Setting 6G Architecture in Motion – the Hexa-X approach

Mårten Ericson[1], Stefan Wänstedt[1], Merve Saimler[2]
Ericsson Research
[1]Sweden, [2]Turkey
{marten.ericson, stefan.wanstedt, merve.saimler}@ericsson.com

Damiano Rapone
TIM, Access Innovation Dept.
Turin, Italy
damiano.rapone@telecomitalia.it

Giovanni Nardini
University of Pisa
Pisa, Italy
giovanni.nardini@unipi.it

Hannu Flinck[1], Gerald Kunzmann[2]
Nokia Strategy & Technology
[1]Espoo, Finland, [2]Munich, Germany
{hannu.flinck, gerald.kunzmann}@nokia.com

Antonio de la Oliva, Carlos J. Bernardos
University Carlos III of Madrid
Madrid, Spain
{aoliva, cjbc}@it.uc3m.es

Miltiadis Filippou, Markus Mueck
Intel Next Generation & Standards
Neubiberg, Germany
{miltiadis.filippou, markus.dominik.mueck}@intel.com

Panagiotis Vlacheas, Panagiotis Demestichas
WINGS
Athens, Greece
{panvlah, pdemest}@wings-ict-solutions.eu

Riccardo Bassoli[1], Frank H.P. Fitzek[1,2]
[1]Technische Universität Dresden
[2]Centre for Tactile Internet with Human-in-the-Loop (CeTI), Cluster of Excellence
Dresden, Germany
{riccardo.bassoli, frank.fitzek}@tu-dresden.de

***Abstract*—The most recent cellular generation, 5G, is being deployed on a large scale globally. The capabilities of 5G surpass all previous generations of cellular networks and support many new services compared to 4G. Despite this, at the same time, preparations for 6G have begun since user demands and technical development continuously push the boundaries of what is possible. Demands come not only from users. Also, society sets requirements, e.g., sustainability, coverage, and privacy. To support the necessary features in the network needed to meet the requirements, a new generation of the architecture is needed; one based on the most forward-looking design principles together with trends in networks, use cases, and whatnot. To show that the proposed new features will allow the future network to meet the set requirements, key performance indicators (KPIs) have to be defined. In this paper, we present six of the KPIs that the European 6G flagship project Hexa-X has identified as the fundamental ones to measure the most important aspects of a new 6G architecture.**

***Keywords*—6G architecture, KPIs, cloud**

I. Introduction

Mobile operators need to continuously update their existing networks to match the growing data demands, but also to provide new services which may require lower latency or higher reliability, as well as to support new services provided by 3rd parties. There are two important requirements affecting decisions regarding future network architectures, namely *i)* ensuring backwards compatibility with the already deployed networks when a new functionality is introduced and *ii)* the need to have market-ready solutions as soon as possible. The launch of 5G, whose first commercial introduction was based on the 3GPP Rel-15 E-UTRA NR Dual Connectivity (EN-DC) [1] [2], demonstrated how these requirements were met. This initial network topology, also known as non-Standalone 5G, heavily leverages the network equipment (core and access networks) already deployed for 4G and increases the user data rate by using radio resources provided by the NR base station tightly inter-working with the 4G network. Although EN-DC is suited for enhanced Mobile Broadband (eMBB) applications, it cannot fully support all the 5G-specific services and features, such as, Ultra Reliable Low Latency Communications (URLLC) or network slicing. Especially for URLLC services, additional 5G network topologies have been specified by 3GPP making use of two main key enablers, that is, the NR radio access and the new 5G Core network (5GC), in addition to the evolution of the LTE access.

