On the efficient architecture for 6G System (draft)

Lechosław Tomaszewski* [0000-0003-3836-7900] and Robert Kołakowski*◇ [0000-0002-5451-8847]
*Orange Polska, ◇Warsaw University of Technology

Abstract

This paper discusses the proposed approaches to the 6G System architecture against the background of the fundamental principles of the 5G network and challenges towards 6G network, focusing also on drawbacks of already adopted solutions and their further impact on the future 6G System performance. The user-centric architectural framework, based on a dynamic stateless procedural approach, extending the organic core concept, has been proposed that simplifies and accelerates Control Plane interactions, introducing also the mechanisms of network self-cognition, self-awareness, and self-control into the new core.

Index Terms

Architecture, 5G, 6G, User Plane, Control Plane, Application Plane, Signalling, User-centric, Cell-free, AI, NTN, TN, Network of networks, Programmability, Lightweight, Stateless, Front-end, Procedure-oriented core, Vertical stitching, Horizontal stitching

ACRONYMS

The following acronyms are used in this manuscript:

3GPP	3rd Generation Partnership Project
4G	4th Generation
5G	5th Generation
5GS	5G System
6G	6th Generation
6GS	6G System
AF	Application Function
AI	Artificial Intelligence
API	Application Programming Interface
CN	Core Network
CNF	Cloud-native Network Function
CP	Control Plane
CU	Centralised Unit
CUPS	Control and User Plane Separation
DevOps	Development and Operations
DevSecOps	Development, Security and Operations
DT	Digital Twin
E2E	End-to-End
eMBB	Enhanced Mobile Broadband
ETSI	European Telecommunications Standards Institute
gRPC	Google Remote Procedure Call
GTP	GPRS Tunneling Protocol
ISAC	Integrated Sensing and Communication
IT	Information Technology
ITU	International Telecommunication Union
KPI	Key Performance Indicator
LEO	Low Earth Orbit
LTE	“Long Term Evolution”
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MIMO	Multiple Input Multiple Output
ML	Machine Learning
mMTC	Massive Machine Type Communications
MNO	Mobile Network Operator
NAS	Non-Access Stratum
NBI	NorthBound Interface
NEF	Network Exposure Function
NF	Network Function
NFV	Network Function Virtualisation
NSA	Non-StandAlone
NSI	Network Slice Instance
NTN	Non-Terrestrial Network
O-RAN	Open RAN
QoS	Quality of Service
RAN	Radio Access Network
RIS	Reconfigurable Intelligent Surface
SA	Stand-Alone
SBA	Service-Based Architecture
SBI	Service-Based Interface
SCP	Service Communication Proxy
SDN	Software-Defined Network
SLA	Service Level Agreement
SotA	“state of the art”
TTM	Time to Market
UE	User Equipment
UP	User Plane
UPF	User Plane Function
URLLC	Ultra-Reliable Low-Latency Communication
VNF	Virtual Network Function

ETHER project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union’s Horizon Europe research and innovation programme under Grant Agreement No. 101096526. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

I. INTRODUCTION

It has become an industry custom to change the generation of mobile systems every 10 years. The 3rd Generation Partnership Project (3GPP) 5G System (5GS), also referred to as International Telecommunication Union (ITU) IMT-2020, has been defined to phase out the 4th Generation (4G) – known under the code name "Long Term Evolution" (LTE) (*sic!*). The standardisation of currently latest 5GS is still under development (within the recently launched 3GPP Release 19 and also planned in the Release 20 to be finished tentatively by the end of 2027, while further 5GS evolution in later releases is foreseen [1]). The operational 5GS's capabilities as yet significantly deviate from the enthusiastic visions of 2016-2019, primarily because about 80% of implementations worldwide are still 5GS Non-StandAlone (NSA) [2] – in fact, LTE with "radio on steroids" providing only the Enhanced Mobile Broadband (eMBB) service and lacking the functionalities crucial for the disruptive 5GS capabilities, *e.g.*, network slicing. It would, therefore, be difficult to say that 5GS has entered the "plateau of productivity" phase according to the well-known Gartner Hype Cycle methodology [3]. And yet, since 2018 [4] at least, hence two years before the premiere of the commercial 5GS (*sic!*), initiated are speculations and visions about what 6G System (6GS) would be.

Currently, there is general consensus in the scientific and industrial communities about the expectations and vision of applications, use cases and capabilities of 6GS. ITU has recently published the document on the framework and overall objectives of IMT-2030 [5] defining new usage scenarios, capabilities, and their Key Performance Indicator (KPI) targets for research and investigation. In the 3GPP roadmap, studies about 6GS are tentatively planned in the Release 20, and the first wave of 6GS standardisation is going to start with the Release 21 [1]. The 6GS applications include such futuristic ones as haptic or avatar communication; its capabilities include integrated communication, sensing, positioning, and computing with ubiquitous, immersive, and native Artificial Intelligence (AI), entailing the architecture much more complex than ever. The question, therefore, arises whether the 6GS architecture will be able to bear not only the burden of 5GS backlog and new expectations but, above all, its own weight!

