Beyond 5G White Paper Supplementary Volume on “E2E Architecture”

Version 1.0

March 7, 2024

XG Mobile Promotion Forum

Contents

【Revision History】

Ver.	Date	Contents	Note
1.0	2024.3.7	Initial version	

Preface

In the future society in the 2030s, it is expected that various services that take advantage of the characteristics of Beyond 5G will be realized. In Beyond 5G, coverage expansion and speedup are advanced by various technologies. In addition, such communication functions will be highly virtualized and can be controlled by software. The network architecture of Beyond 5G needs to provide the optimum infrastructure functions that satisfy the performance required by applications and the quality of experience of users without users being aware of the infrastructure and technologies of Beyond 5G, the congestion of communication bands, and security.

In Japan, Beyond 5G is for the advancement of Society 5.0[1], and network architecture is also required to efficiently provide the functions of Beyond 5G infrastructure in order to advance Society 5.0. In addition, Society 5.0 is expected to develop the economy by solving various social issues through CPSs (Cyber-Physical Systems), which are highly integrated virtual and real spaces. CPSs build real spaces on virtual spaces using AI (Artificial Intelligence) and sensors, which are ubiquitous in society, reproduce social activities on virtual spaces, and bring social value by acting on real spaces. In Beyond 5G, from the viewpoint of realizing CPSs, both virtual spaces and real spaces, which realize various Society 5.0 services, can be connected, and controlled. In this way, Beyond 5G can integrate real spaces and virtual spaces and handle ubiquitous AI in order to realize Society 5.0, including optimizing its own network functions. In addition to network functions, it is expected to be an architecture that can optimally provide computing resources that make AI available.

This Supplementary Part describes the following E2E architecture-related technologies in order to realize the vision for 2030 and provide Beyond 5G infrastructure functions in response to the diverse needs of users.

Digital-Twin for and by Beyond 5G

Digital-Twin can be both digital representation of real-world and digital representation of network objects. Federating and jointly optimizing both real-world and network objects is essential to realize new services in the era of Beyond 5G.

Optimum collaboration of network functions and computing resource

Introducing a technique that enables the optimal use of computing resources by using computing resources located in different network domains (access network, MEC, core network, cloud) as one virtualized computing resource.

User-Centric Network

Introducing an overview of "user-centric networks", an innovative approach designed to provide stable, high-speed, and continuous communication services to each user at anytime and anywhere.

Intent-based operational plan generation for business utilization of autonomous networks

Proposing a method to generate a network operation plan from the user's intention, which allows the user to activate the operation plan after verifying its validity.

Task-Oriented 6G Native-AI Network Architecture

Introducing a task-oriented native AI network architecture to natively support network AI, which integrates AI into the network.

Envisioning Architectural Transformation towards 6G

Introducing system architecture migration options to 6G, interworking aspects with legacy generations, and the recommended design of the overall system architecture, including RAN-CN functional split and logical RAN architecture.

REFERENCE

[1] Society 5.0 https://www8.cao.go.jp/cstp/english/society5_0/index.html

This White Paper was prepared with the generous support of many people who participated in the White Paper Subcommittee. The cooperation of telecommunications industry players and academia experts, as well as representatives of various industries other than the communications industry has also been substantial. Thanks to everyone's participation and support, this White Paper was able to cover a lot of useful information for future business creation discussions between the industry, academia, and government, and for investigating solutions to social issues, not only in the telecommunications industry, but also across all industries. We hope that this White Paper will help Japan create a better future for society and promote significant global activities.

Kentaro Ishizu, NICT
Mitsuhiro Azuma, NICT

Digital-Twin for and by Beyond 5G

Hideyuki Shimonishi, Osaka University
Kentaro Ishizu, Koji Zettsu, NICT
Toshirou Nakahira, NTT
Dai Kanetomo, NEC

Abstract*—In this article, we discuss that a Digital-Twin can be both digital representation of real-world, i.e. Digital-Twin by Beyond 5G, which could be enabled by advanced Beyond 5G capabilities, and digital representation of network objects, i.e. Digital-Twin for Beyond 5G, which helps to enable advanced Beyond 5G capabilities. Federating and jointly optimizing various Digital-Twin instances of both real-world and network objects is essential to realize new services in the era of Beyond 5G, and thus propose a Digital-Twin architecture which manages various Digital-Twin instances in a common way so that any Digital-Twin applications can easily utilize them. We then introduce probabilistic Digital-Twin, which can improve both efficiency and safety of many Digital-Twin use cases by considering uncertainties inherent in the real world, and cross-domain orchestration of Digital-Twins, which will be a key to realize the digital-first services. Finally, we introduce some of examples of the Digital-Twins discussed in this article, including radio communication environment, human-robot cooperation, and smart sustainable mobility.

1 Introduction

Digital-Twin is a digital reproduction of objects in physical space (cars, jet engines, people, buildings, cityscapes, etc.), or potentially in virtual space or so called Metaverse. It is expected to be an important technology for various ICT systems in factories, aviation, connected cars, smart cities, smart buildings, etc., in realizing advanced Cyber Physical Systems. The concept of Digital-Twin has been introduced in various literature since the 2000s, and widely accepted in recent years when several literatures, such as 0, has been known. Also, several articles, such as [2], provides an extensive survey on Digital-Twins, including enabling technologies and technical issues.

As an extension of classical notion of Digital-Twin, which is a one-to-one correspondence between objects in physical space and objects reproduced in virtual space, Digital Twin Network [3] has been proposed to represent of networks of multiple objects so that various objects in real and virtual space share information and cooperate to perform specific tasks in connected cars, smart cities, etc. In addition, Cognitive Digital Twin [4] focuses on knowledge representation, called ontology, to handle various types of objects in the real world. This is expected to enable more sophisticated applications by sharing digital twins between different systems.

Digital-twin, sometimes called Network Digital-Twin, is also used to plan, design, manage, simulate, operate, and control networks, as discussed by ITU-T [5], IOWN Global Forum Digital [6], and TMforum [7]. In this case, the Digital-Twin is a digital reproduction of any network devices, edge/cloud computing resources, terminal devices or robots which has network interfaces, radio communication environments measured or estimated by various sensing devices, or even logical networks and services.

2 Digital-Twin Platform Architecture with Beyond 5G

The key concept of the proposed Digital-Twin platform architecture is: 1) Digital-Twin instances of both real-world and network objects can be handled freely without any distinctions, and

2) various Digital-Twin instances can be easily federated and jointly optimized, so that any Digital-Twin applications can easily utilize them to realize new services in the era of Beyond 5G.