One of the main innovations in the 5GC design is the so-called Service Based Architecture (SBA), introduced by 3GPP on top of the traditional reference point and interface-based approach, in which Network Functions (NFs) communicate with each other by using a pre-established peer-to-peer signalling interface. With SBA, 5GC NFs can be implemented as a set of software-defined services, each service being provided by a service producer and consumed by one or more service consumers. Another intrinsic SBA feature is that it copes with NFs' load distribution by design, in the sense that, during the process of NF selection, the dynamic load of the candidate NF instances is taken into consideration. SBA also allows for better scalability via a 'plug-and-play' approach. Without SBA, whenever a new function is introduced in the system, selected existing NFs need to be enhanced to support the new functionality and a new peer-to-peer interface needs to be defined between the new function and the existing NFs that communicate with it. Conversely, with SBA, the services provided by a service producer, although initially defined for a specific service consumer (or a set of service consumers), can later also be made available to additional consumers, if needed. One important trend in current networks is cloudification, which is driven by efforts to pool, share and scale functionality efficiently. The move of NFs to cloud is made possible by the increased computing power provided by data centers and virtualization technology. As a result of cloudification new companies are entering the market, usually companies focussed on data centers and network integration. Like cloudification, automation of network operations will be used to further improve the network performance. This is then a main driver for a trend on introducing Artificial Intelligence (AI)/Machine Learning (ML) into cellular networks. There are also attempts to use AI/ML to enhance Radio Resource Management (RRM)-near functions, e.g., mobility and prediction of traffic characteristics. By means of cloudification and automation, networks become more flexible. One important reason is that with cloudification the networks are configured so that functions can be deployed in different ways, e.g., in the cloud or as a monolith. Applying cloudification and automation to mobile networks, is expected to lead to lower costs, both for deployment and operation.

With the increased performance provided by 5G, it is possible to support new services making use of this increase. Since devices evolve with advances in technology and, thus, enable new services with new, higher demands on internal processing and external data transport, user applications are a trend that drives the need for even finer performing cutting edge cellular networks. One service trend today is a 3D digital world, often called the Metaverse[1] where users, among others, will communicate in a virtual 3D environment. This 3D environment will need, at least, 5G technology to work over wireless and a different type of UE, i.e., probably an eXtended Reality (XR) headset. As the metaverse evolves, the cellular system needs to evolve too. One anticipated type of service that is not supported by today's networks, is joint communications and sensing. Sensing is different from most existing cellular services, since, in some cases, the network is believed to be used as a radar, i.e., being able to determine the position of an object by using reflections of transmitted radio waves used for communication purposes.

Finally, meeting the enhanced performance demands on different generations of cellular networks in many cases relies on additional spectrum being made available. To deliver cellular transport with the requirements needed for some of the expected 6G services, wide bands of spectrum will need to be available.

Hexa-X is a European 6G Flagship project with the goal to define the direction of the 6G research [3]. In this paper, part of the Hexa-X process in defining a 6G architecture is described. In particular, the paper discusses KPIs used to determine how well new building blocks fit in the architecture, why these KPIs have been selected and, briefly, why these KPIs are useful.

II. Architectural Gaps

From a gap analysis with respect to current network architectures [4], considering the high demands of a future network, e.g., limitless connectivity and global coverage, the Hexa-X consortium has identified a set of areas and features to study in more detail. In particular, the 5G architectures are not fully cloud-native and designed following the SBA approach. Anticipated new features such as full network softwarization and network-cloud convergence are important aspects to consider when designing a future architecture. Naturally, the new architecture needs to handle Artificial Intelligence (AI) models efficiently. Further, the gap analysis shows that to support new use cases beyond what current networks offer, there is a need for flexible adaptation to new network topologies, along with proper integration of heterogeneous network types, such as, private networks and non-terrestrial networks. For example, programmability of network nodes and devices, allowing easy upgrade of the equipment with new features, should be considered to meet requirements. The evolution of network management and orchestration techniques towards supporting the terminal to Edge to Cloud continuum also requires attention in the 6G architecture design.

Moreover, the development of the 6G architecture needs to consider environmental sustainability (e.g., by improving energy efficiency) and to help addressing economic and social challenges in sectors other than Information and Communication Technologies (ICT). The new 6G architecture will likely consider the new European Commission's Radio Equipment Directive (RED) requirements for safety, health, electromagnetic compatibility, and the efficient use of the radio spectrum.

III. Hexa-X 6G Architecture Vision

Hexa-X defines eight architectural principles for 6G in four different areas (see TABLE I.) where the architecture and the network need to be improved. The principles are described in detail in [4] and are briefly presented below. The principles relate to common cloud platform (P1, P2), network of networks (P3, P4), NFs (P5, P6), and network interfaces and transport (P7, P8), respectively.

Principle P1: Exposure of capabilities: The 6G architecture will expose network and computing capabilities to end-to-end applications providing them with enhanced network features which can be leveraged for upcoming 6G services.