This paper presents a novel 6GS architectural vision, which leverages service-based architecture and procedural approach to improve system flexibility, modularity and scalability while decreasing the complexity of operations. First, outlined is the key 5GS design principles (cf. Section II) and arising demands and challenges that 6GS is envisioned to target. Section III presents the emerging concepts regarding 6GS architecture development and improvement mechanisms. Their applicability and rationale in the context of carrier-grade networks is then discussed (Section IV). Based on the analysis, a novel 6GS framework (Section V) is proposed that meets the emerging mobile network challenges and addresses the key deficiencies of the "state of the art" approaches. Finally, Section VI summarises and concludes the paper.

II. MOBILE NETWORK – STATE OF THE ART AND NEW DEMANDS

5GS and its architecture [6] has been founded on the following principles, making a breakup with the legacy of previous mobile network generations:

P1 Partial withdrawal from the "All IP" principle – separation of services from access to the network, *i.e.*, mobile network to be a wireless "last mile" with access to the IP network (public or private); user communication over Non-Access Stratum (NAS) is possible to support extremely low energy consumption User Equipment (UE) devices with no User Plane (UP) implemented.

P2 Network softwarisation – decoupling of network elements' function now implemented in software from underlying hardware and simultaneous transition to commodity hardware used in Information Technology (IT); this approach applies to both Radio Access Network (RAN) and Core Network (CN) layers.

P3 Virtualisation – decoupling of software and hardware life cycles through virtualisation layer where virtualised compute, memory, and storage resources are exposed to the software; virtualisation technology inherently supports flexible vertical and horizontal scaling; the 3GPP vision of 5GS management architecture is complementary with European Telecommunications Standards Institute (ETSI) Network Function Virtualisation (NFV) Management and Orchestration (MANO) [7].

P4 De-monolithisation of network architecture entities (in conjunction with the former two) – while LTE architecture was composed of 7 types of hardware-based nodes, 5GS CN architecture is based on portable softwarised Network Functions (NFs) of much higher granularity; additionally, Control and User Plane Separation (CUPS) has been implemented in CN enabling functional and spatial decoupling of both planes and hence their placement more optimised in terms of user and signalling traffic distribution.

P5 Non-monolithic approach to softwarised CN Control Plane (CP) through its implementation in the Service-Based Architecture (SBA) model [8]; NFs expose their services and consume others' services through REST Application Programming Interfaces (APIs) utilising JSON-encoded HTTP/2 interactions – direct or via Service Communication Proxy (SCP) [6]; the CN signalling takes place via a logical bus implemented by the message broker service; additionally, a service discovery mechanism is defined to be used by SBA-native CP NFs; however, the signalling exchange between RAN and CN (N2 interface) as well as between UE and CN (N1 interface via NAS proxied by RAN) still takes place through the mobility management entity acting as a signalling termination point and coordinator of procedures within CN.

P6 Mobile Network capabilities exposure to external systems – other architectural frameworks like ETSI Multi-access Edge Computing (MEC) or 3rd party systems, *e.g.*, vertical industries – as CP mechanisms availability via external interface of special Network Exposure Function (NEF) or mechanism of Application Function (AF) embedded inside CP and acting as an external system "embassy"; 3GPP has standardised the frameworks of common NorthBound Interface (NBI) API [9] and service enabler layer for verticals [10].

P7 Unlike the previously uniform UP architecture – unable to simultaneously meet the divergent needs of services with different specificities, 5GS in its Stand-Alone (SA) implementation introduces network slicing – free shaping of User Plane Function (UPF) according to the use case; UE applications can request access to specific Network Slice Instance (NSI), *i.e.*, UPF chain.

To sum up: 5th Generation (5G) has brought an in-software communication network, the basic architectural framework of which is composed of more than 40 different CN NFs of high granularity, exposing software services (from 1 to even 36 depending on NF type) [6]. There are more than 500 5GS procedures involving from two to even a dozen NFs composed by even several dozens of point-to-point NF interactions [11], [12], [13]. The use of an approach appropriate for the development of information systems facilitated the development of the software itself and its life cycle management. At the same time, it sharply increased the number of instances of architectural elements, signaling interfaces and exchanges that are subject to monitoring and analysis.

However, on the borderland of generations, the mobile system architecture faces new trends and challenges that drive its further development. The fundamental pressure is the industry-wide trend to migrate towards cloud-native solutions – whether they are placed in a public (at hyperscalers), private or hybrid cloud, it will be determined by the business-driven choice of Mobile Network Operators (MNOs). It is followed by the evolution of virtualisation technology away from virtual machines and Virtual Network Functions (VNFs) towards much lighter containers and Cloud-native Network Functions (CNFs), which has also been addressed by the commonly industrially recognised ETSI NFV framework in its 4th release [14]. However, the greatest challenges come from visions of the future 6GS [5], including:

C1 AI natively built into the mobile network and working for CP, UP and the Management Plane – applying cognitive mechanisms from hardware up to application layers.