Figure 1 shows the proposed framework. As above discussed, target physical object includes any real-world and network objects. Other logical objects such as logical networks could also be included. Any measurement data are collected from them to reconstruct them in the Digital-Twin space, using any sensing/control devices like cameras, radio monitors, ISAC (Integrated Sensing and Communications), and legacy network managers such as syslog or EMS. Raw data analysis, such as object detection from video image, could be done here to extract meaningful information from the raw data. Then, common Digital-Twin functions would be provided by the platform so that Digital-Twin applications can utilize the Digital-Twin instances through common and open APIs. The functions include 1) device connectivity function to connect any kinds of sensing/control devices through common interfaces like WoT or MQTT, 2) device abstraction function to easily utilize various versions of devices in a common way, 3) data management functions to manage the data of any Digital-Twin instances, 4) various analysis functions, such as probabilistic inference discussed below, commonly useful for many Digital-Twin applications, and 5) Data isolation and access control function so that different applications can share the data and federate each other. Those functions may use common open source platform like Eclipse Ditto [8].

Figure 2 shows the implementation of Digital-Twin in distributed infrastructure. As its nature, physical devices, sensing/control functions, as well as raw data analysis are implemented on local devices or edge computing devices. Digital-Twin functions are applications are implemented on distributed computing and storage infrastructure and mutually interconnected via high-speed and low-latency communication infrastructure. We also note that all those functions shown in Figure 1 and Figure 2 are enabled on an advanced Beyond 5G communication, computation, and storage infrastructure.

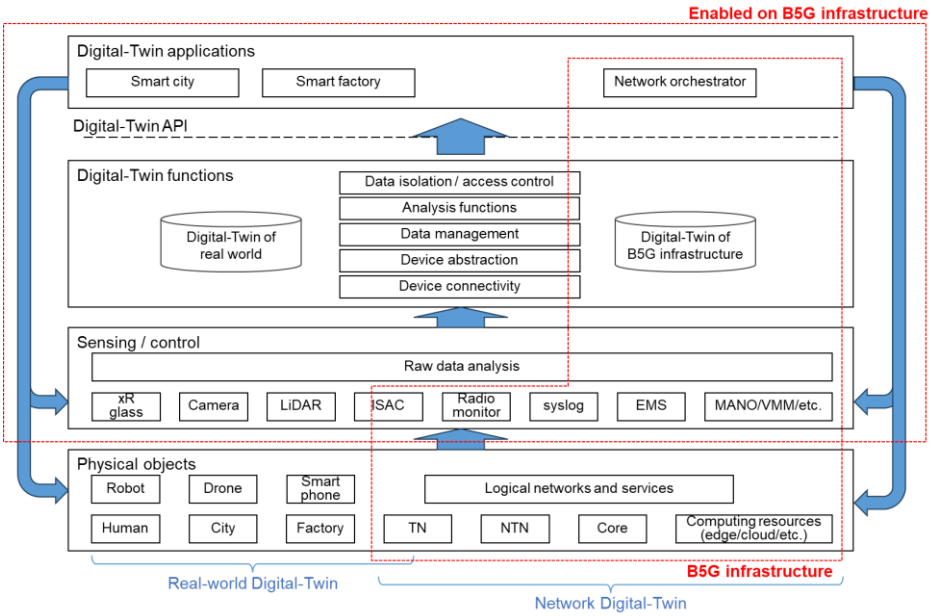


Figure 1 Digital-Twin Platform Architecture

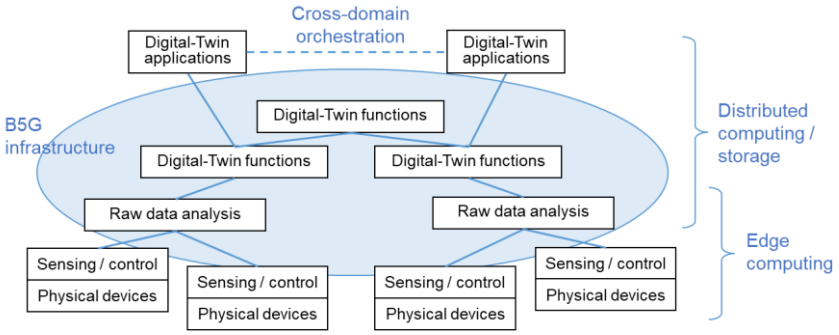


Figure 2 Digital-Twin Implementation in Distributed Infrastructure

3 Functional Design

3.1 Probabilistic Representation of Digital-Twin

Safety and trustability are critical for many Digital-Twin applications in many real-world use cases and thus we propose Probabilistic Digital-Twin in which risk management can be better handled through probabilistic representation of the real-world, and probabilistic prediction and probabilistic control based on the probabilistic representation. Use cases of the probabilistic Digital-Twin includes following examples.

- Human Robot Collaboration, autonomous robot, automatic driving (Figure 3 left): Risk sensitive path/speed/behavior control with probabilistic information for Improved/optimized/controlled safety.
- Remote-controlled robot (Figure 3 center): Move slowly in unstable radio condition, avoid unstable signal area for Robust operation, avoid risks, etc.

- Network design and control (Figure 3 right): BS (Base Station) location design and beam forming control based on probabilistic radio environment map for cost effective, robust, application aware network design and control.

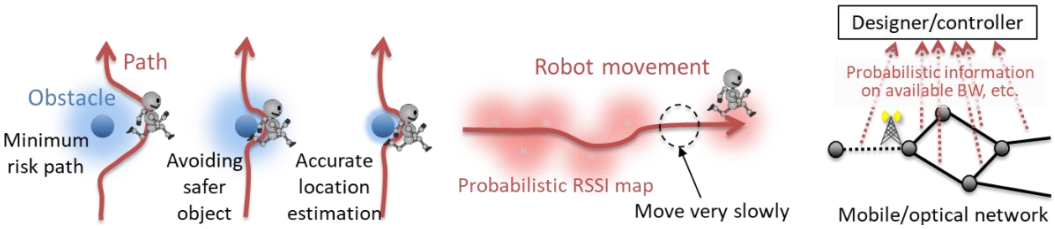


Figure 3 Digital-Twin Implementation in Distributed Infrastructure

Figure 4 shows an example of a data structure for probabilistic Digital-Twin. In between Digital-Twin applications and devices, a space-time structure is composed of many objects in the 4D space. These objects may be a physical object in the space (orange dot) or a status of a lattice point in the 4D space (green dot). Properties of these objects, such as occupancy status of the lattice point, location of the objects, identity/class of the objects are expressed as a probability distribution, rather than a specific value.

