer to and directly perceived by end users. The significance of softwarization on the RAN is self-evident.

As an essential element for 5G systems, the centralized, collaborative, cloud and clean RAN (C-RAN) [3] was proposed as early as 2010.

A C-RAN system centralizes different processing resources together to form a cloud in which the resources could be

managed and dynamically allocated on demand. With virtualization implemented, standard IT servers are used as the general platform with computation and storage as the common resources, on which run different kinds of applications. The indispensable applications in C-RAN are those to realize different radio access technologies including 2G, 3G, 4G and future 5G. In addition, the C-RAN platform could provide a set of standard APIs opening the opportunity for new service provision and deployment.

In the era of 5G, C-RAN itself needs to evolve to accommodate new features, new use cases and new requirements to better support 5G. On the way towards C-RAN softwarization in 5G, there exist several challenges.

First, the C-RAN architecture itself needs to evolve. Although C-RAN embodied the softwarization spirit from the very beginning, it used to be thought more as a means of network implementation (e.g., using Commercial-Off-The-Shelf (COTS) platforms rather than proprietary platforms). It does not change the network architecture of 2G, 3G and 4G defined in 3GPP. However, for 5G with many more requirements such as ultra-low latency, network slicing, and extreme flexibility, the design of the 5G architecture requires co-design of the C-RAN architecture to support such features. Examples include how to support control/user plane separation, and how to incorporate SDN controller and multi-RAT controller.

The fronthaul issue comes next. A fronthaul (FH) link is typically a connection between the baseband unit (BBU) and a remote radio head (RRH). As traditional FH protocols such as Common Public Radio Interface (CPRI) [4] process any shortcomings as constant high data rate without taking account of dynamic wireless traffic, low transmission efficiency, poor scalability, etc., there are increasing concerns that they are not suitable for large-scale C-RAN deployment in 5G networks, especially when massive multiple-input multiple-output (MIMO) is introduced. Several schemes have been proposed to either improve CPRI itself or even redefine the fronthaul interface. One such scheme is the Next Generation Fronthaul Interface (NGFI) concept proposed by China Mobile [5][6]. The essence of NGFI is to redesign the fronthaul interface to make the FH data stream traffic-dependent (therefore dynamic) and antenna-independent. Based on this the underlined transport networks could be designed more efficiently. The key ways towards NGFI include redesign of the BBU-RRH function split and packetization of FH data. By decoupling the FH bandwidth from the antenna number, NGFI can better support large antenna technologies. In addition, the cell-processing functions should be decoupled from the UE-processing functions to make NGFI traffic-aware, which can exploit the statistical multiplexing gain to improve efficiency and further reduce power consumption. It is also suggested that the function split schemes for downlink and uplink could be different to improve flexibility and efficiency. The use of Ethernet for NGFI transmission brings the benefits of improved reliability and flexibility due to the packet-switching nature of Ethernet. In the meantime, jitter, latency and accurate timing distribution mechanisms remain the key difficulties to overcome to realize NGFI transportation.

Virtualization implementation to realize resource cloudification is another challenge. Due to the characteristics of intensive computation and extremely strict real-time requirements on wireless communications, especially on the physical layer process, implementing virtualization technology to realize radio access technologies such as LTE is not an easy task, not to mention the future 5G new radio technologies. Fortunately there has been extensive pioneering work on this front. For example, China Mobile has successfully demonstrated a virtual machine-based LTE implementation running on COTS platforms in field trials. Despite the demonstrated functionality and desirable performance, there is still much room for improvement, including further enhancement of real-time performance, seamless live migration for the sake of energy saving, and standardizing the interface. In addition to the virtual machine-based virtualization technology, there are many other new promising technologies such as container which are also worth further investigation.

Software architecture is another important aspect for C-RAN softwarization in 5G. Traditional wireless network design follows “cell-centric” principles, i.e., resource allocation, mobility management, cell planning and optimization, etc. are on a cell basis. In 5G, there is a paradigm shift from cell-centric towards “user-centric”. The user-centric design depends on several key technologies including data/control plane separation, UL/DL decoupling and C-RAN is deemed to facilitate the realization of user-centric networks [7]. However, in traditional base stations, the system software architecture is

designed based on traditional vendors' proprietary platforms consisting of Digital Signal Processing (DSP), Application Specific Integrated Circuits (ASICs), etc. to meet the cell-centric purpose. In C-RAN in 5G, the systems would operate based on COTS platforms consisting of standardized IT servers, switches, and storage. All the resources are in the cloud and allocated on demand according to user needs. Thanks to the difference between the COTS platforms and traditional DSP-constituted platforms, and more importantly, due to the difference in the design principles from cell-centric to user-centric, the whole software system architecture in C-RAN needs to be reconsidered to exploit the cloud computing features and capabilities of COTS platforms as much as possible. The idea could be strengthened as far as the 5G requirements such as high agility, flexibility and scalability are concerned. In addition, network slicing, which is one of the key features of network softwarization, requires the cloud resources be reconfigured in a fast, agile, dynamic and cost-effective way. This also imposes requirements on careful software architecture design. In this sense, software architecture redesign is a critical issue for future study.

Last but not least, the introduction of SDN in C-RAN should not be neglected. Traditionally the concept of SDN mainly applies in the transport/routing area with the basic idea of control/data plane decoupling to realize the programmability of the control plane. With FH transport networks, in particular when NGFI is introduced, it is natural to extend the SDN concept to C-RAN. There should be an SDN controller located in the C-RAN cloud, deciding on the optimal FH routing path. This work could be coordinated with the management system or orchestrator in the cloud. The system architecture, the interface, the data flow and the coordination among the SDN controller and other control units are all worth further study.

In summary, as the essential element of 5G, the concept of C-RAN is firmly in line with the essence of "Softwarization" of telecommunications. On the one hand, C-RAN claims benefits such as facilitation of signal joint processing, deployment of mobile edge computing, multi-RAT coordination, and user-centric network realization. On the other hand, to achieve these benefits requires careful and optimal design of C-RAN from various aspects, including the architecture, FH transportation, virtualization technologies, software architecture redesign, SDN, management, and orchestration.