C2 New set of usage scenarios: Immersive Communication, Massive Communication, Hyper Reliable and Low-Latency Communication – scenarios extending respectively: eMBB, Massive Machine Type Communications (mMTC), and Ultra-Reliable Low-Latency Communication (URLLC) of IMT-2020/5GS – and new ones, Integrated Sensing and Communication (ISAC) – objects detection and tracking/environment monitoring/motion monitoring including gestures [15], AI and Communication, and Ubiquitous Connectivity; for the latter, assuming 100% coverage of the Earth – unified terrestrial and non-terrestrial network based on ground, aerial, and satellite base stations, to be additionally supported by Reconfigurable Intelligent Surfaces (RISs) [16] acting as beam reflectors, both ground ones (*e.g.*, installed on facades of buildings) and located on aerial objects (*e.g.*, drones).

C3 Extremely dynamic network topologies, in which mobility of network nodes (*e.g.*, base stations in non-geostationary orbit satellites) will be much higher than UEs mobility (27,000 km/s satellite tangential velocity versus 1,000 km/h velocity of an aircraft). The speed, together with orbit parameters (altitude, inclination, *etc.*), limit the visibility window to several minutes, *e.g.*, Low Earth Orbit (LEO) satellites at 1,000 km altitude and 60° inclination are visible for less than 15 min [17].

C4 Fusion of physical and cyber worlds through support of Digital Twin (DT), including also network DT.

C5 Evolution of mobile network through integration of computing, AI, communication, and sensing within the network towards an application enablement platform with easy End-to-End (E2E) integration, exposing "XaaS" capabilities (communication, sensing, AI, compute, DT, *etc.* as a service) for shorter Time to Market (TTM).

C6 Extended positioning capability – cm-level, supporting also other capabilities and mechanisms (*e.g.*, sensing, RIS, beamforming, *etc.*).

C7 Network of networks – 6GS is envisioned to interconnect and integrate various types of networks, including the aspects of technology, network segment (terrestrial, non-terrestrial, transport, *etc.*), and ownership (*i.e.*, multi-provider ecosystem), to establish a comprehensive, seamless and optimised communication environment; the expected dynamicity of network composition poses several challenges in terms of complexity, unification and simplification of network interfaces and protocols, network modularisation or MANO support (*e.g.*, dynamic network attachment/detachment procedures).

C8 Reduction of heterogeneity and complexity, also lowering the costs.

C9 The above pose extended technological challenges related to privacy, security, trustworthiness, and safety as well as energy and spectrum (including the cognition of electromagnetic environment) efficiency of 6GS. In the case of energy efficiency, the wider context of sustainability is given, which also includes less extensive and more efficient hardware, modulation and coding, processing algorithms, software infrastructure, traffic load shaping and balancing, *etc.*

III. APPROACHES TO 6G ARCHITECTURE

While pending standardisation studies developing proposals for the architecture to be adopted by the future 6GS, numerous approaches have been proposed towards the expected development and improvement of the mobile architecture.

A1 Cell-free RAN with distributed massive Multiple Input Multiple Output (MIMO) antenna systems – with dense deployment of base station antennas, it is possible to shape the individual UE beam with a distributed antenna system based on clusters of multiple base stations [16], [18], [19]. This way, the traditional cell-based approach (base stations serving specific areas) will be abandoned, enabling suppression of inter-UE interferences and elimination of signal degradation at cell boundaries. Implementation of this concept needs coordinated processing of signal for all antennas within the cluster, which can be performed within the Centralised Unit (CU) of disaggregated RAN (assuming one CU per cluster). For inter-cluster coordination (*i.e.*, avoiding problems at cluster boundaries), inter-CU coordination would be needed. In the case of RIS technology application, forming of the dedicated UE beam will need active and passive elements coordination (antennas and reflectors).

A2 User-centric RAN – while the placement of disaggregated RAN entities (closer to the edge or core) with the 5GS advancement is going to be flexibly adapted according to service class specificity, it is proposed in 6GS to place the individual per-UE radio processing unit according to the demand type (*e.g.*, low latency – at the edge, high capacity – close to CN) [16], [20].

A3 Individual mobile networks per UE – as a further step from A2, it is proposed to implement an individual network for each UE to provide higher network operation reliability and agility – as the failure impact is limited to a single UE, instant restoration by a simple restart of an individual network instance, avoiding perturbation of others is possible, instead of time-consuming root cause analysis and complicated procedures of restoration (smooth functional instances off-loading, traffic redirection, measures for protection against failure escalation and signalling storm, *etc.*) [19], [20].

A4 Programmability – the postulate of programmability of networks and services is implemented at various levels – by SBA, network slicing [6] the AF mechanism in CN, the xApps and rApps mechanism embedded in controllers of the Open RAN (O-RAN) framework that can also be applied on a per-slice level [21]. The UP full programmability can be provided by incorporation of Software-Defined Network (SDN). Phasing out the GPRS Tunneling Protocol (GTP) will also improve network UP reconfiguration speed, reduce data overhead and simplify UP management. Such implementation, however, is still an open question [22]. Moreover, over-the-air programmability of UE by the network is proposed to provide add-on functionalities [18].

A5 "Parallelisation" of signalling flows – while current mobile systems' procedures are organised as a linear flow of steps performed by two or more functional entities, it is proposed to enable branching signalling into parallel requests in order to optimise the overall signalling speed [18]. Additionally, it is proposed to begin the UP communication simultaneously with CP signalling to reduce service availability delays [16].