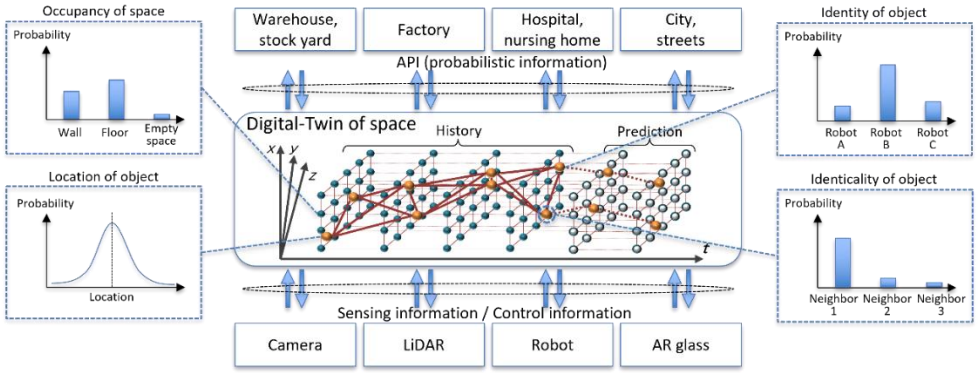


Figure 4 Data Structure for Probabilistic Digital-Twin

3.2 Cross-Domain Orchestration of Digital-Twins

Today, many smart cities are introducing digital twins, using IoT sensors to collect and monitor urban data. Conventionally, they have put much effort to digitize physical space on cyber space in order to analyze and simulate the real world through data for situation monitoring and decision making. From now, we will focus on feedback to the real space to implement the results of analysis and simulation, which is called "digital-first" paradigm[9]. The digital twin collaboration will be the key to realize the digital-first services.

The beyond 5G/6G functional architecture[13] is an open platform to receive a diverse set of functions. One of the key components is the **orchestrator**, which is responsible for

finding the right combination of system components and linking them together to meet the requirements from a CPS service. The orchestrator facilitates the coordination of digital twins across industries to realize a myriad of new value-added services.

To facilitate information sharing and interaction between digital twins among different domains, orchestrators are required to have the functions shown in Figure 5. The **federation function** configures and manages federations of digital twins that update a shared virtual model while maintaining privacy data generated by physical objects within individual digital twins. The **translation function** facilitates the formal and semantic transformation of communications between digital twins in different domains. The **brokering function** identifies and authenticates digital twins, relays data transmission and reception, performs data filtering, real-time delivery, and guarantees delivery. The **synchronization function** synchronizes many-to-many interactions between physical space entities and cyberspace models between digital twins. **The registry function** registers and discovers digital twin components based on their feature information

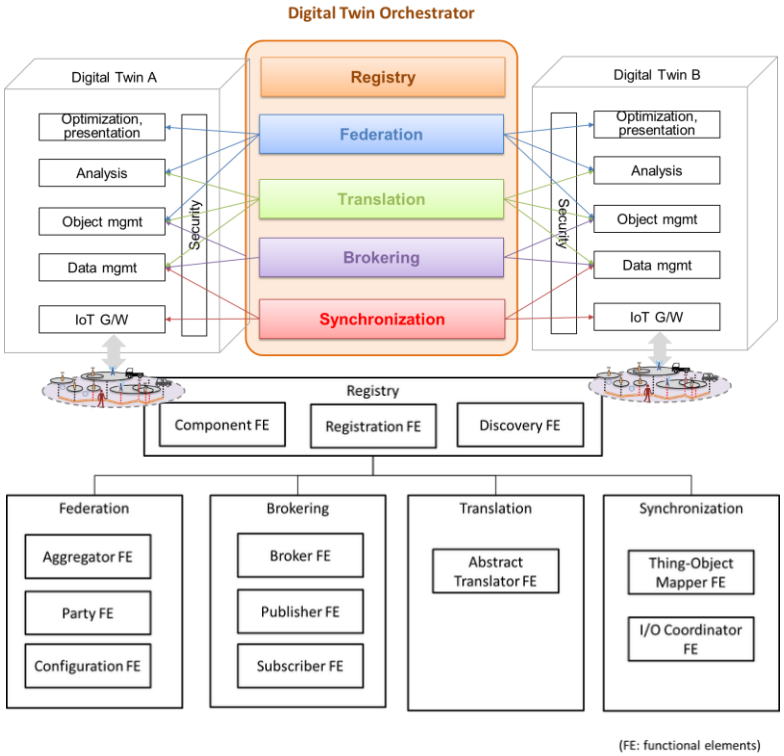


Figure 5 Functional Architecture of the Digital Twin Orchestrator

4 Use Case Examples

4.1 Digital-Twin for Radio Communication Environment

This use case is to manage the radio communication network via a Digital-Twin. Stability of radio communication is so crucial for mission critical communications such as remote robot operations or connected car operations. However, radio communication is heavily affected by radio environment especially when high frequency radio like mm-Wave is used for mobile communications. So, it is very important to understand the radio environment in detail as a Digital-Twin to manage radio communications.

As discussed in Section 3.1, it is quite difficult to monitor, estimate, and predict the radio environment, so we have we proposed a method for constructing the Digital-Twin using probability distributions [15]. We proposed using a graphical model known as Markov Random Field (MRF) to depict the spatial structure of the RSRP (Reference Signal Received Power) map. We then evaluated the proposed method in our own 28 GHz Local 5G environment and extensively studied the characteristics of indoor millimeter radio communication. We confirmed that each estimated point can be well estimated by probability distribution using the proposed method, as shown in Figure 6. We also considered a radio network design in which estimated RSRP value becomes larger than required value plus certain margin at each location. We proposed to set the margin based on the inferred probability distribution, rather than to set a uniform margin. As shown in Figure 7, when the target coverage rate, i.e. the ratio of points that the observed RSRP value is within the margin, is set to 90%, the proposed method can achieve this by using a 1.1σ interval as the margin at each point, whereas the conventional method requires a uniform 1.9σ interval average as the margin.