Principle P2: Designed for (closed-loop) automation: The 6G architecture will be designed to support full automation of network and service management operations, utilizing distributed AI/ML agents to be able to manage and optimize the system without human interaction.

Principle P3: Flexibility to different topologies: The 6G network will be a network of networks integrating multiple technologies and topologies. The network will be able to automatically adapt its processes to deal with a variety of topologies without loss of performance.

Principle P4: Scalability: The 6G system will be able to scale from very small to very large-scale deployments by scaling up and down network resources.

TABLE I. Identified Network areas to be improved

Network area	Improvement area
Common cloud platform runs the network functions and services	There is a need to improve this platform to support better interactions between the network and the applications and for higher level of automation.
Network of networks incorporates different (sub)network solutions into one network, and can easily (flexibly) adapt to new topologies.	The network needs to be more flexible to support the current and future deployments in an adaptable way.
NFs comprising RAN and Core Network (CN) functions and services, e.g., RAN scheduling mobility, session handling, etc.	The architecture and the network functions must be designed for higher resilience and availability as well as reliability. Placement of the NFs should be based on run time latency, processing and scalability needs.
Interfaces and transport, i.e., signalling between the different network nodes and NF, including UEs.	The signalling and procedures need to be more efficient and simplified.

[1] https://en.wikipedia.org/wiki/Metaverse

Principle P5: Resilience and availability: The 6G network will show unprecedented levels of resiliency and service availability by enabling multiple inter-connections between the different elements of the network (leveraging the network of networks concept). Wide adoption of Control Plane (CP) and User Plane (UP) separation and multi-connectivity will reduce service failures.

Principle P6: Exposed interfaces are service-based: Following the concept of serverless design and the improvement of the network flexibility through service separation and reuse, the 6G network interfaces will be service-based, where appropriate.

Principle P7: Separation of concerns of NFs: This leads to minimal dependency with other network functions, so that network functions can be developed and replaced independently from each other.

Principle P8: Network simplification in comparison to previous generations: By using cloud-native RAN and CN functions, fewer parameters to configure, and fewer external interfaces, the 6G network will be simpler and easier to design, deploy and maintain.

IV. OUTSIDE HEXA-X

For the period 2021-2027, the European Commission (EC) has undertaken the Joint Undertaking (JU) on Smart Networks and Services (SNS) initiative[2], aiming not only at boosting 5G deployment in Europe, but also at fostering Europe's 6G technology sovereignty. Possible 6G technology comprises strategic areas of the networks and services value chain, from edge- and cloud-based service provisioning to market opportunities in devices beyond smartphones. According to the 5G Infrastructure Association (5GIA), the scope of the 6G Architecture is expanded beyond the RAN and CN to include terminals and data centres to assure complete, end-to-end resource awareness and native support of AI. Europe is targeting the first 6G deployments by the end of the decade. Fig.1 summarizes the main worldwide activities on 6G.

For 6G, the REINDEER[3] project will create hyper-diversity by developing a new sort of smart networking platform, as well as robust and scalable real-time and real-space interactive applications using cell-free protocols and distributed intelligent processing. The objective of the DAEMON project[4] is to develop and implement novel and practical methods for Network Intelligence design that result in high-performance, sustainable, and exceptionally reliable zero-touch network systems for Beyond 5G. Dedicat6G[5] envisions transforming Beyond 5G networks into a smart connectivity platform that is dependable/resilient, highly adaptive, ultra-fast, green for supporting securely innovative, human-centric applications. AI@EDGE[6] will provide a connect-compute fabric for generating and maintaining robust, elastic, and secure end-to-end slices, based on a serverless paradigm focusing on AI for closed-loop automation. It will further provide a distributed connect-compute platform, provisioning of AI-enabled applications, a hardware-accelerated AI/ML serverless platform, and cross-layer, multi-connectivity, and disaggregated radio access. For Beyond 5G, MARSAL[7] aims for an architecture with new levels of flexibility and closed-loop autonomy at all tiers of the infrastructure, as well as dramatically enhanced spectral efficiency through cell-free networking.

Fig. 1. The main world activities on 6G.

