Mandate-driven Networking Eco-system: A Paradigm Shift in End-to-End Communications

Ingrid Moerman[*], Djamal Zeghlache[†], Adnan Shahid[*], Joao F. Santos[‡], Luiz A. DaSilva[‡], Klaus David[§],
John Farserotu[¶], Ad de Ridder[‖], Wei Liu[*], and Jeroen Hoebeke[*]

[*] imec - IDLab, Department of Information Technology at Ghent University, Ghent, Belgium
[†] Télécom SudParis, Institut Mines-Télécom, Institut Polytechique de Paris, France
[‡] CONNECT - Trinity College Dublin, Ireland
[§] Chair for Communication Technology (ComTec), University of Kassel, Kassel, Germany
[¶] Centre Suisse d'Electronique et de Microtechnique S.A.,Jaquet Droz 1, CH-2007, Neuchatel, Switzerland
[‖] Hermes Partnership, Zwijndrecht, Netherlands

Abstract—The wireless industry is driven by key stakeholders that follow a holistic approach of "one-system-fits-all" that leads to moving network functionality of meeting stringent End-to-End (E2E) communication requirements towards the core and cloud infrastructures. This trend is limiting smaller and new players for bringing in new and novel solutions. For meeting these E2E requirements, tenants and end-users need to be active players for bringing their needs and innovations. Driving E2E communication not only in terms of quality of service (QoS) but also overall carbon footprint and spectrum efficiency from one specific community may lead to undesirable simplifications and a higher level of abstraction of other network segments may lead to sub-optimal operations. Based on this, the paper presents a paradigm shift that will enlarge the role of wireless innovation at academia, Small and Medium-sized Enterprises (SME)'s, industries and start-ups while taking into account decentralized mandate-driven intelligence in E2E communications.

Index Terms—End-to-end Communications, Network Orchestration, Decentralized Intelligence, Mandate-driven View.

I. INTRODUCTION

Nowadays, network and service providers, as well as telecom-equipment manufacturers, are expanding their market beyond consumers by targeting the professional market, especially industry, to generate new revenue flows [1]. Key stakeholders realize this expansion using a holistic "one-system-fits-all" philosophy for serving industry verticals, leading to a central core specification with a multitude of interoperable options, that will survive or not on the market [2]. This approach is gradually leading to very complex systems, making it very hard for tenders to select the best options, and turning impossible for new and smaller players to bring in their innovations. To satisfy the future demands of multiple tenants, e.g., private and public service providers, as well as application developers on one side; and end-users, both consumers and verticals, on the other side, we see the need for open architectures, systems and solutions. The tenants and end-users must become active players of this evolution to bring in their own requirements and innovations.

In addition, there is a clear trend of moving radio and network functionality towards the core and cloud infrastructures; this dependency on the core and the cloud may be a limiting factor for smaller deployments and dynamic environments.

E2E communication involves different network segments, each having its own characteristics, community and domain-related expertise. Driving E2E solutions from one specific community may lead to undesirable simplifications and high-level abstractions of other segments, resulting in sub-optimal network performance [3].

Based on this assessment, in this paper, we propose a paradigm shift for broadening and reestablishing the role of innovative wireless research at academia, SME's, industries and startups in E2E communications. We target the professional market for integrating private network and services providers in a new standardization and specification paradigm fostering cooperation, interfaces and openness.

II. THE PARADIGM SHIFT

Our proposed paradigm shift is driven solely by the E2E performance between communicating applications, where E2E networks and services are dynamically composed through cooperation between the most appropriate combination of network segments and network services. Additional criteria for our envisioned evolution comprises E2E system optimizations towards higher resource efficiency, in particular spectrum and energy resources, as well as higher security and improved overall cost efficiency. Our vision promotes evolved standardization processes and focuses on Expressive Application Program Interface (xAPI), expressing user and device-specific profiles, requirements, services and intentions that are both technology and Operating System (OS) agnostic. This approach leads to absolute freedom for implementation and distributed optimization, and hence, maximizes the role of academia, SME, industries and startups in the innovation cycle of E2E communication ecosystems.