A6 Further disaggregation of 5GS NFs into separate cloud-enabled microservices providing distinct implementation of current standardised services of these NFs (cf. services *Nabc_ServiceName* in 3GPP TS series 29.5*nn*, where *abc* is a 5GS NF acronym, and *n* is a decimal digit) [23].

A7 Extension of CN SBA to RAN is proposed for replacing the N2 interface [24]. It might be used for registration and de-registration of the RAN instances in the common network functions repository, RAN services discovery and authorisation within CP, and direct CP NFs interaction with RAN services.

A8 Distributed NAS termination in CP – abandoning of UE-CN signalling termination at the mobility management entity is proposed to enable direct CP interaction of UE with CN NFs, including CP services discovery [25].

A9 Signaling load mitigation by the dynamic placement of NFs and providing their direct non-SBA interaction – as high-granularity functional split entails an increase in signaling traffic and processing effort to handle web protocols (*e.g.*, HTTP messages parsing), it is proposed to dynamically replace the SBA messaging through internal messaging within "components" – groups of NFs mutually communicating via shared memory; such group would maintain the SBA interfaces towards the CP messaging bus [19].

A10 Front-end/back-end model-based approach to CN CP – it is proposed to access the core functionality of CP (*i.e.*, its back-end) via front-ends adapted to interconnected party specificity (specific front-ends toward UE, RAN, CP NBI, *etc.*). The front-end would also play the role of a load-balancer for the distributed back-end. Additionally, a translation to Google Remote Procedure Call (gRPC) interface may be used [26] for internal CN CP exchange to provide very low latency due to the binary serialisation of messages.

A11 Lightweight NF instances – instead of monolithic software functions implementing all functionalities regardless the use case, it is proposed to provide their further dividing into separately "switchable" elements (selectable sub-functions), which will enable quicker life cycle control, fast distribution and startup/shutdown processes, due to their acceleration of by smaller software instances [19]. This approach may be applied both to CP and UP through vertical stitching [27] in which the functional expansion of software instances is done through gradual piggybacking of upgrades on lightweight basic functions.

A12 Stateless NFs – in this approach, decoupling of data from computation is proposed; each CP NF, when processing a request associated with a specific procedure in the context of a specific UE, queries this UE state in the unified UE state storage and returns it to the storage if it has been changed due to the processing [19], [28]. In the case of NF failure, its recovery is very simple. As a result, NFs' "cloudification" as well as higher resiliency and scalability of the network are provided at the cost of longer procedure delays due to the involvement of the unified storage at every step of the signalling procedure.

A13 Procedure-oriented CN architecture – the 5GS CN is here rearranged by regrouping the functionalities (services of the individual 5GS CP NFs) into new E2E services implementing 5GS procedures [29]. While individual procedures can duplicate the same fragments of code (*i.e.*, certain functionalities), combined with statelessness, the total gain on computing resources is around 50%. The approach enables a reduction of at least 40% in CP exchange and up to 50% faster execution of procedures, especially more complex ones.

A14 Organic network [30] is the comprehensive approach that also includes other concepts (A7, A8, A10, A12, A13). CN is organised as a set of workers where stateless E2E procedures are launched on demand of front-ends. These latter implement the reduced 3GPP SCP [6] functionality, select appropriate procedure services and workers to be called based on the incoming requests, and are acting as CN border gateways providing interaction with entities external to CN CP: RAN nodes, UEs, UPFs, other CN CPs (*e.g.*, for roaming support), with necessary protocols translation and load balancing. The procedures are organised as 4 groups of services: (i) Access Control, Authentication and Authorisation Service, (ii) Connectivity Management Service, (iii) Mobility Management Service, and (iv) Session Management Service. A special Subscriber State Service, *i.e.*, a UE state store containing static/dynamic subscription profile data, policy/charging data and dynamic state (session, mobility, *etc.*), is defined to support the statelessness of other services. The inner communication of the organic core network is provided according to the SBA principle. For the most complex 5GS procedures (*e.g.*, registration), the reduction of signalling messages from 42 to 8, and the reduction of state fetching and updating (for statelessness handling) from 44 to 4 is even reported. However, it should be noted that the organic core is purely "reactive", *i.e.*, responds to external requests. Additionally, being user-centric, it lacks the network perspective (internal network processes and, above all, network self-awareness).

IV. DISCUSSION

The presented approaches focus on various and different aspects and challenges. The cross-analysis of approaches, architectural principles, and challenges described above is provided in Tab. I. The more indications in the second column, the greater the impact of a given architectural approach. The approaches can be clustered into two groups related to user-centricity (A1–A4) and the demand of CP efficiency to respond to new challenges (A5–A14). While the first group can be interpreted as a new paradigm, the second group is a set of attempts to deal with the consequences of the architectural choices made when defining 5GS.

TABLE I
CROSS-INTERACTIONS BETWEEN PRINCIPLES, CHALLENGES, AND APPROACHES.