B. An end-to-end vision for 5G

5G era is aiming at an End-to-End (E2E) vision that includes the evolution of the RAN, the Next Generation (NG) core, and a management/control plane that extends User Equipment (UE) to the core and beyond.

As mentioned 5G is much more than an air interface beyond current LTE-4G. 5G will include evolutionary components of current generations of mobile networks (under a unifying umbrella). It also includes revolutionary components that will enable energy and spectral efficiency, a new resilient framework (i.e., responsive, auto-manageable QoS/QoE, secure, survivable, traffic and disruption tolerant) for services to everyone and everything (applications and machines).

5G requires a complete revamping of the E2E architecture, new service capabilities, rethinking of interfaces, management and control frameworks, access and non-access protocols and related procedures, functions, and advanced algorithms (e.g., Authentication, Authorization and Accounting (AAA), auto-maintenance and management of services) and any resource types (both physical and virtual).

Several challenges are still in the process of being addressed to meet stringent performance targets set out by the 5G community. These include 1000 times higher mobile data volume per area, 10 times to 100 times higher typical user data rate, 10 times to 100 times greater number of connected devices, 10 times longer battery life for low power devices, and five times reduced E2E latency. Moreover, the infrastructure needs to be highly flexible and scalable thus meeting foreseen and unknown requirements. Resiliency and responsiveness must be built into the design. Complexity is a big issue that needs to be measured and evaluated as part of this comprehensive redesign.

Service Providers (SPs) and network operators are currently deploying transformative approaches to provide network functions in appropriate infrastructures (using both centralized and distributed flexible architectural concepts) and thus providing flexible and scalable capabilities according to required use cases and their traffic demands. This flexibility will be achieved using a software-defined ecosystem and NFV technologies as well as data path programmability. The target architecture has to be cost and resource efficient as well as auto-managed and flexible for new innovations.

Significant adoption of Cloud Edge-Fog computing, SDN, NFV demands new thinking in various key areas to be able to fully utilize and monetize the capabilities presented: e.g., distributed system architecture, provide minimum "state" information, elastic and scalable systems in a consistent way, loose coupling and necessary event handling. Auto-management and control of mobile networks using new and innovative paradigms will be crucial.

IEEE SDN argues that any aspects require new and innovative work. These include context management (e.g., related to service, network, and device information); a Control, Orchestration, Management and Policy (COMP)[2] details related to eRAN; spectrum management; E2E resilient service composition; mobility management; low power-long range and various Machine-Type Communications (MTC); Device-to-Device (D2D) services; Radio Resource Management (RRM); and modular radio interfaces and a new protocol independent layering. Similarly, a new set of devices would also have modular capabilities developed around context, interference and Radio Access Technology (RAT) management, since end devices would be an integral part of RAN.

Most of the research and innovation efforts need to be in place in the next few years so that large field trials and testing can occur for early deployments to happen in 2020. This can be realized only through global collaboration and investment in key technologies and related fields. Since the required set of capabilities is very broad, mobile and wireline ecosystems need to be established that will allow global participation through open frameworks.

C. Open Mobile Edge Cloud

Various efforts are in progress in the RAN and core areas to address the architectural principles outlined above. In the RAN space, one of the promising architectures was identified as Cloud RAN (C-RAN, various flavors) as it provides a transition path to the cloud computing-based architecture. C-RAN architectures have been in trials in various countries and research labs for the past few years to determine the major benefits, challenges and solutions. The major challenges are fronthaul requirements (e.g., delay, jitter, cost, technology) and the ability of centralized baseband units (BBUs) to provide adequate signal processing in performance targets which basically determine the required spectral and energy efficiencies.

Several variants of C-RAN are proposed to address the fronthaul restrictions; one of the promising architectural directions is to decouple user and control planes and progress using the SDN strategy. This also allows a major rethink of the mobility edge (and subsequently the converged wireline/wireless edge). In this framework, a deconstruction of basic functions of RAN and core networks is followed by the definition of new architectural elements using the deconstructed functions.

IEEE SDN argues that one key area of this exercise is the introduction of a new functional node as an intersection point of these functions in order to create a future proof architecture.

This functional node, called Open Mobile Edge Cloud (OMEC) node, will be deployed to provide seamless coverage and execute various control plane functions as well as some of the "core functions" currently placed in various nodes of the Evolved Packet Core (EPC). More functionality related to compute and storage will be added to enable true cloud capabilities in closer proximities. Since many location-based applications are on the rise (social, analytics, video, etc.) fronthaul load will be considerably higher in the future. Requirements on local storage, compute and networking processing of "edge" services almost forces a new architectural direction.

---

[2] **COMP**: many SPs have example implementations of Control, Orchestration, Management and Policy frameworks that are part of a larger ecosystem that specifies standardized abstractions and interfaces that enable efficient interoperation of the ecosystem components. They are collections of software components which collectively are responsible for the efficient control, operation and management of capabilities and functions.

ETSI's Mobile-Edge Computing (MEC) Industry Specification Group[3] and "Fog computing[4]" are efforts in this direction; trying to address similar issues and identifying that a substantial amount of storage, communication, control, configuration, measurement and management should be placed at the "edge" of a network, in addition to the current cloud paradigms. This idea is based on the premise of certain extensions of Cloud computing architectures to the network edge, up to the Users' equipment/terminals.

Figure 2 shows an example of functional decomposition of NG UE, RAN and core functions for an E2E architecture of mobile networks in the 5G era. All these are related approaches but much more needs to be done.

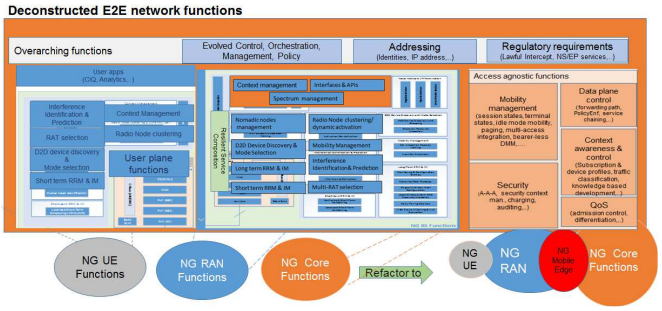

Figure 2: The Functional Decomposition of NG UE, RAN, Core and End-to-End Functions

The deconstruction of functions is a prelude to reconstruction and optimal placement of functionalities, much more than networking, to refactored nodes to address all the considerations outlined above. It is envisioned that NG Mobile edge (subsequently converged edge) will be the center of all 5G era networks with compute and storage functionalities attached. OMEC is an architectural paradigm based on this framework.