Our mandate-driven E2E vision is illustrated in Figure 1. It is inspired by several limitations identified in current network designs and based on several considerations, which are detailed in the following paragraphs.

Consideration 1: *Current network designs fall short in considering the end devices as a full part of the network.*
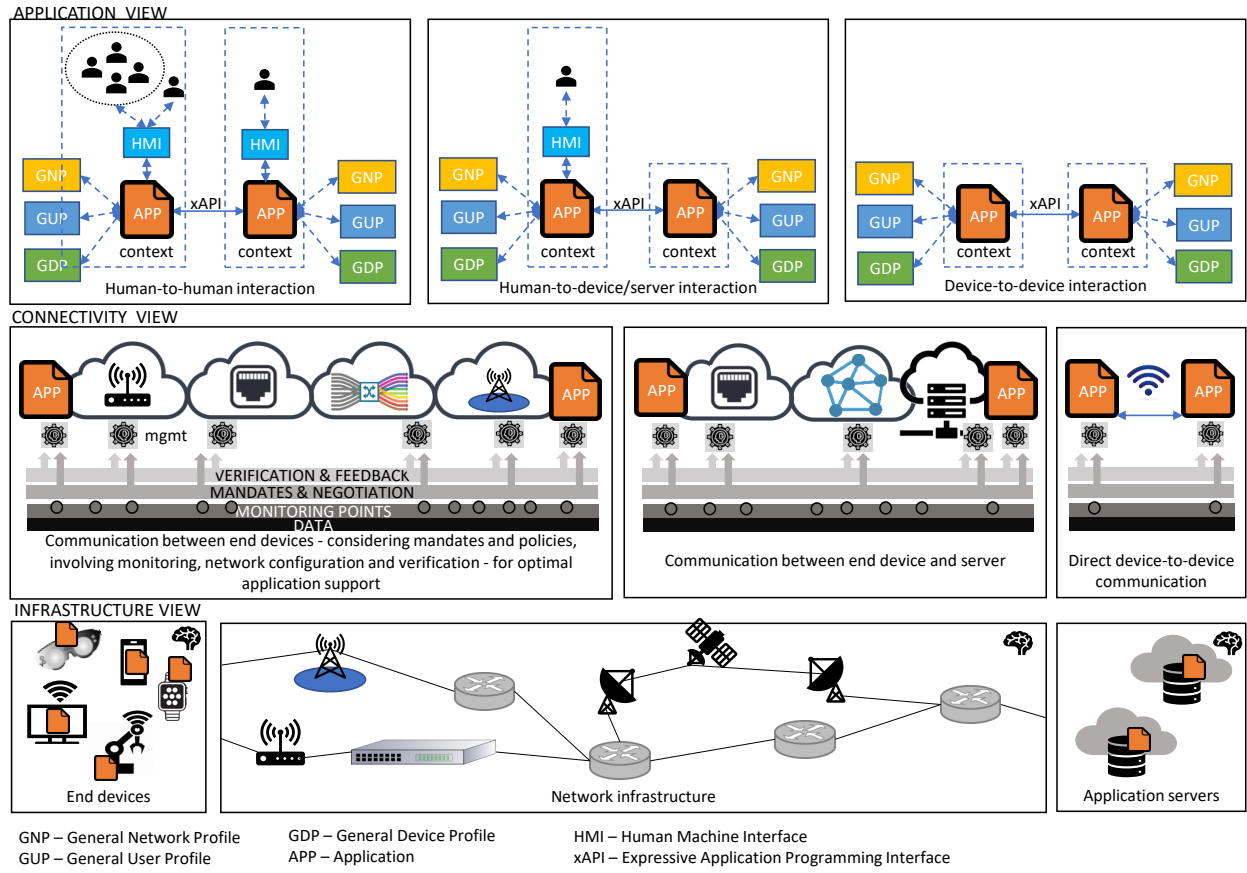
Fig. 1. Mandate-driven E2E vision.