Approach	Related principles and challenges	Conflicting features
A1	C2, C3, C5, C6, C7	C8
A2	P2, P3, P4, C2, C3, C5, C6, C7	C8
A3	P2, P3, P4, C2, C3, C7, C9	C8
A4	P2, P3, P5, C2, C3, C4	
A5	P4, P5, C2	C7, C8, A12⊕
A6	P3, P4, P5, C1	C8, A9, A12⊙, A13⊗, A14⊗
A7	P5, C3, C4, C6, C8	
A8	P1, P5	C8
A9	P2, P3, C3	P5, C8, A6
A10	P2, P3, P5, P6, C3, C5, C7	
A11	P2, P3, P7, C1, C2, C3, C8	
A12	P2, P3, P4, C5	A5⊕, A6⊗
A13	P2, P3, P4, C5, C8	A6⊗
A14	P2, P3, P4, P5, P6, C1, C3, C5, C7, C8	A6⊗

⊕ Potential UE state conflicts due to affecting it in parallel at the same time.
⊙ Risk of signalling latency significant growth due to multiplication of interactions with UE state storage.
⊗ By a principle, the procedural and NF chain approaches are antagonistic.

An issue of particular importance is the consequences of adopting to mobile networks the IT frameworks that have proven successful in implementing web applications operating in a highly distributed model and based on cloud technologies. The deep disaggregation of the functional architecture of the mobile network implemented in the software model resulted in a sharp increase in the complexity of the network, the number of interactions between increasingly smaller functional blocks, and thus, an increased impact of the used protocols on the time duration of procedures in 5GS (*i.e.*, due to the CP-induced delays). It should be noted that web systems operate on a different time scale than communication networks, so the overhead associated with individual API handling (*e.g.*, the order of 12 ms [31] under light-load) has fundamental consequences for the dynamics of network control, especially if additional delays related to virtualisation or containerisation of individual functions are concerned [32], [33]. Therefore, with the finer granularity of the functions (A6), the signaling latency will scale up. The microservice-based approach will also have a catastrophic impact on the CP traffic increase, which may require additional infrastructural investments to mitigate congestion and potential additional latency. The related increased complexity can be

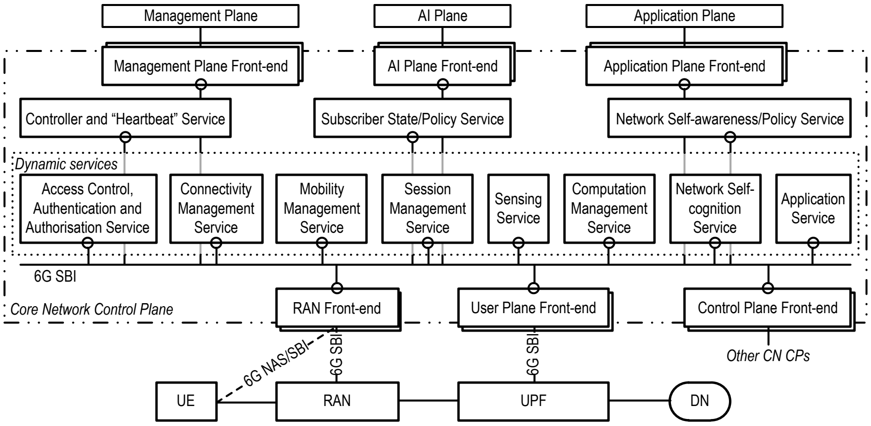

Fig. 1. High-level 6GS framework with focus on CN CP

another nail in the coffin of high-performance 6GS. The large number of services that can concurrently use/modify user-related data raises the issue of data integrity maintenance and requires governance across services to synchronise the operations. Moreover, an increased number of services will require a specific approach to management and orchestration to: i) ensure the proper versioning of NFs and independent upgradeability; ii) establish testing and verification environment for different NFs compositions (both versions and NFs types) to ensure compatibility and proper operation. Today, however, there is no mature and unified approach to Development and Operations (DevOps), let alone the integration with ETSI NFV MANO, the inclusion of security features, *i.e.*, Development, Security and Operations (DevSecOps), or AI-specific requirements into the MANO processes. Nonetheless, some of the cloud-native and microservice principles are vital to establishing a modular, scalable, and flexible 6GS due to, *i.a.*, provided development agility, easier code base management, technological flexibility, simplified root-cause analysis and network-wide scalability. While some conservative approaches (A5, A9) can provide a more or less efficient palliative response, the extensions of SBA (A7, A8) using current protocols may increase the perimeter of the problem. Disruptive approaches – efficient protocols (A10) or application of E2E network procedures processing (A13 and derivative approaches) – seem to provide a significant response to the problem.

Finally, the aspect of 6GS integration with DT is not well addressed yet. While each vision will have an impact on the interactions with DT, their scope and exchanged data are vague, potentially limiting the DT models' consistency with real network conditions and decreasing the performance and usability of AI models due to increased risk of Service Level Agreement (SLA) violations (due to wrong AI model understanding of the optimised network process).

In conclusion, there is a need to define the 6GS architecture that will: i) leverage the beneficial properties of cloud-native and microservice-based approach; ii) take into account their impact on the mobile CP and diminish their drawbacks; iii) establish an open, scalable and flexible ecosystem to facilitate the integration of external components (*e.g.*, DT frameworks) or business entities to allow the implementation of the most advanced AI-driven 6GS use cases; iv) exploit protocols adapted to the dynamicity of 6th Generation (6G) network.