There are multiple variants of this key idea that essentially pushes some applications, analytics and computing, content and storage to the edge (including the edge devices). What is being done in the networking space should converge with similar ideas in the compute and storage dimensions. To include the UEs and other Customer Premise Equipments (CPEs) in this methodology requires many new collaboration capabilities to be developed to execute traditional RAN functions (e.g., RRM) along with others such as mobility management and security in this architectural framework as well as content delivery, storage and compute functionalities.

Given the vision outlined above, and the strategic role it would play as an intersection of NG Core and RAN functions, the edge needs to be properly defined. In the IEEE SDN workshop that took place in November 2015[5] on OMEC, it was defined as:

An open cloud platform that uses some end-user clients and located at the "mobile edge" to carry out a substantial amount of storage (rather than stored primarily in cloud data centers) and computation (including edge analytics, rather than relying on cloud data centers) in real-time, communication (rather than routed over backbone networks), and control, policy and management (rather than controlled primarily by network gateways such as those in the LTE core).

Note that this definition substantially re-architects the whole network. Key components are an application delivery framework on a cloud-based system with key functionalities refactored from NG RAN and Core.

The broad set of use cases outlined in various research and standards bodies points to a new set of applications that are limited by human physiology and psychology[6].

What differentiates the 5G era networks is the ability to address varying degrees of requirements (in delay, throughput, types and quantities of devices, etc.) concurrently with a unified framework. This almost dictates a new architectural component that is in close proximity to end users/devices with at most 10km distance to provide the new control and steering applications brought by new use cases.

[3] http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

[4] http://www.openfogconsortium.org/

[5] http://sdn.ieee.org/pre-industrial

[6] http://eandt.theiet.org/magazine/2015/03/tactile-internet-5g.cfm

These requirements are already very well known within the industry and there are approaches within the 5G community to address them.

All architectural work should be based on a set of Non-Functional Requirements (NFRs) on technology, business, and quality. These are the high level requirements and constraints that determine the evolution direction. Then the architecture[7] becomes the functional implementation of these requirements based on these constraints.

A key question that needs to be answered for the OMEC architecture is to determine how to merge various activities at the edge on a common platform based on the NFR constraints.

III. Security, Policy and Regulation Issues

SDN and NFV have much to learn from existing security best practices in enterprise and the cloud, but there are specific challenges that arise with large-scale "softwarization" projects, particularly in the telecommunications industry.

Rather than attempt an exhaustive survey of the issues that arise, three specific areas are examined to give an overview of some of the key issues facing service providers: "Planning, Policy and Regulation", "Infrastructure" and "Operations".

A. Planning, Policy and Regulation

In a traditional, non-virtualized environment, the domains of control of various entities are generally fairly clear: network administrators manage the network, systems administrators manage the hardware systems, OS stacks and patching, storage administrators manage the storage, and so on. With the advent of widespread virtualization of all the various components of the data center, the separation between different components becomes less clear. However, the importance of maintaining appropriate authorities and responsibilities does not diminish, even if the sets of components managed by different administrators changes. It is a key security requirement that trust domains are defined between, for instance, the various components – virtual machines, containers or non-virtualized hardware – of a Virtual Network Function (VNF), and that the administrators of this trust domain are not confused with – nor have control over – separate trust domains such as the hardware and software infrastructure on which the virtualized components execute.

It is not only virtual network functions which are being virtualized, of course: service providers are also enjoying the benefits of the softwarization of Management and Orchestration (MANO) components. Trust domains need to be considered not only for the MANO components, but also for the infrastructure – physical and virtual – that underpins them. SDN controllers fall firmly within this category, as compromise of an SDN controller may mean loss of control of significant portions of a service provider's network.

The complexity of managing the network at various different layers opens another issue: that of network topology. Network topologies should be an expression of operational – and security – policy, but in a world where routing rules can change on a second-by-second basis, there are three specific challenges:

1. The creation of a resilient policy;
2. The mapping and application of the policy to real hardware and software;
3. The visualization and enforcement of the policy, typically through visualization and enforcement tools.

In fact, once any trust domain has been defined, establishing and maintaining it will require use of a variety of monitoring and enforcement tools, including attestation, Intrusion Defense Systems (IDS) and Network Domain Security (NDS), and careful management of software (including vendor-provided image) provenance and integrity. Definition – and the mapping and enforcement – of these trust domains is not simple, and is further complicated by the need for some trust domains which span others. Security monitoring, management and application of policy, for instance, will need to cross multiple trust domains. Added to this complication is the fact that some will span geographical boundaries. This includes components which reside outside the data center, in the case of vCPE and base station equipment, for instance, but also across legislative and judicial boundaries.

[7] See for example Architecture, constraints, and behavior, John C. Doyle and Marie Csete (http://www.pnas.org/content/108/Supplement_3/15624.abstract)

The issue of legislative and judicial boundaries raises the question of regulatory bodies. There are various requirements that are placed on service providers. Examples from the USA include requirements associated with Personal Identifiable Information Protection, the Payment Card Industry Data Security Standard (PCI-DSS), the Health Insurance Portability and Accountability Act (HIPAA), Critical National Infrastructure and Lawful Interception. Many of these have significant impact on the types of trust domain and the controls between them, a particularly notable example being the requirement to keep data in virtualized resources – whether parts of VNF or MANO components – confidential from non-authorized entities, as well as integrity-protected. In the short to medium term, technical controls will not be sufficient to provide all the required protections, reliance will continue to be placed on human and physical controls.