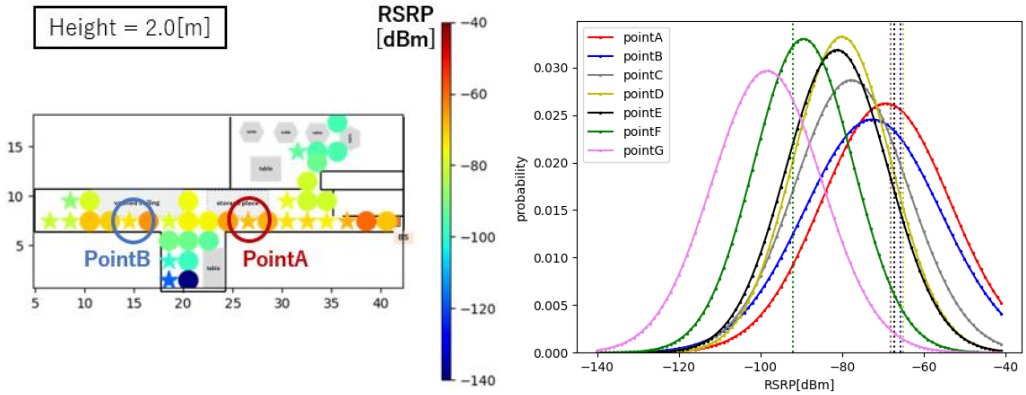


Figure 6 Estimated Radio Environment Map and its Probability Distribution

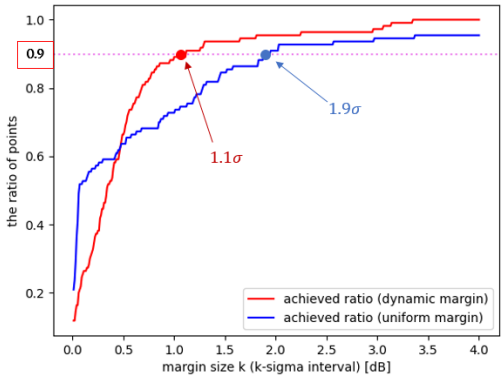


Figure 7 Ratio of points that the observed value is within the margin

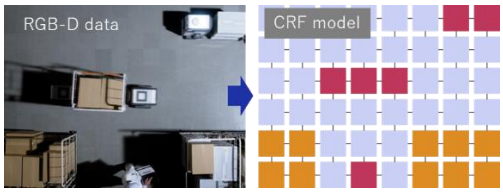
4.2 Digital-Twin for Human-Robot Cooperation

Robots are widely utilized in industrial sites due to the decrease in the number of workers. However, there are many sites where it is difficult to replace all operations with robots due to cost and environmental adaptability, and robots and workers need to coexist. A logistics facility is a such site.

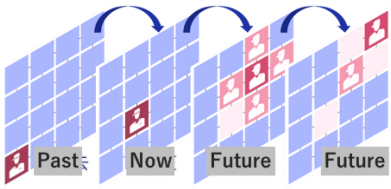
Logistics facilities have become larger and larger in recent years, and robots (AGVs and AMRs) are increasingly used to transfer goods inside them. On the other hand, there are many tasks that are difficult for robots to handle goods directly, for example picking and repacking goods, so manual labor is also indispensable. Therefore, workers and transfer robots coexist. Although there are some sites that separate the space for both workers and robots, it is desirable for both to be able to coexist safely in the same space to increase the efficiency of space utilization in the facility. In such cases, the trade-off between safety and efficiency becomes an issue. A typical transfer robot restrains its speed so that it can stop when an obstacle including workers approaches, and once it recognizes the obstacle, it stops. While this ensures safety, it inevitably reduces transfer efficiency. This trade-off can be resolved by utilizing probabilistic digital twin to predict the future location of the worker and control robots to consider the risk of collision and speed reduction. Each of the location prediction and control techniques is introduced in detail below.

However, sensors inevitably have blind spots, and there is a delay between detection of location of an obstacle and robot control, so obstacle location information at the time when the robot is operating is needed. To solve this problem, the presence or absence of obstacles at each time and point in the robot operation area is expressed as a probability, and based on the observed information, the condition of the blind spots and the future condition at each point are estimated as probabilities (See Figure 8). When estimating the probability, it is important to understand the relationship of obstacles in space and time. In other words, for moving obstacles, if the obstacle is within a certain distance in

the direction of movement from the point where it was observed at the previous time, the probability of its presence is high, but if it is further away than a certain distance, the probability of its presence is low. We represent such a spatial-temporal relationship between the presence and absence of obstacles as a conditional random field, CRF, and by mapping the observed values, we construct a model that predicts the future situation of obstacles from the current obstacle situation based on the observed values[10].



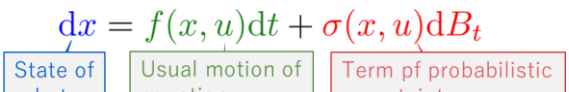
a: Probability of obstacle existence at each point (yellow is high, red is middle, blue is low)



b: Prediction of worker's future location

Figure 8 Prediction of future location of obstacle

After the location of future obstacles is estimated, it is necessary to control the route and speed of the transfer robot to travel safely and efficiently considering real-world uncertainties. Risk-sensitive stochastic control[11][12] solves these problems. In risk-sensitive stochastic control, the robot's motion equation is defined as a stochastic differential equation (see Figure 9-a) because it represents the uncertainties that affect the robot's motion, such as hardware degradation and ground conditions, as a model. An evaluation function is used to choose optimal control inputs, and we design it to evaluate both safety and efficiency, as well as to be sensitive to risk (see Figure 9-b). Although the value of the evaluation function will be a probability distribution because stochastic differential equation is used as equation of motion, it is possible to select the control that reduces both the value that the smaller is better and the variance as the optimal one. To determine the actual control inputs, various control inputs are prepared in advance, and the path and speed determined by solving stochastic differential equations are evaluated with the risk-sensitive evaluation function to select the optimal control (See Figure 9-c).



a: Stochastic differential equation

b: Risk-sensitive evaluation function

c: Safe and efficient path

Figure 9 Risk-sensitive stochastic control

4.3 Smart Sustainable Mobility

Today, the environment and mobility are major issues for many smart cities. Here we assume the following digital twin; the smart environment digital twin monitors air

pollution by collecting air quality data from observation stations, while restricting emissions at major sources when air pollution is expected to worsen; the smart driving digital twin monitors the driving environment of individual cars using in-vehicle sensors, while guiding driving maneuvers and travel routes according to changes of the environment. The eco-driving assistance, an application of digital twin orchestration owned by a city officer, aims to improve the city's environmental quality by recommending environment-friendly driving maneuvers to drivers and autonomous cars in areas with poor environmental quality. Based on the emission restriction plan simulated by the smart environment digital twin, the smart driving digital twin instructs the navigation system to perform driving maneuvers to control emissions. Furthermore, it enhances the air pollution prediction of the smart environment digital twin using environmental sensor data captured by the cars, which enables more effective eco-driving assistance.