V. CONVERGED 6GS ARCHITECTURE VISION

The following section presents the high-level vision of 6GS architecture. The architecture outlined below is based on the following fundamental assumptions:

- Composing the overall architecture of the User, Control, Application (including verticals' environment), Management, and AI Planes – all mutually interacting via Service-Based Interfaces (SBIs); the AI Plane exposes its services to other planes (cf. C1). The support of network DT (cf. C4) will be provided by exposure and interaction of all planes' services.
- Integration of communication, computing, AI, and sensing into one global architecture (cf. C5).
- Simplification of architecture, reduction of signalling interactions and CP latency, avoidance of excessive fragmentation and futile overheads (cf. C8).
- Support of user-centricity of CP and UP including individual network, network slicing, lightweightness, and programmability (cf. P7, A3, A4, A11).
- Adoption of organic core with underlying or supported concepts (cf. A7, A8, A10, A12, A13, A14), particularly CP SBA span extension to RAN, UE, and UPF; its further enrichment and supplementation with missing features.

The proposed architecture is presented in Figure 1. In its centre, there is CN CP composed of inner and outer parts. The inner core comprises dynamic stateless services to be launched by the outer core, implementing E2E (generally real-time) control procedures. In addition to 4 basic organic core services (Access Control, Authentication and Authorisation, Connectivity Management, Mobility Management, and Session Management) [30], the following ones are included:

- Sensing Service – the service responsible for integrated in-network sensing capability [5] including UE positioning.

- Computation Management Service – the service responsible for integrated in-network computing capability (*e.g.*, MEC-like, which is no longer external to the network but natively built-in). In particular, this service will be responsible for seamlessly ensuring computation mobility in coordination with user data path mobility (application and UE handovers).
- Network Self-cognition Service – the service responsible for building knowledge about the network environment (topology, connected entities – *e.g.*, RAN or other CNs, data collection for charging or network analytics, mechanisms for discovering services available in SBA, *etc.*). In the context of integration with Non-Terrestrial Network (NTN), it will also be important to obtain information enabling the determination of future states of the environment, *e.g.*, cyclic topology changes related to the trajectories of satellites with RAN nodes on board.
- Application Service – the service representing both future extendability of CP to accommodate new network capabilities and CP programmability through embedding trusted control procedures needed by applications (cf. mechanism of Application Function in 5GS [6]).

The dynamic services (dynamically instantiated stateless software components to be loaded and executed upon request, then terminated after completing their tasks) cooperate with permanently active Subscriber State/Policy and Network Self-awareness/Policy services (the latter provides storage of network data/state/policies/environment) through fetching and/or updating both storages. As the dynamic services are called in the context of specific UEs, the framework is inherently user-centric. They are also able to consume services of the AI Plane via the relevant front-ends.

The essential job of the outer core is managing calls of control procedures provided by dynamic services. They can be requested externally via front-ends dedicated to specific components of the CN CP environment. The front-ends provide necessary protocols adaptation towards outside (*e.g.*, 5G SBI to legacy 5GS CN CP or other 6GS CNs), and towards inside they select workers to dynamically load the procedures for execution as well as perform workers load balancing. The role of the Controller and “Heartbeat” Service entity can be paralleled to the vegetative (autonomic) nervous system. The controller, operating permanently, launches dynamic services based on schedule, triggering conditions, or criteria contained in the UE (subscriber) state. This way, it can periodically refresh the UE state, *e.g.*, in the part related to location, azimuth and speed, which will mitigate the signalling exchange thanks to the replacement of multiple direct dynamic services execution from NBIs with the UE state exposure to support Management or Application Planes’ processes or context awareness mechanisms to be used, *e.g.*, for mobility management. The controller adopts a similar approach in terms of Network Self-awareness/Policy Service, *i.e.*, by monitoring of network-level data (*e.g.*, data aggregated for multiple UEs from different UPFs and services) can trigger execution of dynamic services in accordance with preset policies. While the Controller is a logically centralised component by design (to enable control of all the services deployed under its jurisdiction), the service itself can be scaled up to many parallel instances (*e.g.*, serving different geographical locations), as long as the operations are coordinated and synchronised across them. The latter is essential to preserve data integrity and synchronise operations on the network- and subscriber-related data as well as launching multiple services of the same type operating on the same contextual data. Such coordination can be achieved by interactions via CP Front-end. The interconnection of different CN CPs enables the straightforward implementation of C7 via horizontal stitching (leveraged by SBA mechanisms) of individual components from different CNs (as long as the protocol translation and synchronisation of procedures is maintained).

UPF can be implemented as user-centric and lightweight with user-specific customisations (cf. A3, A11). The interaction between CP entities and UPF instances occur via the UP Front-end, which – apart from user data transmission-specific support, *e.g.*, Quality of Service (QoS) management and enforcement, mobility handling, *etc.* – will require the specific extensions for in-network computing, *i.e.*, embedding additional user data processing inside UP (*e.g.*, analytics, filtering, packets inspection). The above imply the need for flexible data path manipulation. This feature can be achieved by incorporation of SDN as the underlay fabric for UPF. However, the network-wide exploitation of classical SDN is problematic due to scalability issues and might require necessary adaptation (*e.g.*, introduction of different SDN architectures).