The 6Genesis Flagship Program (6GFP)[8] is a national 6G program funded by the Academy of Finland and led by the University of Oulu to develop, implement, and test key enabling technologies for 6G. Similarly, Germany's Federal Ministry of Education and Research (BMBF) aims to of create "the basis for an innovation ecosystem for future communication technologies around 6G". For this purpose, BMBF supports four technology research hubs, including the 6GEM[9], the 6G Research and Innovation Cluster (6G-RIC)[10], the 6G-life[11], and the Open6GHub[12]. The partners involved in the hubs aim to achieve technological sovereignty and ensure data security by exploiting quantum communications, post-Shannon theory, AI, as well as adaptive and flexible hardware and software platforms.

In North America, the Alliance for Telecommunications Industry Solutions (ATIS) launched the Next G Alliance (NGA[13]), an initiative that is engaged with developing a 6G roadmap. According to NGA, green data centers, virtualization, network management approaches, and Internet of Things (IoT) energy usage are dominant priorities for energy saving, renewable energy transition, and, possibly, self-powered devices. The RAN and CN architecture should be improved using new protocols, AI-based networks, and service automation to reduce idle resources and enable on-demand network connectivity.

The Resilient & Intelligent NextG Systems (RINGS)[14] program, led by the National Science Foundation and the US Department of Defence, promotes research in areas that impact on the next generation mobile communication,

[2] https://digital-strategy.ec.europa.eu/en/policies/smart-networks-and services-joint-undertaking/
[3] https://reindeer-project.eu/
[4] https://h2020daemon.eu/about/
[5] https://dedicat6g.eu/
[6] https://aiatedge.eu/
[7] https://www.marsalproject.eu/
[8] https://www.oulu.fi/6gflagship/
[9] http://www.6gem.de/en/
[10] https://6g-ric.de/
[11] https://6g-life.de/
[12] https://www.dfki.de/en/web/news/open6ghub-foerderung-bmbf/
[13] https://nextgalliance.org/
[14] https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm

networking, sensing, and computing systems, with a focus on improving the resiliency of such systems.

South Korea has launched a five-year plan to create fundamental standards and technology. The IITP, a state body affiliated with the Korean Ministry of Science and ICT, inked a deal with the National Science Foundation (NSF) to conduct joint research in 6G technologies. Major Korean companies have laid out their 6G vision based on a common set of technological building blocks such as use of Terahertz bands, AI, novel antenna systems and advanced duplex techniques to enable advanced and innovative 6G services.

China has officially launched a research and development plan in 2019 that involves how to conduct 6G R&D and the technical aspects of 6G. Additionally, the government maps out a clear guide for the entire industry and offers a head start. Following this, main Chinese companies disclosed their vision publicly aiming to design an intelligent architecture and verify intelligent enablers for 6G.

After announcing a program for 6G R&D in 2019, the Japanese government expects to commission the program to private companies and universities through the National Institute of Information and Communications Technology (NICT). NICT aims to develop an R&D roadmap and implementation plan, where the reality of the B5G/6G world is examined through different scenarios, use cases, constituent technologies, and corresponding circumstances required to achieve them. Also, Japan aims to develop core technologies for the 6G system by 2025 and commercially launch the technology in 2030 with major operators. Japan also established a partnership with the USA [15], Finland and Singapore [16] on 6G for research, development, testing, and deployment of secure networks and advanced ICT.

V. 6G Architecture Key Performance Indicators

Our assumption is that the architecture will have distributed and centralized parts and will leverage on virtualization and cloud technology. Also, a central assumption is the inherent use of AI/ML based network automation mechanisms to optimize and adapt the network architecture to the needs to different deployments assuming a high level of configurability of the architecture. For a successful introduction of a new 6G architecture, there must be relevant and concrete measurements that show that proposed new features meet the requirements. Below we list a set of well-defined architectural Key Performance Indicators (KPIs) tailored for measuring what we believe are the most important aspects of a new 6G architecture.

A. KPI 1: Convergence time needed for the network to adapt to occurring changes

The convergence time KPI covers the time to adapt the network and its constituent elements, traffic routes and radio coverage to reflect the optimization decisions taken by the network management, orchestrator(s) and AI-agents. Convergence time measures how quickly a network can adapt a new network configuration to match changes in the deployment environment. There may be multiple reasons that trigger a need for such configuration changes, including addition of new network or computing resources, changes in traffic patterns, introduction of new services or due to service, device or link failures. An accurate estimation of this KPI is needed to ensure that the reconfiguration of the network does not impact the End-to-End (E2E) service experience. Timely and accurate network reconfiguration is particularly important for the following Hexa-X use cases characterized by changes in the deployment environment:

- E-health;
- Immersive smart city;
- Digital Twins for manufacturing;
- Robots to collaborative robots (or "cobots").