Figure 10 show interactions between these digital twins through the orchestrator functions. The federation function shares the air pollution prediction model of the smart environment digital twin with the smart driving digital twin for federated learning using private data collected by individual car. The brokering function allows application to receive the emission restriction plan generated by the smart environmental digital twin, determines the restriction order for cars driving in the restricted area, and can send the order to the smart driving digital twins of the target cars. The translation function converts the environmental sensor data collected by the smart driving digital twin of individual cars to the format of observation data in the smart environment digital twin to import the "mobile" observation data for denser prediction of air pollution.

Implementation of the orchestrator framework is promoted for individual digital twin platforms as a common interface of inter-platform digital twin orchestration. The first implementation of the orchestrator framework and the use case is being conducted on NICT xData Platform[14] and Testbed. The framework implementation for IWON is also being discussed in IOWN Global Forum based on mapping the orchestrator functions to the IOWN Digital Twin Framework. In addition, integrated architecture of the orchestrator between physical space and cyber space is included in our future work.

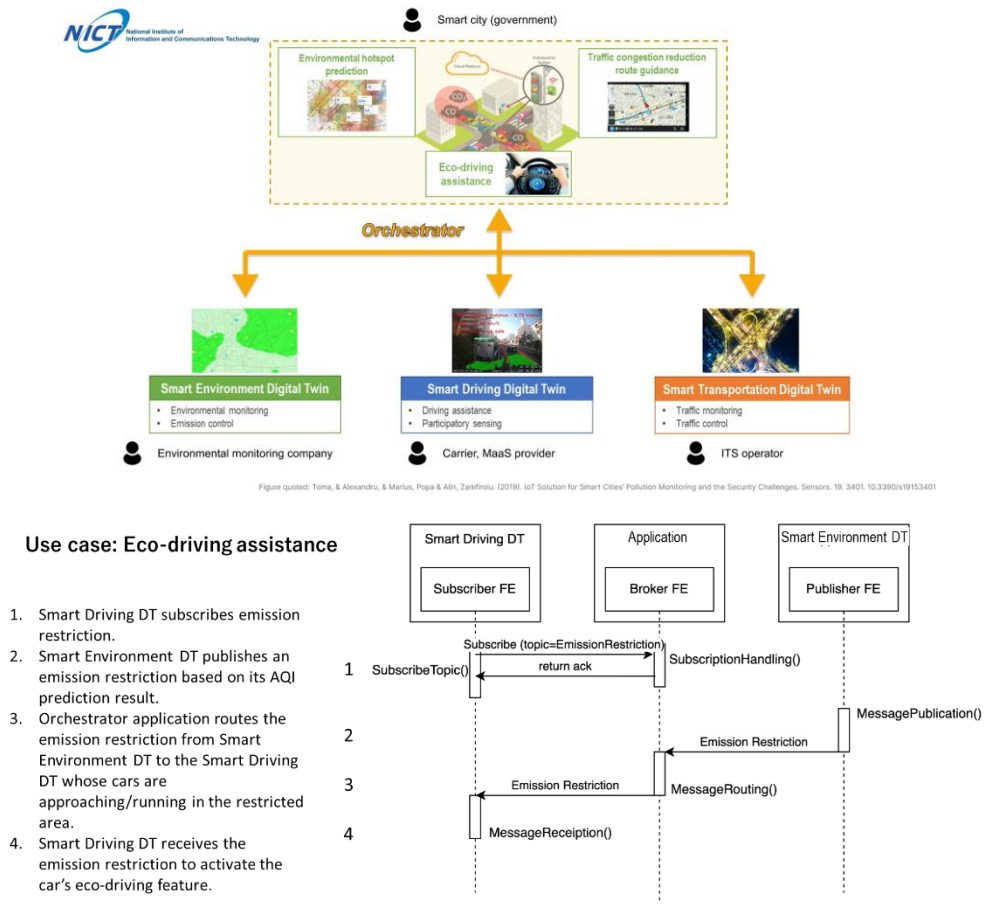


Figure 10: Smart sustainable mobility use case

5 Conclusion

In this article, we argued that a Digital-Twin can be digital representation of both real-world network objects. Based on this, we proposed a Digital-Twin architecture which manages various Digital-Twin instances in a common way so that any Digital-Twin applications can easily utilize them. We then introduced probabilistic Digital-Twin and cross-domain orchestration of Digital-Twins, as well as the use cases including radio communication environment, human-robot cooperation, and smart sustainable mobility.

Acknowledgements

This work was partly supported by MIC under a grant entitled R&D of ICT Priority Technology (JPMI00316). Also, this research results were partly obtained from the commissioned research (No.00701) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCE

[1] Grieves, Michael and Vickers, John. “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems”. 10.1007/978-3-319-38756-7_4., 2017.

[2] A. Fuller, et. al., "Digital Twin: Enabling Technologies, Challenges and Open Research," in IEEE Access, vol. 8, pp. 108952-108971, 2020.

[3] Y. Wu, K. Zhang and Y. Zhang, "Digital Twin Networks: A Survey," in IEEE Internet of Things Journal, vol. 8, no. 18, pp. 13789-13804, 2021.

[4] Xiaochen Zheng, et. al., “The emergence of cognitive digital twin: vision, challenges and opportunities”, International Journal of Production Research, 60:24, 2022.

[5] “Digital twin network – Requirements and architecture”, Recommendation ITU-T Y.3090, 2022.

[6] “Innovative Optical and Wireless Network Global Forum Vision 2030 and Technical Directions”, IOWN Global Forum, available at https://iowngf.org/wp-content/uploads/2023/03/IOWN_GF_WP_Vision_2030_2.0-2.pdf (Accessed: Jan. 2024).