GNP – General Network Profile
GUP – General User Profile
GDP – General Device Profile
APP – Application
HMI – Human Machine Interface
xAPI – Expressive Application Programming Interface

E2E communication happens between applications on end-devices and/or application servers. Devices can be very diverse: from tiny sensor devices to high-end servers and storage devices in data centers, from traditional smartphones, laptops and desktops to more sophisticated devices, e.g., AR/VR headsets. Some devices may interact with humans through user Human-Machine Interfaces (HMIs), e.g., keyboards, touchscreens, or more sophisticated interactions via speech, gestures or signatures (face, eyes) recognised by cameras. Other devices, referred to as 'machines', directly interact with other devices without human intervention. Independent of the type of applications and end devices, they should be considered as a first-class citizen in future networks and take up a more prominent role. In order to realize this inclusion of applications in the overall service establishment, application to network interfaces and APIs should evolve into agile services themselves. These network APIs must become an integral part of future network design.

Consideration 2: *There is a gap between applications/users and the networks they are using, preventing the sharing of information about policies, capabilities, etc.*

User involvement and participation in setting policies, rules and preferences should be improved in 6th Generation (6G). Indeed, access and distribution of data should not be deter-mined by the directives of users' terminals OSs or single companies, but by users based on a democratic consensus. Users should be able to express consent, preferences and opt-in and out with service providers and systems providing means to verify and audit the implementation of the user expressed privacy and personal data management [4]. The system must be able to dynamically create and setup new profiles based on context to drive the end-to-end service life cycle management.

Consideration 3: *Currently, there are no mechanisms for applications to communicate their fine-grained requirements to the underlying networks, and conversely, for these networks to monitor the status/health/requirements of their applications in real time*

E2E applications can be started manually by users, or triggered autonomously by events or context. When an application starts running, it should express its communication requirements, i.e., its specific QoS and traffic demands, to the underlying network in a technology-agnostic way. An application should not care by which network interface or technology it is served, as long its requirements are fulfilled and its preferences expressed in profiles are respected. Depending on the application needs and context, different types of network, e.g., wired, wireless, public, private, cellular, wide, local, proximity, and different spectrum bands, can serve

the application. Despite huge standardization efforts for the sake of interoperability, we have witnessed an explosion of new releases and technology standards. The belief that a single technology or operator can manage all today or future communications needs is unrealistic.

Consideration 4: *Networks behave as isolated systems, focusing on optimizing their own operation.*

An E2E connection between applications may be a single link i.e., in case of device-to-device communications, but it will most probably consist of multiple wired/wireless segments crossing different operator or administrative domains. E2E connections may change dynamically to keep fulfilling its requirements, while coping with changing contexts. This should happen transparently to the applications without loosing connection. Context information (location, time, weather condition, presence of other devices/persons in the environment, activity, mood) is also expressed in a generic way, which can be seen as a dynamic profile information that can be accessed to provide context-aware services to applications. Such dynamic information is also protected and secured according to the generic user/device/network profiles.

Consideration 5: *The granularity in supporting application requirements is too limited considering the diverse types of applications that will run over future networks.*

A single network is capable to serve many devices and very different applications requirements in terms of QoS requirements. The application to application communications and networking QoS should be defined in the E2E sense and should be expressed independently of the services and underlying networks that realize the E2E connection between applications. This application level QoS requirements need to be decomposed in QoS objectives for the segments that will collectively lead to the application desired QoS. Realizing the E2E QoS should go beyond typical aggregated capabilities, e.g., overall capacity and number of supported users/devices, by enabling applications to express requirements per flow throughput, different flavours of latency (in-time, on-time, cyclic), reliability, coverage, availability.

Consideration 6: *Applications do not have the ability to verify in detail whether guarantees have been met. Today's toolbox is limited to coarse SLAs and basic end-to-end measurements. Crucial data for end-to-end network management is missing.*

Future applications should have capabilities for verifying their E2E guarantees in real-time, for finding out how each involved network segment or even network component was able to deal with them. In turn, future networks must be supportive in making the required info available and consistency in the format of such information across networks should be considered. Based on the obtained insights, future applications can provide fine-grained feedback to the network, forming the basis for advanced, evidence-based network management.