VI. SUMMARY AND CONCLUSIONS

In this paper, presented are the key visions, trends and challenges for the emerging 6GS architecture. The major drawbacks related to the principles have been discussed, which included, among others, the adoption of web-based approaches to telco systems, network fragmentation into microservices or stateless architecture. The presented concepts will have a negative impact on network complexity (Management, Control, User and AI Planes) and performance (larger volume of CP traffic, increased time to realise CP procedures). Based on the analysis, proposed is the novel 6GS architecture extending the most promising organic network concept exploiting the procedure-based approach. The concept reshapes the 6GS CP from the conventional NF-based approach to modular self-contained dynamically launched services that implement specific network procedures. The original organic framework is extended by services providing in-network sensing and computation management as well as generic application service. Moreover, the CP components for network state cognition and maintenance are introduced. The role of the latter is to handle the network perspective, autonomously launch the dynamic services and provide coordination across different network segments (integrated via SBI and dedicated front-ends). Overall, the presented 6GS architecture improves network flexibility, openness (for external integration, *e.g.*, with DT) and scalability while simplifying CP interactions. Nonetheless, some open issues remain to be addressed, *e.g.*, the design of effective CP protocols, cooperation of 6GS procedures across several CNs, or development of advanced coordination mechanisms to maintain user and network data integrity and support stitching of several 6GS CN. These mechanisms, however, once developed, can be easily integrated into the proposed 6GS framework by leveraging SBI and exposed front-end mechanisms.

Acknowledgements: The authors wish to thank Dr. Marius Iulian Corici (Fraunhofer FOKUS) and Mr. Masayuki Kurata (KDDI Research) for inspiring presentations and discussions during the 11th FOKUS FUSECO Forum (Berlin, 2023).

REFERENCES

[1] M. Papaleo, "3GPP roadmap towards 6G," in *CEPT Workshop on 6G Mobile Communications*, Jun. 2023, Last accessed 16 Apr 2024. [Online]. Available: https://cept.org/files/129806/2%20Marco%20Papaleo%202023-06-6G-CEPT-Workshop.pdf

[2] Global Mobile Suppliers Association, "115 mobile operators now investing in 5G Standalone networks globally," Aug. 2023, Last accessed 16 Apr 2024. [Online]. Available: https://gsacom.com/press-release/115-operators-investing-in-5g-standalone/

[3] Gartner, "Gartner Hype Cycle," Last accessed 16 Apr 2024. [Online]. Available: https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

[4] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, 2018, doi: 10.1109/MVT.2018.2848498.

[5] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond," International Telecommunication Union – Radiocommunication Sector, Recommendation M.2160-0 (11/2023), Nov. 2023.

[6] 3GPP, "System architecture for the 5G System (5GS)," 3rd Generation Partnership Project, Technical Standard TS 23.501, ver. 18.5.0, Mar. 2024.

[7] ——, "Management and orchestration; Architecture framework," 3rd Generation Partnership Project, Technical Standard TS 28.533, ver. 18.1.0, Apr. 2024.

[8] ——, "5G System; Technical Realization of Service Based Architecture; Stage 3," 3rd Generation Partnership Project, Technical Standard TS 29.500, ver. 18.5.0, Mar. 2024.

[9] ——, "Common API Framework for 3GPP Northbound APIs," 3rd Generation Partnership Project, Technical Standard TS 23.222, ver. 19.1.0, Apr. 2024.

[10] ——, "Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows," 3rd Generation Partnership Project, Technical Standard TS 23.434, ver. 19.1.0, Apr. 2024.

[11] ——, "Procedures for the 5G System (5GS)," 3rd Generation Partnership Project, Technical Standard TS 23.502, ver. 18.5.0, Mar. 2024.

[12] ——, "Architecture enhancements for 5G System (5GS) to support network data analytics services," 3rd Generation Partnership Project, Technical Standard TS 23.288, ver. 18.5.0, Mar. 2024.

[13] ——, "Security architecture and procedures for 5G System," 3rd Generation Partnership Project, Technical Standard TS 33.501, ver. 18.5.0, Mar. 2024.

[14] ETSI, "ETSI – Standards for NFV – Network Functions Virtualisation — NFV Solutions," Last accessed 16 Apr 2024. [Online]. Available: https://www.etsi.org/technologies/nfv

[15] 3GPP, " Integrated Sensing and Communication," 3rd Generation Partnership Project, Technical Report TR 22.137, ver. 19.1.0, Apr. 2024.

[16] KDDI Research, "Beyond 5G/6G," KDDI Corporation, White Paper ver. 2.0.1, Oct. 2021. [Online]. Available: https://www.kddi-research.jp/sites/default/files/kddi_whitepaper_en/pdf/KDDI_B5G6G_WhitePaperEN_2.0.1.pdf

[17] I. Ali, N. Al-Dhahir, and J. Hershey, "Predicting the visibility of LEO satellites," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, pp. 1183–1190, 1999, doi: 10.1109/7.805436.