The overall convergence time KPI can be split into the following components, which can be measured separately, and measured in time unites (sec, msec). The subcomponents are presented in the typical order where they appear when changing the network configuration:

1) Detection time to trigger a configuration change. The trigger could come from multiple sources: Operations and Maintenance (O&M), network analytics, policy control, performance monitoring of Service-Level Agreements (SLAs), as well as monitoring of cloud resources.
2) Time needed to issue a reconfiguration decision and selection of impacted network nodes, services, and main functions. If new resources are added or old ones removed from the network, the related AI/ML models may need to be updated, and, in the worst case, retrained. In some cases, the configuration change, most likely caused by addition of a new network node or a new type of service may lead to a time-consuming retraining operation, if none of the pre-trained models can be used for this new configuration.
3) Time needed for propagation and installation of the new configuration and update of the AI/ML agents as needed across the network.
4) Time needed for completion of ongoing network tasks (e.g., ongoing signalling transactions, emptying packet buffers, etc.).
5) Time for the system to reach a new stable desired operation state. This includes path switching times, and new QoS states of the service flows. Performance monitoring of the user flows shall indicate when the desired state has been reached. Closed looped control mechanisms managing affected subsystems require a synchronization phase with sufficient hysteresis to reach stable state of end-to-end performance.

Coordination of these dynamic reconfiguration operations across all network segments is essential to guarantee proper E2E behaviour of the services. This sets requirements on synchronization between the involved orchestrators, network controllers, and AI/ML agents.

B. KPI 2: AI communication and computing overhead

The amount of additional computing and AI communication resources allocated to optimize E2E QoS in comparison to acceptable, albeit static resource allocation, provides a first order metric for this KPI. However, as the

[15] https://asia.nikkei.com/Business/Telecommunication/US-and-Japan-to-invest-4.5bn-in-next-gen-6G-race-with-China

[16] https://echalliance.com/finnish-led-international-6g-technology-cooperation-expands-to-singapore/

resource consumption in 6G is expected to be even more volatile than in 5G due to new resource critical services (e.g., E-health, robots, etc.) a static reference set up to define the relative overhead may not be possible to cover the dynamicity of use cases. Therefore, we must break down the AI communication and computing overhead and consider the consumed resources against the value they provide to the overall service experience. AI needs continuous data logging from network nodes and UEs in order to perform continuous and refined learning together with accurate decision making. Such data logging will be transmitted and cause communication resource overhead in both the UL and the DL. In addition to this, use of AI may cause additional signal processing load at the device side. This processing load can be expressed as a function of number of antennas, modulation bits, coding rate, MIMO-layers and number of physical resource blocks [5],[6]. For example, a large usage of resource blocks will result in higher processing load. The AI communication and data logging will, therefore, increase the UE processing overhead. Here, a critical trade-off arises. In order to reduce latency, in-network intelligence will be used (i.e., for anticipatory networking) and the processing of the softwarized functionalities will be at the edge. However, in parallel, the latency will increase by the increased usage of the RAN link (transmission latency inversely depending on the available capacity) and the processing load. Furthermore, the use of explainable AI models, needed to enhance AI-enabled network trustworthiness, can also lead to additional overhead, due to the need of sending and receiving the explanations for the output provided by the AI algorithms themselves. Thus, supporting in-network AI functionalities makes the need for scalable and flexible communication protocols evident. For example, Federated Learning (FL) processes will need new signalling for joining and/or leaving a group of federated UEs, as well as for training and obtaining AI models. First, the design of such protocols must consider a large, variable number of devices involved in the AI process. Also, protocols should be flexible enough to support different communication paradigms, such as request-response and subscribe-notify, and choose the most suitable one according to the network conditions to minimize the communication overhead. Moreover, AI-related NFs may be required to be moved from mobile devices to the edge/cloud and vice versa to reduce AI communication and computing overhead.