III. Multi-disciplinary Research Challenges

In order to realise the E2E vision, several multi-disciplinary research challenges must be addressed and are detailed in the following subsections.

A. Profiles

Profiles are indispensable part of the E2E paradigm, which leverages communication entities including users, applications, networks, etc. to interact with other entities while respecting privacy, societal, and other related aspects. Users should have Generic User Profiles (GUPs) that keep track of personal data (age, gender, interests, contact info, subscriptions, home/family/professional/social related info) and preferences (security, privacy, consent, generic application preferences, priorities of services/applications). Construction of a user profile is as much as possible automated with minimal user interaction based on generic expressions independent from type of device, OSs, applications and networks the user interacts with and without the need for deep legal or technical knowledge from the user. This approach eliminates the duplication of profiles and preferences across different devices, OSs and applications. Moreover, each device possesses a Generic Device Profile (GDP), concerning, e.g., ownership (single person, company, community), users and Access-Control Lists (ACLs). Finally, E2E networks, and their network segments, have their own Generic Network Profiles (GNPs), containing information related to, e.g., business, security and protection policies; compliance with regulations, e.g., General Data Protection Regulation (GDPR) and limitations with respect to spectrum usage. The GNPs and preferences must be respected by the OS of every node attached to the network, as well as every application and network service, for ensuring that any data exchange is properly protected and secured, and not violating policies and regulations.

B. Social-technical Logical and Design Processes

In order to meet future objectives in privacy protection social, human and legal sciences have to be included at early design stages. These social and technical aspects are extracted from the profiles developed for the involved entities. 6G should provide privacy protection by design with the direct involvement of the end-users in the privacy protection configuration, control and management process. In essence, user intent and consent should drive the composition of privacy protection and security services irrespective of terminals and hosting infrastructures serving end-users. Actually, the same concept could be extended to all the stakeholders where not just dynamic SLAs could be established but also all security and privacy and property protection can be adjusted on situation, context and environment conditions.

C. Integration of End-users and Applications in the Network Design Process

The main focus of wireless system design today is still on increasing the capacity (mostly expressed in aggregated throughput and the number of simultaneously supported users

or devices) at the physical level. An enormous R&D effort is spent to push the physical boundaries, e.g., by increasing modulation rates (up to 4096-QAM), by introducing massive (distributed) Multiple-Input Multiple-Output (MIMO), moving to the higher frequency bands (60 GHz and beyond), extremely small cells. However, just throwing in more capacity when targeting higher throughout applications, e.g., ultra high resolution (8K) and multi-view video needed for AR/VR applications, does not solve the strict E2E reliability and latency requirements for most other consumer and professional applications that do not require such ultrahigh data rates and can also rely on less complex and less costly wireless infrastructures and lower (sub-6 GHz) spectrum bands.

Today, wireless networks are operating as black boxes towards the applications. Current QoS mechanisms, defined at the network level, e.g., Type of Service (TOS) classes in IP networks, Access Categories (AC) in Wi-Fi, or network slices in 5G based on high-level Service Level Agreements (SLAs), only offer aggregated network services and still cannot guarantee E2E QoS demands by individual applications. Until today, there exist no scalable mechanisms for an individual application to express its individual application-level QoS desires to the network. In our vision, we believe that:

- applications must become an integral part of network management to produce the appropriate E2E network configurations based on the expression of the applications' requirements (also called mandates);
- networks should become more measurable, giving applications insights into how their data is treated and performance guarantees are met;
- end-users' preferences and consent have to be reflected automatically in the way the networks are configured, controlled, and managed, so applications degrade gracefully in decreasing or limited radio coverage and battery depletion.

Radically new network protocol designs are needed that allow (i) to extend the notion of management, control and data plane separation to the application level, (ii) applications to make their mandates known to the networking stack not only in the originating node, but across the entire E2E journey of a packet to the destination node, and (iii) include measurability as an inherent part of every data transfer.