B. Infrastructure

The provisioning of appropriately secure infrastructure is a keystone to any securely-managed deployment. With open source software making up a significant part of many deployments, it is important to have a good view of the security of such software. The first point that should be made is that, for many service providers and operators, the provision of open source software will not be directly by them, but by a vendor who will undertake to support it. To some extent, then, liability for the security of the software will lie with another party. A firm understanding of the liability and support arrangements behind the use of open source software is important in any deployment, but where that software has particular security functions, it is even more vital. It should be stressed, of course, that no system – either software or hardware – should be considered completely secure. Many of the security functions required for full softwarization – virtualization, containers, vSwitches, cryptographic libraries, etc. – are very complex, and seemingly minor mistakes in implementation may have major and far-reaching impacts on the service offering. What is more, the worldwide security community has shown time and time again that "security through obscurity" as practiced by some commercial vendors can be next to worthless. This does not mean, however, that the openness of open source software necessarily guarantees its security. There have been several examples of key security functions being shown to be incorrectly implemented, and even some cases of the public repositories in which such software is stored having been compromised, leading to concerns about the trustworthiness of the available code. The most robust code comes not from "many eyes", but from multiple *expert* eyes.

Operators planning to deploy open source software have both the opportunity and a responsibility to ensure that sufficient due diligence has been performed over that software, particularly when it supports core security functions.

One approach to mitigate security-related implementation errors or bugs in software that can be applied to either open source software or proprietary software is the provision of heterogeneous systems within a single deployment. Although this may be considered to increase the attack surface of a deployment, by introducing more systems, in reality it can reduce the impact of a single vulnerability in a key piece of widely-deployed software. There is, of course, a trade-off between manageability and security, but even when proprietary software is being used, when that software implements open interfaces or protocols, opportunity exists for consolidated management of the different systems – though this, in itself, may introduce a single point of failure which is unacceptable to service providers.

It has become clear, given the various vulnerabilities that software inevitably introduces, and whether proprietary or open source software is employed, that an approach rooted in hardware measures is required to provide sufficient defense in depth to satisfy a number of the requirements for a secure platform for both VNF and MANO components. Use of hardware-based attestation mechanisms can improve the trust in particular platform instantiations and agglomerations of systems, but run-time protection is more complex. Hypervisors already make use of chip-level hardware instructions to provide memory and process isolation between virtual machines, but protection of the administration layer from malicious or compromised workloads, and of the workloads from a malicious or compromised administration layer, will require further hardware measures. Containers in their standard Linux implementation currently make little use of hardware isolation. Hardware-mediated execution environments are expected to provide capabilities to allow isolation between layers of execution such as the hypervisor, vSwitch and virtualization components.

C. Operations

As noted above, management and maintenance of a deployment with multiple trust domains is a complex undertaking. A set of security policies, management capabilities and monitoring capabilities to support them is vital. Monitoring must be able to detect a variety of issues, of which reaction to malicious attacks is the most obvious. In order to do this, telemetry, agents and probes will be required in various positions (logical and physical) within the network and infrastructure: without aggregation across various layers, these inputs will be of significantly less utility. Malicious attacks are not, however, the only type of event to which operational reactions must be made. Probing – passive or active – of parts of the network or infrastructure may be the precursor to a full-out attack, and may occur at various levels: again, without aggregation and pattern-matching, such probing may not be detected.

An actual failure may occur for one of several reasons – one of which is a malicious attack – and may or may not have an impact on the security posture of the deployment. One of the opportunities offered by SDN is the ability, at least in some cases, to reconfigure the network to mitigate against such failures. Such reconfigurations should be in line with topology policies. NFV also offers opportunities for mitigation of failures, as VNFs – or their components – may be redeployable to nodes and hosts which are not affected – or less affected – by the failure.

In all these cases, there are likely to be options for different mitigation strategies. In some cases, the most secure is to "fail safe" – which may involve closing down a service. However, one alternative model – well supported by the SDN and NFV approaches noted above – is to accept a degradation of service, balancing impacts in various metrics such as security, performance and reliability whilst maintaining some levels of service. Although a common approach in the enterprise, reconciling this sort of degradation with the Service Level Agreements (SLAs) usually associated with telecommunications services will be a challenge. Another alternative may be to embrace the ability to sacrifice certain services (or degrade them to a larger degree) to maintain other, more critical services. These choices are enabled by softwarization, and require careful preparation and policy design.

One final point is the importance of managing trust models once they have been established. By default, trust relationships should be assumed to degrade over time. Neither can it be assumed that a trust relationship in one direction should be mirrored in the other direction: trust relationships are rarely symmetric. To give one example, the level of security controls implemented for an SDN controller will typically be higher than for the switches that it manages. It is therefore quite feasible – even probable – that vSwitches will fail or be compromised, and the SDN controller should expect these events and any trust model should take them into account when policies are being designed and implemented. The failure – or worse still, compromise – of an SDN controller is an altogether more complex problem to detect, let alone manage, and the ability of vSwitches to cope with such an event is likely to be much lower.

In summary, there are a number of areas where softwarization brings new challenges, or at least complexities, to security planning, operations and management. Some of these areas can be addressed with existing techniques whereas others – the use of hardware-mediated execution environments, for example – require new mechanisms and approaches. There are also opportunities: increased telemetry from NFV hosts and infrastructure will allow for mitigations as more traffic (North-South and East-West) is recorded, alongside performance and state metrics from various components of the deployment. The scale of these benefits and the challenges are yet to be discovered: security is still an area of very active research within both SDN and NFV.

IV. Strategic Role of Open Source Software

"Softwarization" of telecommunications opens up different strategies: vendor proprietary software, in-house operator development or open source software, and often a mix of these. But with cloud, NFV and SDN the number of open source initiatives is increasing. They typically leverage upon relatively mature IT projects such as Linux, KVM, libvirt, OVS and OpenStack [8] but also enterprise SDN with OpenDaylight, OpenContrail, ONOS [9] and expand towards specific needs of telecom operators, with Data Plane Development Kit (DPDK) for instance. The Open Platform for NFV Project (OPNFV) federates multiple upstream projects into one reference implementation for telecom networks. These projects, initially focused on the virtualized SDN enabled infrastructure, in line

with corresponding standards, essentially ETSI NFV for NFV and ONF for SDN, are now evolving to the management stack. As the technology and the market progress, parallel and sometimes concurrent projects have also appeared, such as OpenMANO, Tacker or Open Baton for implementing the ETSI NFV MANO stack (Figure 3).