To find the optimal operation point between communication and computing overheads in an AI-native network, one should also consider the overall cost (both billing and energy-related) relating to wireless (learning) data transfer and the occupation/sharing of storage, memory and processing resources. To achieve the goals of a sustainable and intelligent 6G network, the Compute-as-a-Service (CaaS) approach can be exploited. CaaS aims to facilitate discovery of compute resources external to the calling entity, such as an edge infrastructure node or a user device, through a well-defined open interface. Given such discovery capability, a workload orchestrator can decide (or, at least recommend) upon offloading a processing workload, based on the available resources of discovered network nodes and considering requirements, such as the incurred latency from task generation to output acquisition by the CaaS consumer, the energy footprint and monetary cost of the task delegation, the trustworthiness of the entity to host the workload and its robustness to software and hardware failures. Of course, the energy footprint, service cost and added signalling complexity of the CaaS approach needs to be compared against "vanilla" network configurations lacking it.

C. *KPI 3: Reliability/robustness for network of networks*

This KPI describes the ability of the network to minimize the radio link failures while maximizing QoS (even if some QoS requirement may not be reached). The KPI can be a combination of two or more metrics such as minimizing the downtime of a connection (radio link failure) while maximizing the possible QoS (or maintaining a minimum QoS). The KPI is averaged over all users in a network. The new 6G enablers to improve this KPI are several based on the deployment type. The main enabler for this KPI is the concept of network of networks. 6G will consist of many sub-networks, i.e., networks of networks. Therefore, there is a need to efficiently integrate the different sub-networks. In addition, the mobility aspects must be enhanced to support 6G strict requirements on reliability and availability. This includes a new improved multi-connectivity solution and integrating Non-Terrestrial Network (NTN) inherently in 6G. Integrating NTN inherently in 6G enables an improved reliability and robustness. For example, a fast-moving UE can quickly switch to an NTN network if the terrestrial network becomes bad. At the same time, a new enhanced multi-connectivity solution with decoupling of UL and DL as well as decoupling of UP and CP can also improve the reliability. In DL, it can be advantageous to send data via several cells to one UE, while, in the UL, it is typically better to send data only to one cell (to avoid splitting the UE transmit power). Other features that help improve the reliability can be to have multiple CP connections over the air and possibly over different infrastructure.

D. *KPI 4: Network flexibility*

This KPI describes the ability for the network architecture to perform well over a wide range of deployments and network states, i.e., it is a combination of selected KPIs over several different deployments. This means any standard KPI must always be fulfilled regardless of network deployment. One 6G enabler to achieve the network flexibility is to develop a flexible topology for a "network of devices". There is a need for an advanced architecture with network of devices as an option to address higher coverage/computation needs, lower latencies, reliability, security, and decentralization. The enablers for this are:

- Design algorithms for discovering and selecting the best possible and "trusted" nodes and far-edge devices, as well as the best connectivity options.
- Unified modelling of far-edge nodes and devices and definition of interfaces to control and interact with them.
- Delegation of computational tasks to edge (or cloud) nodes.

Another enabler is the campus network which exploits the interconnection of several local area networks within a limited geographical area [6]. Unprecedented challenges arise from the realization of a network of networks, which consists of public and non-public networks (e.g., campus), hosting and interconnecting heterogeneous technologies and services. 6G also envisions the design of an efficient and effective 3-dimensional combination of campus and public networks.

This implies the careful design of Virtualized NF (VNF) placement and of a programmable protocol stack to ensure required E2E performance in the 6G 3-dimensional architecture. In fact, reliability (both for hardware and software), previously discussed, becomes a function of the increased flexibility. And the increase in flexibility inversely affects the software reliability (for example the one of microservices running in virtual containers). Next, intelligence used for network management and especially prediction of future network states is a non-deterministic approach to networking. The 'autonomous' flexibility obtained by softwarization and intelligent prediction implies a price to pay in terms of reduced ensured reliability. This approach of 6G requires investigation to ensure predictions, which match the 100% network availability and very high reliability that is expected by the very sensitive services hosted by future 6G networks like ultra-reliable communications.

E. KPI 5: Separation of concerns of network functions

To achieve flexibility, as mentioned in some previous KPIs, the network needs to be able to adapt to different situations, e.g., some situations may require functionality not needed in other situations. One way to define this KPI is to measure the number of dependencies to other NFs. When adding new functions, effort should be put in the process to try to separate the functions as needed. There must be clear division of responsibility, especially in multi-vendor networks. For example, there are situations in current networks where many nodes need access to UE context, like the multiple solutions for handling IoT, where some are in the CN, and some are in the RAN. In such situations it is not clear who "owns" the responsibility and should therefore be avoided.