In generating the mandates, applications can take into consideration of the profiles. For instance, based on user preferences and the device capabilities, e.g., camera resolution, processing capacity, etc., and the available network information, a conference call application can compose which end-to-end treatment it requires regarding continuous throughput, latency, jitter, etc. Once active, in-band telemetry can be exploited to continuously follow-up on whether QoS is met, across different network segments [5]. In case of context changes, adaptations can be made.

Hence, applications must exchange key information to produce better network configurations and applications adaptations based on profile, preferences and context information. This assumes that users, devices, network and applications

should reveal the minimal required information available at each level without compromising or conflicting with stakeholders' interests by respecting their nondisclosure requirements.

D. New Terminal Designs

End devices like user equipments (UEs) should conceptually be an integral part of the overall 6G system and offer reconfigurability as predicted in [6]. This should facilitate the following innovations:

- Bring the terminal into the design process, with more open air interfaces to the terminal/ end-device, reprogrammability, and a radio application store in which one can just download the next air interface innovation.
- Radios should be more flexible, software definable, compared with nowadays where each radio technology has a specific radio interface and transceiver design.
- Personalized, context-aware and graceful degradation of terminal (radio and associated applications), due to spectrum or any other computing resources, e.g., memory, processing or battery resources.

E. End-to-End Slice Management and Orchestration Involving Multiple Segments and Domains

Our proposed mandate-driven network ecosystem involves the combination of multiple network segments and administrative domains for realizing the E2E communication between humans and/or devices. However, each network segment and administrative domain may possess different purposes, e.g., mobile coverage, point-to-point connection, and indoor offloading; and may be subject to different constraints, e.g., resources, SLAs, and policies. As a consequence, the composition of E2E services across such diverse infrastructure components easily becomes overly complex or even intractable. Therefore, we must adopt a decomposition of E2E network management, allowing each separate network segment in every administrative domain to be independently managed by a separate specialized orchestrator, which reduces the overall complexity and simplifies the design of resource allocation directives [3].

Our paradigm shift encompasses the introduction of a hyperstrator, either centralized or distributed, for leveraging and coordinating multiple specialized orchestrators to deploy E2E services [3], as shown in Figure 2. Each specialized orchestrator can be designed by domain experts to include Artificial Intelligence (AI) and Machine Learning (ML) solutions to gain deeper insights in the operation of its respective network segment, enabling optimal configuration and resource allocation, assisted by the novel monitoring and auditing capabilities. The hyperstrator ensures a cohesive allocation of resources across network segments for guaranteeing a consistent E2E QoS for the E2E services. This functionality requires aligning network functions and services descriptors (e.g., tenant slices, service function chains, radio and network functions) proposed by various communities to ensure interoperability and cooperation of multiple orchestrators. In addition, there is the challenge
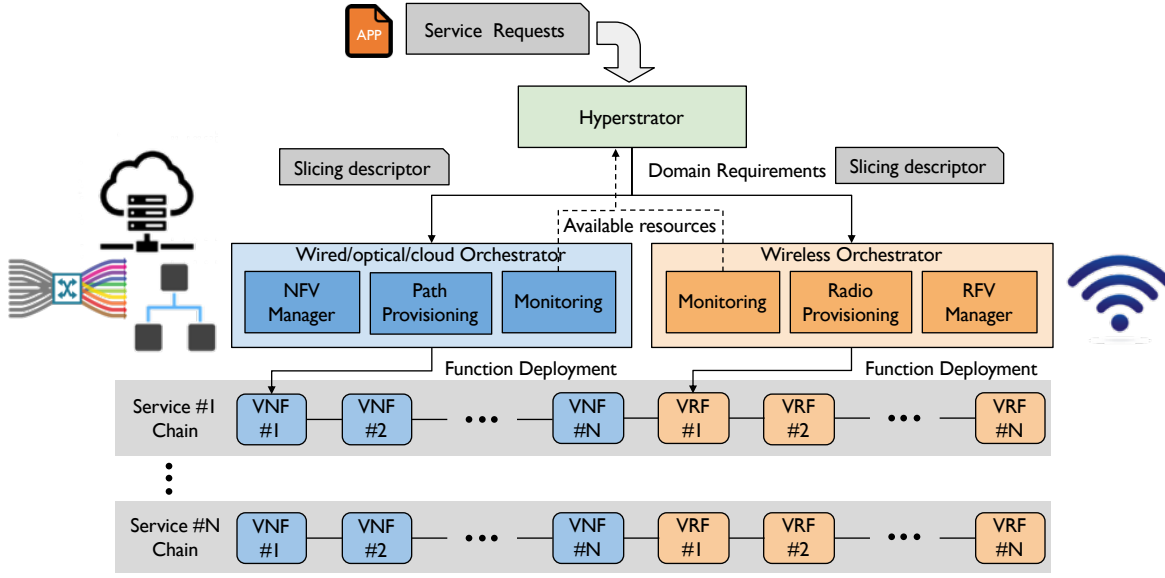
Fig. 2. Decomposition of the E2E network management using a hyperstrator and specialized orchestrators for managing different network segments.