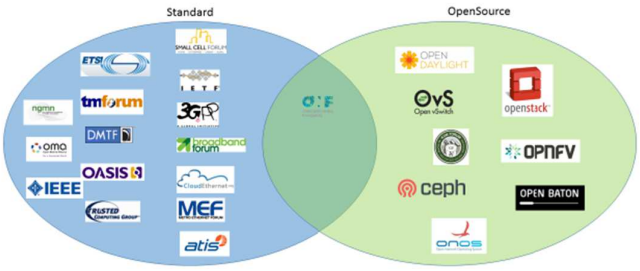

Figure 3: SDN-NFV Standards and Open Source Software

Some are driven by a very large community, and robust implementation platforms and processes, typically with OpenStack or Linux Foundation OPNFV, and OpenDaylight. Most of the code of these open source projects is released under Apache 2.0 license, a non-copyleft license, which opens up to different business models. These projects lead to software releases that are expected to be robust enough to support commercial deployment. But other projects have different objectives. They may be led by a much smaller ecosystem, with a less mature continuous integration mechanism and target different goals, more experimental. Open source can be interesting as collaboration projects to design some tests, or data models. Some standard organizations, such as IETF or ONF, look into open source projects as a way to validate specification with a reference implementation. Open source can also be initiated by a given vendor that decides to change business model, move from closed software to open software, and share his asset under a free open source license. More recently, an operator, Telefonica, released the OpenMANO NFV orchestrator of his testing platform under free license and initiated action to aggregate external actors to this community. Open source is a way to drive innovation by granting easy access to the code and creating an open ecosystem

The benefits of open source for the telecom market are numerous: first it is a unique tool for a broad community, with operators, vendors, universities, to agree on requirements, use cases and prototype a solution quickly. Full consensus is often not necessary to start coding, as the open source model is based on an iterative approach (Figure 4). Experimentation and community expansion bring new requirements, and more robustness.

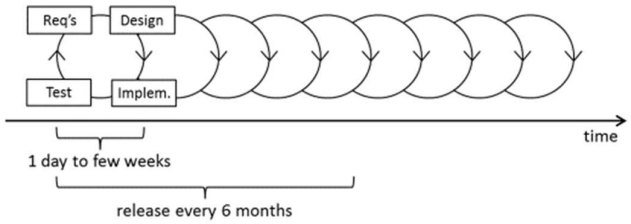

Figure 4: Open Source Iterative Model

Second, it is a great tool for research and universities: with software and open source, the entry barrier to build an experimental telecom network is getting lower and lower.

Third, it brings a common baseline across the industry and fosters better interoperability across vendors that adopt this technology and *de facto* across operators.

However the concern is that even now few operators adopt open source. Not only do operators have very few resources skilled to produce code and contribute to open source, but they are also very cautious of software coming from a community that does not aim to provide SLAs and standard assurance as proprietary solutions do, nor a clear roadmap as it is built iteratively from a kernel of subprojects. They would rather ask vendors to package open source software into a robust solution they can commit to deploy and support. Liability and risk remain key showstoppers to open source versus vendor proprietary software. Moreover open source is not only a question of technology; it becomes more and more a question of organization. As open source software is transforming the industry, it is difficult for non-native software companies to adopt its paradigms.

Nevertheless, open source is now clearly identified as complementary to standards to validate specifications with real neutral community-driven implementation. Open source is also the easiest and fastest way to fuel innovation across a broad ecosystem. With all-IP networks and 5G networks, more and more actors in the value chain become consumers and

producers, including producers of open source code that is live tested with a few virtual machines and enhanced on the fly by the community. Of course security is a big topic, but many studies and much work are also underway to cope with this.

V. SCENARIOS AND USE CASES

Multiple scenarios and use cases can be envisioned around softwarization of the network. Software-based solutions include the use of different types of virtualization, typically hypervisor or Linux container-based, NFV architecture, SDN control plane, and open APIs. They enable decomposition of the network and service layers into subcomponents that allow modular and multi-vendor architecture and software as service models. These allow service providers, infrastructure providers, and application vendors, to share services on the fly and expand towards new business models, as described in Figure 5.

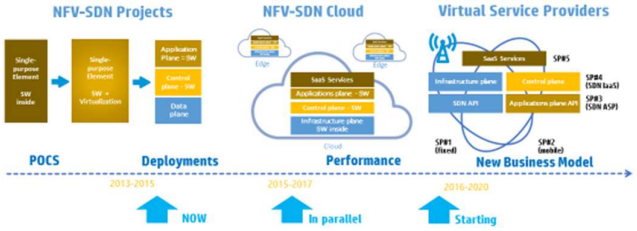

Figure 5: Softwarization, a Phased Approach

ETSI NFV defined nine use cases [10] mainly focused on the evolution of existing typical architectures for broadband or mobile networks towards virtualization, without disrupting the current status quo, i.e., 3GPP and some more specific SDN-NFV ones. But the industry is already exploring beyond this model, expanding into a combination of SDN and NFV, decomposition of the network functions, hybrid deployments with edge and cloud set-up, cross domain – multi operator environments, open management and service APIs including SLAs and monetization, etc.

A. Cloudification Scenario: from Core Network Optimization to VNFaaS Use Case

NFV is about virtualizing network functions, from residential customer set-top box to enterprise CPE, and network core functions such as Evolved Packet Core (EPC) and IP Multimedia Subsystem (IMS), and deploying them in a carrier grade NFV enabled cloud (Network Function Virtualization Infrastructure (NFVI)) across multiple data centers. On top of that SDN is implemented on the connectivity layer decoupling data and control planes and bringing extra flexibility at the packet forwarding level. All in all to reduce cost and adapt quickly to market dynamics as shown in case#1 Figure 6.