[7] TMforum. available at https://www.tmforum.org/ (Accessed: Jan. 2024).

[8] Eclipse Ditto: Digital Twin framework of Eclipse IoT, available at https://eclipse.dev/ditto/index.html (Accessed: Jan. 2024).

[9] Toru Yamada "The Digital Twin and its Evolution from the International Standardization of Smart Cities.", A1-07, Interop Tokyo 2023 (2023)

[10] Y. Ohsita, S. Yasuda, T. Kumagai, H. Yoshida, D. Kanetomo, and M, Murata, “Spatio-temporal model that aggregates information from sensors to estimate and predict states of obstacles for control of moving robots,” IEICE Proceedings Series, 2022

[11] S. Yasuda, T. Kumagai and H. Yoshida, ”Cooperative Transportation Robot System Using Risk-Sensitive Stochastic Control,” 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp.5981-5988.

[12] S. Yasuda, T. Kumagai and H. Yoshida, ”Safe and Efficient Dynamic Window Approach for Differential Mobile Robots With Stochastic Dynamics Using Deterministic Sampling,” in IEEE Robotics and Automation Letters, vol. 8, no. 5, pp. 2614-2621, May 2023.

[13] Beyond 5G/6G white paper version 3.0, ISBN 978-4-904020-32-6, 2023, available at https://beyond5g.nict.go.jp/download/index.html (Accessed: Jan, 2024)

[14] NICT xData Platform, https://www.xdata.nict.jp/en/ (Accessed: Jan, 2024).

[15] Daiki Kodama, Kenji Ohira, Hideyuki Shimonishi, Toshiro Nakahira, Daisuke Murayama, and Tomoki Ogawa, “Enhancing Indoor Millimeter Radio Communication: A Probabilistic Approach to RSS Map Estimation”, in Proc. of IEEE Consumer Communications & Networking Conference (CCNC), 2024.

Optimum collaboration of network functions and computing resource

Minoru Matsumoto, NTT network service systems laboratories.
Hiroki Baba, NTT network service systems laboratories.
Shiku Hirai, NTT network service systems laboratories.

Abstract— In the Beyond 5G era, network operators will have higher flexible deployment methods of network functions via the virtualized RAN/core network, and the optimum usage method for distributed computing resource collaborated with network functions will be important. The reason is that high real-time/broadband communications (e.g., VR/AR) and AI inference will become popular applications, and require higher performance of arithmetic processing power using several types of computing resource (e.g., Memory, CPU, FPGA, and GPU).

To achieve these targets, the network architecture has been studying provided the virtual computing resource. The technology of network transport is employed APN (All Photonic Network), the computing resource deployed in different network domain (access network, MEC, core network, cloud) can be used as virtualized one computing resource via APN, so the network operator can use optimum computing resource on demand from requirements of communications and applications.

This paper reports evaluation its feasibilities of the proposed architecture compared to the existing CPU-based computing one.

1 Introduction

In the B5G era, “integration and cooperation” will progress in the four areas of “network and computing,” “cyber space and physical space,” “analog and digital,” and “mobile communications and fixed communications” as both communication services themselves. This progress in multifaceted integration and cooperation will generate an even greater need for end-to-end and seamless linking of information processing and information distribution for terminals, the network, and applications, which had been performed separately in communication services using the 5G network. This is expected to be led to new communication services in the form of Cyber Physical Systems (CPS), various environments that use AI integrated communications (cyber space and real space), and communications that overcome terminal/location limitations and functional limitations [1],[2].

This paper reports new concept of network architecture on optimum collaboration of network functions and computing resource in the B5G era.

2 Network Requirements and Technology Trends in the B5G era

A variety of use-case scenarios are proposed for features of B5G network such as extremely high data rates, ultra-low latency/jitter, and expanded coverage in several SDOs (Standards Developing Organizations). These use-case scenarios include ultra broadband communication, ubiquitous sensing, mission critical communication, universal coverage, ultra massive connection, and intelligent connection from viewpoints of network requirement [3]. To achieve these use-case scenarios, several SDOs have been studied technology trends in the B5G era shown in Table 1 [4],[5],[6].

Table. 1 B5G Technologies Trends from SDOs

1	Addition of computing and data services
2	Network simplification, protocol-stack reduction
3	Distributed cloud computing including that within the network
4	Integration and operation of multiple access technology methods
5	Transmission/data-exchange control in the core network to support ultra-low latency and high reliability
6	Wide-area time synchronization and deterministic communications
7	Enhanced security
8	Providing services with privacy-conscious user sensitive information
9	Improved robustness and resilience

3 Concept of Proposed Architecture

3.1 Background

Operators have already provided audio-voice, video content, and internet data as main communication services before the 5G era. Information-processing applications were deployed on servers on the cloud at a data center, and on terminals. Information-processing functions were centralized-located on the cloud and terminals, so that operators could provide in a form that minimized the amount of information exchanged between terminals and server over the network.

The data compression/encoding can make reduce the amount of information exchanged and the synchronization processing that performed between server and terminals interacted with each other over the network. However, computing resources and functions/processing-performance are generally varied for each terminal, so it needed to develop server applications on the cloud in accordance with network and terminal's functions/processing performance. Therefore, applications have any rest of problems that could be used limited by the difference of terminals' specification, and limited processing performance.

The network virtualization has been progressing since the 4G era, and the 5G architecture have been adopted cloud-native technologies such as Cloud-native Network Functions (CNFs) are being implemented to the 5th Generation Core (5GC) network. For low-latency applications, operators need to deploy server applications performing information processing at edge points serving as entrances to the network without going through the Internet using Multi-access Edge Computing (MEC) technologies [7]. Virtualized technologies have been used as a common foundation for MEC, and the virtualized Radio Access Network (vRAN) will also enable virtualized ones to be applied even to equipment installed closer to terminals. Therefore, the technology of commonality is expected to make MEC applications to be deployed not only on the edge of the network but also to a wide area including RAN where it can be shared in resources of the computing platform.

3.2 Concept of Proposed Architecture

The concept of proposed architecture can achieve a computing service by mixing and closely linking network functions, and application functions on the computing platform that has a ubiquitous presence on such a network. This computing service can provide to combine in a composite manner dispersed computing resources such as several kinds of computing functions (terminal, network, and cloud) and integrates the information-processing and communication functions required for services. In a network at the 5G era, a delay is generated from communications between terminals and server applications on the cloud, and associated information-conversion processing. However, if terminals and server applications for services on the cloud are to be consolidated in an exclusive computing platform, these communications/processing can become unnecessary make real-time interaction.