of operating in the presence of heterogeneous interfaces be them east-west or north-south. These interfaces should become an integral part of the service composition and should be seen as service components involved in the overall process. The introduction of generic and semantically-rich interfaces for description and discovery of services, components, and resources, can enable a more agile and automated composition and deployment using knowledge-based reasoning and inference.

Realizing effective cooperation between network segments and administrative domains towards a global E2E efficiency is an interoperability and compatibility challenge, which must be specifically addressed in future open and cooperative systems. In addition, there is also the design of a language that can express the both the diverse set of assets, e.g., network resources and components, and constraints, e.g., policies and QoS requirements, while sharing the minimal amount of data. For instance, at the highest orchestration level, the hyperstrator, does not need be aware about how individual wireless resource blocks are allocated, however, it must know that available wireless link capacity, throughput and reliability. Harnessing the intelligence in the different levels, e.g., device, network segment, and E2E network is a major concern to design a cohesive, stable, reliable and resilient system.

F. Context and Purpose Aware Composition and Deployment of Services

In order to achieve maximum flexibility and efficiency, both devices and the network infrastructure must cooperate for selecting the best technology and configuration, e.g., air interfaces (Visible Light Communication (VLC), WiFi, Device-to-Device (D2D) and mobile), spectrum bands (licensed or unlicensed spectrum), and bearers (wireless or wired), and

network segment, based on the context and purpose of the deployment. The context may include relevant information beyond the GUP, e.g., location, environment and speed of the user, but also the requirements of the desired services. Based on the context, the mandate-driven networking ecosystem should infer the necessary network resources and functionality for meeting the services' requirements, and autonomously select the most appropriate technologies and configuration support these services and applications. With such capabilities integrated into the 6G network design, we can reduce or eliminate need for human interaction for configuring and optimizing the network infrastructure to support new types of services. In addition, it also fosters innovation, creating space for SMEs, industry and start-ups to develop better automation and decision tools for achieving higher efficiency and performance.

G. Machine Learning and Artificial Intelligence for Addressing System Complexity

The surge in the constant changing and dynamic relationships of various components, e.g., node, links, flows, and functions, makes it difficult to optimally configure the wireless networks. Traditionally, it is done by network experts that use their extensive knowledge of network topology, mobility, coverage, etc., in devising policies for configuring the network. In 6G, these network elements and their dynamic behaviour grow tremendously along with a huge number of network configuration options can lead to varying impact on different performance objectives. Such dynamicity and extreme level of configurability become a difficult challenge for network experts unless AI and ML are introduced to assist the experts in configuring and managing the networks in an optimal and autonomous manner. In addition, the wireless networks are

becoming highly distributed and it becomes essential to take decisions locally. This is because centralized management poses formidable complexity in the network due to increased information exchange overhead. ML or AI can be used for enabling intelligence locally and in real-time, leverages on the monitoring information from in-band telemetry. For instance, [7] uses a non-cooperative game for performing resource allocation in a decentralized way in ultra-dense small cell networks, where game theory is essential to enable AI in multi-agent environments. In such multi-agent environments, a deep reinforcement learning technique can be employed in which an agent (which can be a node, a base station, or an access point) can take network configuration decisions while meeting certain design objectives locally without having any interaction with a centralized controller. The latter can nevertheless, adjust and steer, the local decisions, though the injection of new policies, rules and eventually drive in a non compelling manner the distributed, local and autonomous agents (to enhance learning and the related exploration and exploitation processes).