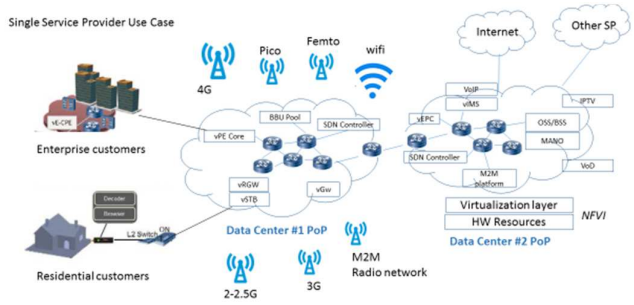

Figure 6: NFV and SDN to Optimize a SP Network

Further use cases define some VNFs that will be dedicated to certain enterprise customers. This is ETSI NFV VNFaaS use case to enterprise.

But VNFaaS can also be provided to other SPs. SP#1 may have a virtualized infrastructure and VNFs, and offer some functions to a 3rd party SP#2 such as vHSS (Home Subscriber Server) for a Mobile Virtual Network Operator (MVNO), or functions such as virtual media service function for voice mail, audio-video conference or transcoding, for other SP, MVNO or Over-The-Top (OTT) service providers. The VNF can be deployed on shared or dedicated resources and capabilities offered to SP#2 are up to the multi-tenancy capabilities of SP#1: configuration, scalability, monitoring, usage-based charging, etc. (Figure 7).

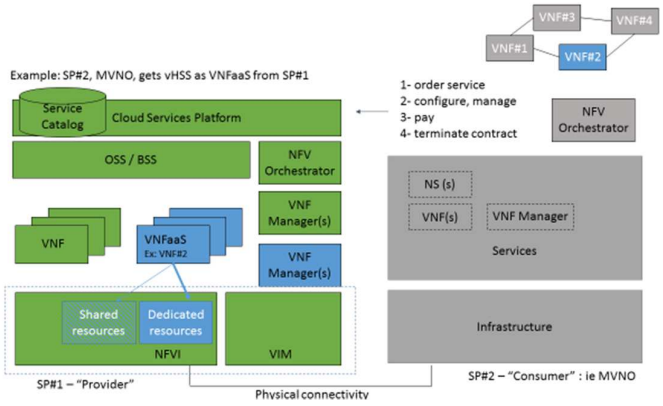

Figure 7: Virtual Network Function as a Service across SPs

B. "NFV-SDN Decomposition" Scenario: Virtual Mobile Core

Given the traffic growth on mobile networks and the impact of Machine-to-Machine (M2M) devices, virtualization of the

mobile core is one of the top NFV use cases to deploy in small instances, for example dedicated to certain businesses such as M2M, or to bring flexibility to adapt to traffic variations and scale up and down programmatically and rapidly. But SDN is also being explored: not only for cost reduction but also to introduce granular programmability at the data plane level to bring new capabilities, such as dynamic routing of traffic per user or application, Openflow-based Wi-Fi offload, or reducing the Signal to Noise plus Interference Ratio (SNIR) by dynamically selecting base stations. Leveraging SDN and NFV can optimize low or ultra-low latency providing placement of the SDN and virtual functions is designed properly (Figure 8).

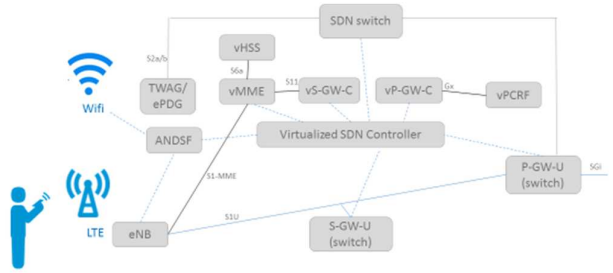

Figure 8: Virtualization and SDN in the Mobile Core

C. "Edgification" Scenario: vCDN

Virtualization of base stations and vCDN introduces virtualization to elements that are deployed at the edge of the network. While cloudification is a big trend to leverage cloud infrastructure and mutualization of resources, edge resources remain of high interest for services that require low latency or repeatable content to be distributed to end users, typically streaming blockbuster movies. Virtualizing end points such as CDN edge caching or mobile base stations to host some OTT or M2M vendor applications offers new capabilities to service providers and opens up new business models. It also processes some data at the edge and reduces the traffic being carried to the back end data centers (Figure 9).

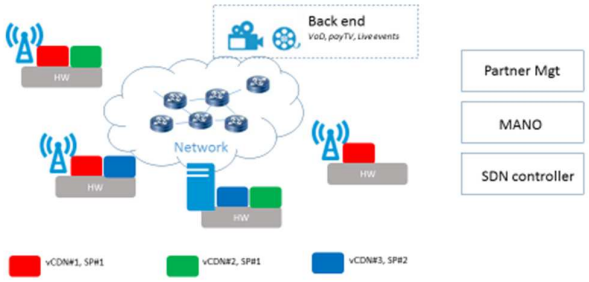

Figure 9: Virtualization of CDNs and Edge Deployment

D. "Autonomous Machines"

Robots, drones, autonomous machines, and Artificial Intelligence (AI) interfaces will be the 5G terminals of the future (Figure 10). The development of more and more complex cognitive capabilities through advanced terminals (increasingly powerful and sophisticated) attached at the edges of the 5G infrastructure, offers interesting opportunities not only to automate processes and optimize costs, but also to develop new service scenarios (Cognition as-a-Service). This will pose challenging requirements for ensuring ultra-low latencies in closing the interaction "loop".

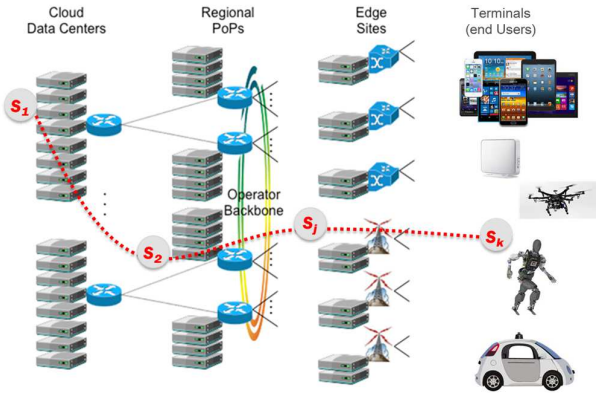

Figure 10: Robots, Drones, Self-Driving Vehicles becoming the "Terminals of the Future"

Today, the local computing power of a robot is not enough (for reasons of consumption, space, dissipation, etc.) to implement strong cognitive characteristics of autonomy. Tomorrow, thanks to 5G, it will be possible to make use of cloud robotics solutions which offer a huge amount of resources at low cost through cloud/edge computing. In fact, with 5G, the data collected from the several sensors of robotic systems, thanks to high bandwidth connections with very low latency, will be

transmitted to destinations where there are adequate computing power and memory resources – with an appropriate centralization/distribution balance. These data will then be transformed quickly into cognitive "decisions", transferred back locally and actuated by the robot within a few milliseconds. It will be possible, for example, to perform highly accurate operations at a distance, to optimize real-time control of any transportation system, and to manage business processes in a highly dynamic and flexible fashion.