The proposed network architecture provides distributed computing functions, so that various connections methods can be supported such as ones using optical wavelengths, and the conventional Virtual Private Network (VPN) connection on an All-Photonics Network (APN). In specially, an APN is expected to make it possible to integrate computing functions to perform communications and run information-processing applications with higher level of performance and lower power consumption, that is difficult to achieve by existing protocols (e.g., TCP/IP). Moreover, the proposed network architecture can use suitable volume of hardware accelerators according to requirements of the information-processing applications and network functions to perform higher data rate and lower latency processing that combines these applications and network functions in a composite manner. Therefore, the proposed network architecture can be provided higher performance for all network regions. The comparison of network architectures in different generations is shown in Fig. 1 [8].

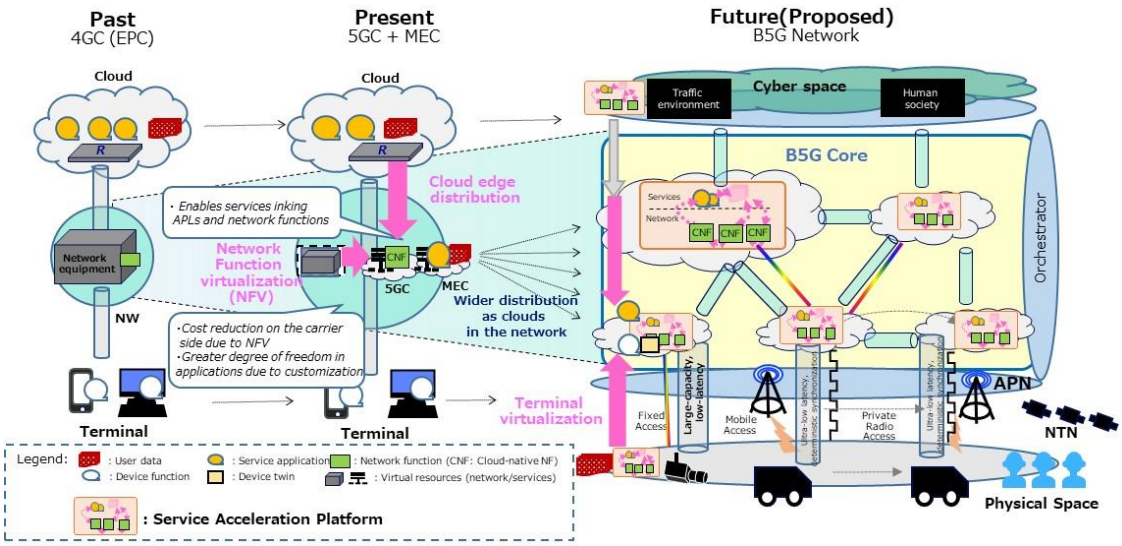


Fig. 1 Comparison of network architectures in different generations

4 Proposed Network Architecture

The proposed network architecture can make use of both terminal and cloud computing resources in addition to those of the network, and configure cloud/terminal application functions and network ones in an integrated manner. The computing service functions of this network architecture are to provide computing resources to terminals, and are being studied for the B5G era. The communication services have been provided by the network up to the 5G era, but these are expected to achieve on a distributed computing platform in the B5G era (shown in Fig. 2).

The proposed network architecture can also provide Control-plane function to control the communication session collaborated with computing functions as a computing service (shown in Fig. 3). And, it can also provide Data-plane function to add conventional User-plane function that transfers audio-voice, video, and data as a computing service. In the B5G era, it can also be expected to provide a service-mesh function to enable Data-plane to process multiple microservices network-wide in a chain-like manner within the distributed computing method by using a variety of microservices.

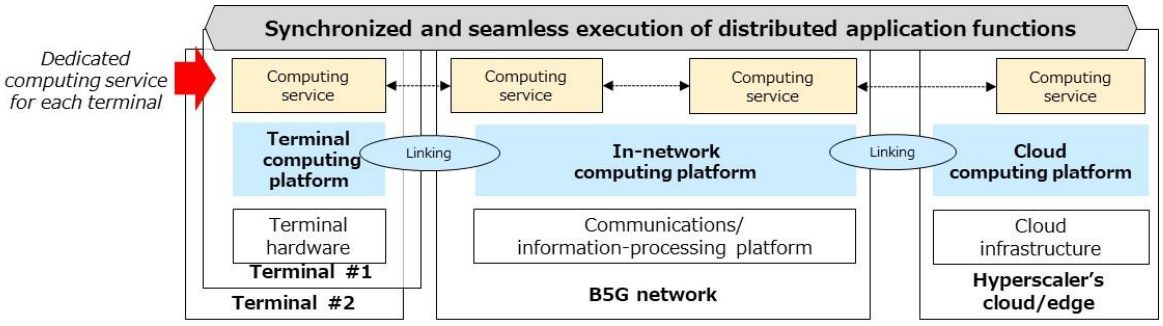


Fig. 2 Distributed computing platform in the B5G era

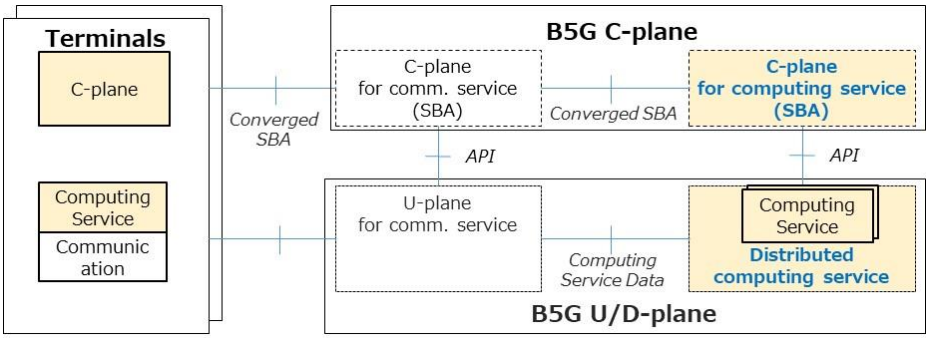


Fig. 3 C-plane and U/D-plane in the B5G era

5 Evaluation

5.1 Assumed Use-Case

This paper assumes one use-case which relays ultra-high-resolution video in real time or deliver 3D video in the metaverse (shown in Fig. 4). These use-cases are a delivery service for ultra-high-resolution video or 3D content with performing advanced information processing such as video synthesis or 3D rendering based on the user's viewing conditions (e.g., terminal capabilities, resolution, and viewing angle). In this situation, the terminal needs to support advanced video processing functions and/or hardware-level performance to play back 3D content or render high-resolution video according to viewing conditions.