Decentralized learning has marked advantages over centralized such as spectrum efficiency, privacy, low latency, high scalability, energy consumption, etc. Decentralized learning leverages to maintain the privacy of users by not sending raw information to the edge. We call this decentralized learning as *on-device learning* and it becomes feasible now because modern smart phones and sensors are equipped with high end computational units such as A11 bionic chip in IPhone X [8], Snapdragon neural processors (e.g., Snapdragon 855 [9]), and Raspberry Pi 4 [10], which paves the way towards on-device learning. In order to realize decentralized learning, a new ML paradigm has emerged named as Federated learning (FL) – a concept introduced by Google in 2016 [11]. It leverages decoupling of data processing and computation at the central point and uses on-power device capabilities for model training from locally acquired datasets. FL is in its infancy and a lot fundamental challenges need to be solved including different sizes and distribution datasets, issues in model weight sharing, scheduling, on-device constraints, communication channel impairments, etc.

IV. CONCLUSION

The growing complexity of current and future networks calls for the direct involvement of tenants and end-users in the overall design of E2E communication. Because E2E solutions driven by one specific community moving every functionality towards cloud infrastructures, when actually many decisions can be taken locally, limit overall system efficiency and slow down the introduction of innovations by small players. Based on this analysis and observation, we propose to broaden research and innovation from the current telecom-centric view with limited actors, to a E2E mandate and purpose-driven view steered by the needs of professional markets also targeting private networks involving multiple actors and academia. Further, the proposed paradigm shift is technology and actor neutral and can be implemented independently of any specific technology generation and business relationships.

REFERENCES

[1] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5g mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[2] M. Iwamura, "Ngmn view on 5g architecture," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. IEEE, 2015, pp. 1–5.

[3] J. F. Santos, J. van de Belt, W. Liu, V. Kotzsch, G. Fettweis, I. Seskar, S. Pollin, I. Moerman, L. A. DaSilva, and J. Marquez-Barja, "Orchestration next-generation services through end-to-end networks slicing," 2018.

[4] E. Mansour, A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, and T. Berners-Lee, "A demonstration of the solid platform for social web applications," in *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 223–226.

[5] J. Haxhibeqiri, I. Moerman, and J. Hoebeke, "Low overhead, fine-grained end-to-end monitoring of wireless networks using in-band telemetry," in *CNSM2019, the 15th International Conference on Network and Service Management*, 2019, pp. 1–5.

[6] "Intel and softbank beware. open source is coming to the chip business," https://www-bloomberg-com.cdn.ampproject.org/c/s/www.bloomberg.com/amp/news/articles/2020-01-22/open-source-transformed-software-the-chip-industry-is-next, accessed: 2020-02-13.

[7] A. Shahid, V. Maglogiannis, I. Ahmed, K. S. Kim, E. De Poorter, and I. Moerman, "Energy-efficient resource allocation for ultra-dense licensed and unlicensed dual-access small cell networks," *IEEE Transactions on Mobile Computing*, 2019.

[8] Apple Inc., "The future is here: iphone x," Available at https://www.apple.com/newsroom/2017/09/the-future-is-here-iphone-x/.

[9] Qualcomm, "Snapdragon 855 mobile platform," Available at https://www.qualcomm.com/products/snapdragon-855-mobile-platform.

[10] Raspberry Pi 4, Available at https://www.raspberrypi.org/products/raspberry-pi-4-model-b/.

[11] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan *et al.*, "Towards federated learning at scale: System design," *arXiv preprint arXiv:1902.01046*, 2019.