An example of an application use case is the radio remote control of an industrial robot or any Autonomous Machine (AM). It implies actuating the orchestration of the service logic execution of the AM. A user wishes to control a remote AM performing a certain service (X-as-a-Service) in a certain area (e.g., a museum or a firm). The AM is equipped with sensors and actuators and has some local (but limited) processing and storage capabilities. The AM is connected to a low latency radio network, the area (where the AM is acting) also has radio access to processing and storage capabilities (both in edge points of presence and in the cloud), which might be required to execute complicated tasks.

In agriculture, for example, AMs can be used for tasks like crop inspection, targeted use of water and pesticides, actions and monitoring to assist farmers, as well as in data gathering, exchange and processing for optimizing the production and distribution processes. Cloud robotics and Industry 4.0 paradigms are full of other potential use cases. In general, these are ideal contexts where an OS can control and operate AMs in real-time (as they were nodes) for a number of different applications. Interestingly, APIs can be opened to end users and third parties to develop new types of services.

Besides agriculture and industry, it is likely that we'll see robotic applications also in the domestic environment: It is estimated that by 2050-2060 one third of people in Europe will be over 65. The cost of the combined pension and health care system will be over 29% of the European gross domestic product (GDP). Remotely controlled and operated robots will enable remote medicine and open up a new world of domestic applications which may be available to the entire population (e.g., cleaning, cooking, playing, and communicating).

E. Autonomous and Self-Driving Vehicles

The automotive sector is expected to pose very challenging requirements for 5G, with several use cases based on vehicle-to-vehicle communications and vehicle-to-cloud/edge. Other use cases concern services based on augmented reality dashboards installed in the vehicles: for example, displays that overlay information on top of what a driver is seeing through the front window and can identify objects in the dark or in fog by showing the distances and movements of objects.

Many car manufacturers are already adding driver assistance systems based on 3D imaging and built-in sensors and the first prototypes of self-driving vehicles are being tested. The technical requirements for self-driving cars call for ultra-low latencies and ultra-high reliability.

In fact, at the end of the day, a self-driving vehicle is a sort of complex robotic system equipped with sensors, actuators and ICT capabilities. Driving a car in real traffic is a very challenging task for machine intelligence: reaction times in milliseconds are required to avoid sudden and unpredictable obstacles, and maybe some form of "common sense" is also necessary. This demands considerable computing power (to minimize the application latencies) and very low network latencies.

Today, the local computing power that can be embedded in a vehicle is limited for reasons of space, dissipation, and cost. It is not enough for executing machine learning, heuristics or AI methods required to exploit such levels of autonomy. But the availability of enormous computing and storage power in the cloud encourages us to consider locating the "cognition" of the vehicle in the cloud.

VI. SOCIO-ECONOMIC IMPACTS ON THE VALUE CHAINS

Technology trends are pushing the competition to move towards OPEX-based business models (e.g., Pay as you go), radically changing the current value chains. This will be reflected in a general convergence process of IT systems and networks and (in the medium to long term) in the gradual disappearance of the distinction between the network and what connects to it, i.e., terminals, machines, smart things, etc.

Technology is going to become accessible to all enterprises in any part of the world on an equal basis, further reducing any competitive advantage due to location. Hence, the real

differentiator will be the capacity to innovate continuously. More and more the economics will shift from the economics of resources (becoming commodities) to the economics of data/information (and its related context). This will result in lower barriers to entry and thus lead to a larger number of players.

The impact of softwarization on the telecommunications industry can be seen as a substantial reinforcement of this general convergence process into converged infrastructure for all services (e.g., voice, internet access, and 'over the top' typical services). The emerging paradigm will be X-as-a-Service.

There is likely to be a split between Infrastructure Providers owning and operating the converged infrastructure, and the Service Enablers which offer the connectivity and network and service functions that enable Service Providers to develop and provision end-user services (e.g., retailing services). This split is also likely to drive a separation in the vendors supplying the infrastructure providers, the service enablers and the service providers.

The most likely merger will be between wholesale telecoms supply and data center hosting. In particular, the capital investment required to enter the data center hosting sector is likely to remain lower than that required to enter the access connectivity industry sector. In fact, many players offering global connectivity services are already also significant players in data center hosting.

The converged infrastructure can host a wide variety of network and service functions. Some of these services may not need to be executed in data centers and could instead be run in the middle of the network, involving virtualized functions to carry out intermediate information processing.

Some examples of these functions include:

- Content distribution networks;
- Content repurposing/recoding;
- Authentication, authorization, and access control;
- Content policing and filtering, content-based routing, content-based QoS management (e.g., Deep Packet Inspection (DPI));
- Intrusion detection;
- Firewall; and
- Content-based performance acceleration and bandwidth optimization (WAN acceleration).

These functions and others can be dynamically combined into complete services by constructing a specific chaining of functions – called service function chains. The orchestrator can be seen as a key system of such infrastructure. It would manage the different steps involved in the provisioning of virtual functions and services, such as creating and removing logical resources as well as installing, configuring, monitoring, running and stopping software processes in the logical resources. In this sense, the orchestration of these network and service functions is more linear (chaining) than traditional service orchestration which is based on a more articulated combination of service logics.

It seems likely that the retailing of traditional telecommunications services as a separate industry sector is going to disappear. Traditional telecommunications services will become packaged with other services such as voice with internet access and premium TV. Telecommunications retailing is likely to join with OTT service providers as voice becomes just another OTT service.