This KPI is very much related to how many and what kind of transaction (earlier known as "procedures") a certain NF functional split results into. The behaviour and interfaces of a function need to be clearly defined, so exchangeability of functions is practically enabled. While in the "legacy/old" telco world, external interfaces were the focus of standardization, as functional entities were implemented by physical boxes and could not easily or dynamically be replaced. With the virtualization paradigm and the SBA architecture refocus, now it is critical to be able to define functions in a way that allow them to be re-used, while not adding a lot of dependencies. This may need extending existing function descriptors and templates to also characterize dependencies and concerns. Orchestration entities will also need to be enriched to manage those types of additional constraints.

F. KPI 6: Ease of adding new functions in future

The process of cloudification of a cellular network should include a study of how well actions for UEs perform in different settings, e.g., how many nodes or NFs are involved in an action. This KPI is defined as the number of involved nodes and the number of involved interfaces. A bad NF design leads to that many nodes and interfaces are involved which may introduce a lot of signalling. Another advantage of isolating functionality in NFs, i.e., limiting the functionality, is that, in theory, it will be possible to add or remove NFs without really affecting the system in general.

Dependencies can be either non-UE associated, i.e., characterizing relations between the RAN node and the Access and Mobility Management Function (AMF). However, dependencies are more often UE specific, relating to UP, security, UE context management/mobility, UE – NF instance binding, etc.

To be able to characterize NFs based on how efficiently they serve processes of involved devices there needs to be relevant measurements. One way to measure these dependencies is to see the number of specifications that need to be updated for the addition of a new NF. Another way to measure this is to see how many other NFs need to be changed when adding the new NF. However, this is also related to the amount of separation mentioned in the KPI in the previous section. The mentioned measurement will not be compared to some threshold but should instead be used to compare different alternatives. However, the overall objective is, of course, to minimize the number of dependencies.

VI. Conclusions and Future Directions

For a successful introduction of a new 6G architecture, concrete and relevant measurable criteria need to be defined to verify that proposed new features meet the envisioned requirements. A network architecture reconciles requirements and expectations from multiple parties of an ecosystem, containing end users, content service providers, new verticals and network equipment vendors, etc. In this paper we present six KPIs, defined in the Hexa-X project, that are used to compare and identify trade-offs between alternative approaches. These KPIs are tailored for measuring what we believe are the most important aspects of the forthcoming 6G architecture. Further, we scrutinize each of these six well-defined KPIs by underlining the new trade-offs and intrinsic limitations that may arise. In this sense, the article highlights important scientific challenges in the design of the 6G architecture, providing guidelines for the ongoing 6G research.

Acknowledgment

This work has been funded by the European Commission through the H2020 project Hexa-X (Grant Agreement no. 101015956).

References

[1] 3GPP TS 37.340 - NR; Multi-connectivity; Overall description; Stage 2, Version 16.8.0, December 2021.

[2] 3GPP TS 38.331 - NR; Radio Resource Control (RRC); Protocol specification, Version 16.7.0, December 2021.

[3] M. Uusitalo; P. Rugeland; M. Boldi, et. al, "6G Vision, Value, Use Cases and Technologies From European 6G Flagship Project Hexa-X," in IEEE Access, 2021, Vol. 9, DOI: 10.1109/ACCESS.2021.3130030.

[4] Hexa-X, Deliverable D5.1 "Initial 6G Architectural Components and Enablers", V 1.0, December 2021.

[5] T. Werthmann, H. Grob-Lipski, S. Scholz, and B. Haberland, "Task assignment strategies for pools of baseband computation units in 4G cellular networks," in 2015 IEEE International Conference on Communication Workshop (ICCW), June 2015, pp. 2714–2720. DOI: 10.1109/ICCW.2015.7247589.

[6] R. Bassoli, F. H.P. Fitzek, E. Calvanese Strinati, "Why do we need 6G?", ITU Journal on Future and Evolving Technologies, Volume 2 (2021), Issue 6 - Wireless communication systems in beyond 5G era, Pages 1-31, DOI : 10.52953/IROR5894.