[18] M. Ericson, M. Condoluci, P. Rugeland, S. Wänstedt, M. S. Abad, O. Haliloglu, M. Saimler, and L. Feltrin, "6G architectural trends and enablers," in *2021 IEEE 4th 5G World Forum (5GWF)*, 2021, pp. 406–411, doi: 10.1109/5GWF52925.2021.00078.

[19] B5GPC, "Beyond 5G white paper supplementary volume on "E2E architecture"," Beyond 5G Promotion Consortium, White Paper, Mar. 2024. [Online]. Available: https://b5g.jp/w/wp-content/uploads/2024/03/Beyond-5G-White-Paper-Supplementary-Volume-E2E-Architecture.pdf

[20] M. Kurata, "Directions and key enablers toward 6G mobile network for easy operation and management," in *11th FOKUS FUSECO Forum*, Sep 2023, Track 1 – Beyond 5G and 6G network technologies and enablers: RAN, core network disaggregation, NTN, edge computing, virtualization, network automation, and Machine Learning (ML).

[21] S. Kukliński, L. Tomaszewski, and R. Kołakowski, "On O-RAN, MEC, SON and network slicing integration," in *2020 IEEE Globecom Workshops (GC Wkshps)*, 2020, pp. 1–6, doi: 10.1109/GCWkshps50303.2020.9367527.

[22] Tomaszewski, L. *et al.*, "ETHER: Energy- and cost-efficient framework for seamless connectivity over the integrated terrestrial and non-terrestrial 6G networks," in *Artificial Intelligence Applications and Innovations. AIAI 2023 IFIP WG 12.5 International Workshops*, I. Maglogiannis, L. Iliadis, A. Papaleonidas, and I. Chochliouros, Eds. Cham: Springer International Publishing, Jun. 2023, pp. 32–44, doi: 10.1007/978-3-031-34171-7_2.

[23] H. J. Einsiedler, "SBA and cloud native principles will shape the future telco world – some ideas for discussion!" in *11th FOKUS FUSECO Forum*, Sep 2023, Track 1 – Beyond 5G and 6G network technologies and enablers: RAN, core network disaggregation, NTN, edge computing, virtualization, network automation, and ML.

[24] Hexa-X, "Analysis of 6G architectural enablers' applicability and initial technological solutions," Hexa-X project – A flagship for B5G/6G vision and intelligent fabric of technology enablers connecting human, physical, and digital worlds, Deliverable D5.2 ver. 1.0, Oct. 2022. [Online]. Available: https://hexa-x.eu/wp-content/uploads/2022/10/Hexa-X_D5.2_v1.0.pdf

[25] Next G Alliance, "6G technologies," Next G Alliance, Next G Alliance Report, Jun. 2022. [Online]. Available: https://nextgalliance.org/wp-content/uploads/dlm_uploads/2022/07/TWG-report-6G-technologies.pdf

[26] Magma (The Linux Foundation Project), "Implications of the Magma Architecture; Interoperability, scale and resilience," Last accessed 16 Apr 2024. [Online]. Available: https://magmacore.org/wp-content/uploads/sites/5/2022/10/WP_Implications_of_Magma_Architecture_092822.pdf

[27] S. Kukliński, L. Tomaszewski, R. Kołakowski, and P. Chemouil, "6G-LEGO: A framework for 6G network slices," *Journal of Communications and Networks*, vol. 23, no. 6, pp. 442–453, 2021, doi: 10.23919/JCN.2021.000025.

[28] S. Sthawarmath, E. Renault, and T. Lejkin, "Designing stateless control-plane for next-generation telecom networks," in *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, 2022, pp. 154–159, doi: 10.1109/MeditCom55741.2022.9928749.

[29] E. Goshi, R. Stahl, H. Harkous, M. He, R. Pries, and W. Kellerer, "PP5GS – an efficient procedure-based and stateless architecture for next-generation core networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3318–3333, 2023, doi: 10.1109/TNSM.2022.3230206.

[30] M. Corici, F. Eichhorn, H. Buhr, and T. Magedanz, "Organic 6G networks: Ultra-flexibility through extensive stateless functional split," in *2023 2nd International Conference on 6G Networking (6GNet)*, 2023, pp. 1–8, doi: 10.1109/6GNet58894.2023.10317754.

[31] L. Nathaniel, G. V. Perdana, M. R. Hadiana, R. M. Negara, and S. N. Hertiana, "Istio API gateway impact to reduce microservice latency and resource usage on kubernetes," in *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2023, pp. 43–47, doi: 10.1109/ISITIA59021.2023.10221035.

[32] D. B. Oljira, A. Brunstrom, J. Taheri, and K.-J. Grinnemo, "Analysis of network latency in virtualized environments," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6, doi: 10.1109/GLOCOM.2016.7841603.

[33] M. Sollfrank, F. Loch, S. Denteneer, and B. Vogel-Heuser, "Evaluating docker for lightweight virtualization of distributed and time-sensitive applications in industrial automation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3566–3576, 2021, doi: 10.1109/TII.2020.3022843.