At the same time, there will be some merging in the supply of hardware between traditional telecommunications equipment suppliers and IT equipment suppliers. Some telecommunications equipment suppliers will reposition themselves as principally software supply companies. This will require a significant shift in the business model.

Many of the OTT service providers have no practical restrictions, be they technical, legal, or commercial, which means that they do not have to focus on a local national market. Many of these companies are truly global. The marginal cost of entering a new country is very low, assuming infrastructure is in place and is available to the OTT service provider. The introduction of SDNs and NFV enhances this situation, making the marginal cost of geographical extension even lower. Indeed, softwarization makes it possible to be present in a geography without having to have any physical infrastructure at all, neither people nor physical equipment.

VII. CONCLUSIONS AND IEEE SDN PLANS

Techno-economic drivers are creating the conditions for a change of paradigm in the design and operations of future telecommunications networks and services. SDN, NFV, Cloud and Edge-Fog Computing can be seen as different dimensions of a systemic transformation termed the "Softwarization" of Telecommunications.

This transformation will likely find concrete expression in the 5G network and services infrastructure, which will be much more than a direct evolution of current 4G networks (i.e., beyond merely an increase of bandwidth and reduced latency): 5G will be the "nervous system" of the digital society and economy.

The IEEE SDN Initiative has produced a white paper that provides: 1) an overview of the major techno-economic drivers steering the "Softwarization" of Telecommunications; 2) an introduction to the Open Mobile Edge Cloud vision; 3) the key challenges concerning security, policy and regulation; 4) the potential role of open source software; 5) some probable use cases; and 6) the main socio-economic impacts being produced by "Softwarization" [11].

The next step is a companion white paper describing the Open Mobile Edge Cloud paradigm in more detail.

The IEEE SDN initiative is also detailing the progress in the development of an open catalogue of software platforms, toolkits, and functionalities, aiming at a step-by-step development and aggregation of test-beds/field-trials on SDN-NFV-5G. This will prepare the ground for developing new ICT ecosystems, improving the quality of life and facilitating the development of the new digital economy.

REFERENCES

[1] White paper on "Software-Defined Networking: The New Norm for Networks" https://www.opennetworking.org/;

[2] White paper on "Network Functions Virtualisation" http://portal.etsi.org/NFV/NFV_White_Paper.pdf;

[3] C. M. R. Institute (2014) C-RAN: The road towards green ran. [Online]. Available: labs.chinamobile.com/cran

[4] CPRI, "Common Public Radio Interface (CPRI) Specification (V6.0)", Tech. Rep. Aug. 2013. Online: http://www.cpri.info.

[5] C. M. R. Institute (2015) White Paper of Next Generation Fronthaul Interface. [Online]. Available: labs.chinamobile.com/cran

[6] Chih-Lin I, Yannan Yuan, Jinri Huang, Shijia Ma, Ran Duan and Chunfeng Cui, "Rethink Fronthaul for Soft RAN", IEEE Commun. Mag. 53(9): 82-88

[7] Chih-Lin I, Rowell, C., Shuangfeng Han, Zhikun Xu, Gang Li and Zhengang Pan (2014) Toward green and soft: a 5G perspective. IEEE Commun. Mag. 52(2): 66-73

[8] Openstack: http://www.openstack.org/;

[9] ONOS: http://onosproject.org/;

[10] ETSI NFV Use Cases GS NFV 001.

[11] IEEE SDN Initiative http://sdn.ieee.org

[12] Ahmed, A., Ahmed, E., A Survey on Mobile Edge Computing - 10th IEEE International Conference on Intelligent Systems and Control, (ISCO 2016), DOI: 10.13140/RG.2.1.3254.7925

[13] Heilig, L., Voss, S., A Scientometric Analysis of Cloud Computing Literature, IEEE Transactions on Cloud Computing, Vol 2, Issue 3, April 2014, DOI: 10.1109/TCC.2014.2321168

GLOSSARY

3GPP	3rd Generation Partnership Project
5G	Fifth Generation
AI	Artificial Intelligence
AAA	Authentication, Authorization and Accounting
AM	Autonomous Machine
API	Application Program Interface
ASIC	Application Specific Integrated Circuit
BBU	Baseband Unit
CDN	Content Delivery Network
COMP	Control, Orchestration, Management and Policy
COTS	Commercial-off-the-shelf
CPE	Customer Premise Equipment
CPRI	Common Public Radio Interface
C-RAN	Cloud RAN
D2D	Device-to-Device
DPDK	Data Plane Development Kit
DPI	Deep Packet Inspection
DSP	Digital Signal Processing
E2E	End-to-End
EPC	Evolved Packet Core
eRAN	evolved RAN
ETSI	European Telecommunications Standards Institute
FH	Fronthaul
GDP	Gross Domestic Product
HIPAA	Health Insurance Portability and Accountability Act
HSS	Home Subscriber Server
ICT	Information and Communications Technology
IDS	Intrusion Defense Systems
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem

IMT	International Mobile Telecommunication
IT	Information Technologies
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KVM	Kernel-based Virtual Machine
LTE	Long Term Evolution
M2M	Machine-to-Machine
MANO	Management and Orchestration
MEC	Mobile-Edge Computing
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MTC	Machine-Type Communications
MVNO	Mobile Virtual Network Operator
NDS	Network Domain Security
NFR	Non-Functional Requirement
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NG	Next Generation
NGFI	Next Generation Fronthaul Interface
OMEC	Open Mobile Edge Cloud
ONF	Open Networking Foundation
ONOS	Open Network Operating System
OPEX	Operational Expenditure
OPNFV	Open Platform for NFV Project

OS	Operating System
OTT	Over-The-Top
OVS	Open vSwitch
PCI-DSS	Payment Card Industry Data Security Standard
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RRH	Remote Radio Head
RRM	Radio Resource Management
SDN	Software-Defined Networking
SETI	Search for ExtraTerrestrial Intelligence
SLA	Service Level Agreement
SNIR	Signal to Noise plus Interference Ratio
SP	Service Provider
UE	User Equipment
UL/DL	Uplink/Downlink ratio
VNF	Virtual Network Function
WAN	Wide Area Network