On the other hands, it would be desirable to provide abovementioned services to all kinds of terminals including smartphones that are limited in functionality and performance beyond any locations and functional limitations from customer's viewpoints.

To achieve customer's demands, this paper considers a method to carry out video processing, 3D-content rendering, and customized functions to be executed on the terminal on a system other than the terminal such as network or cloud, and instantly display processing results on the terminal. This method makes release the limitations in functions or performance such as a small screen to handle a new service independent of terminal functions and performance as in the case of a terminal incapable of 3D displays. If a terminal has relatively high performance capable to display 3D video and service requirements are defined to deliver 3D video with low-latency in real time at maximum resolution, the transmission of materials from the cloud can be synchronized between memory devices. Moreover, it can be also performed both advanced storage access, and synthesized on a GPU for 3D video. This architecture can provide the video delivery using an uncompressed, real-time transmission protocol exploiting the features using a protocol-independent abstract API layer without considering the upper-level application.

Therefore, it can provide an environment that facilitates to develop high-definition 3D video applications.

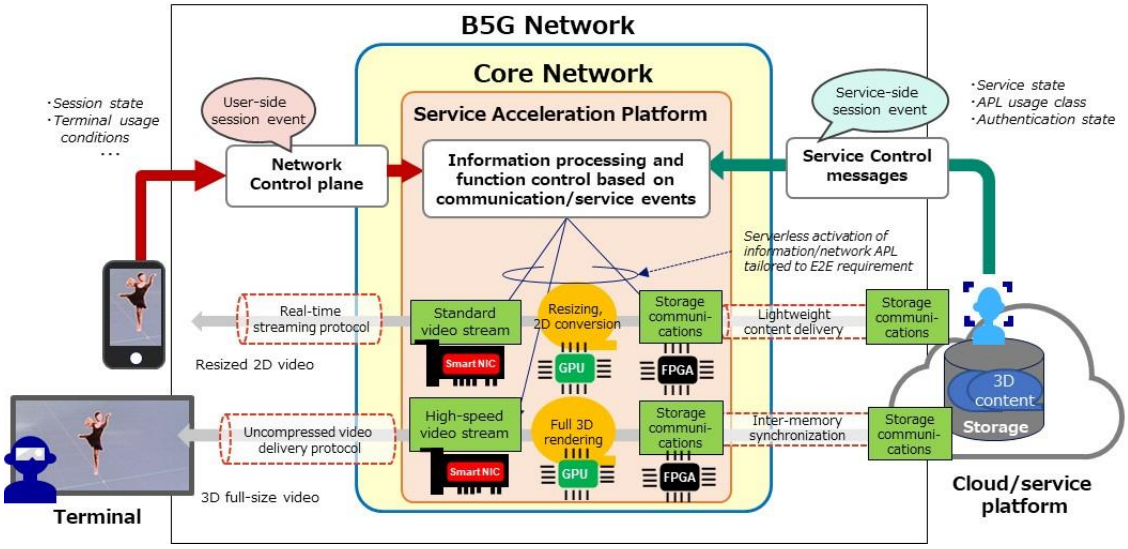


Fig. 4 Assumed Use-Case (low-latency, real-time delivery of 3D video to multi-types of terminals)

5.2 Evaluation Results

This paper simulates abovementioned assumed use-case via the proposed network architecture is to connect and chain multiple workloads on multiple accelerators (Smart NIC, GPU, FPGA and etc.) over network or computer bus connection. And this paper evaluates validities of proposed “virtualised accelerator pool” between different network domain via accelerating End to End data transfer by offloading applications and network workload, and chaining those workloads with hardware connection as shown in Fig. 5.

Evaluation results are three points (throughput, latency, and PDV: Packet Delay Variation) compared to the existing CPU-based model as shown is Fig. 6. Each result is confirmed extremely improved compared to CPU-based model.

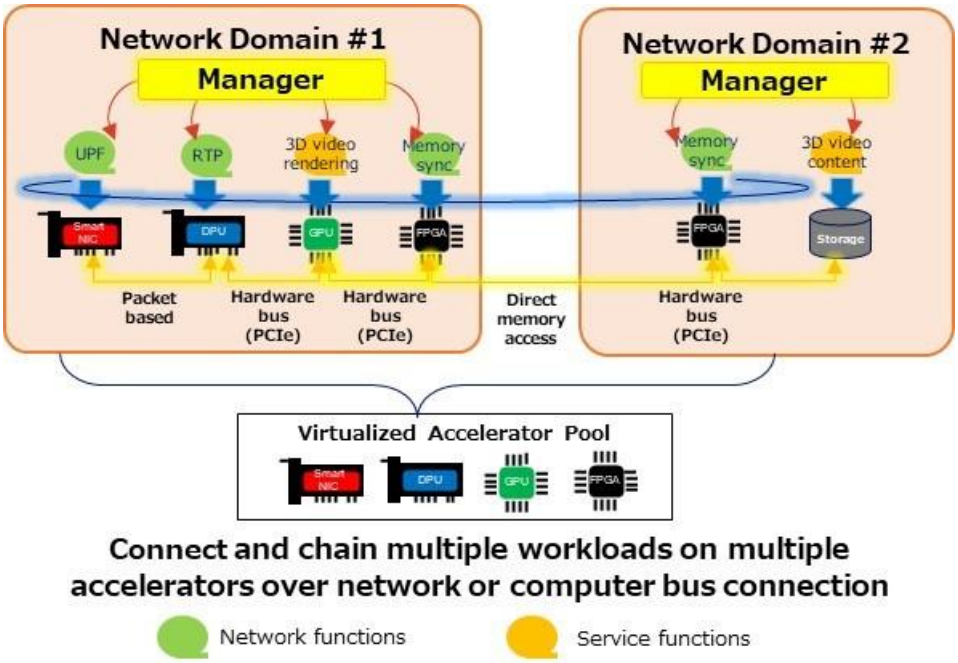


Fig. 5 Proposed "virtualised accelerator pool" between different network domain

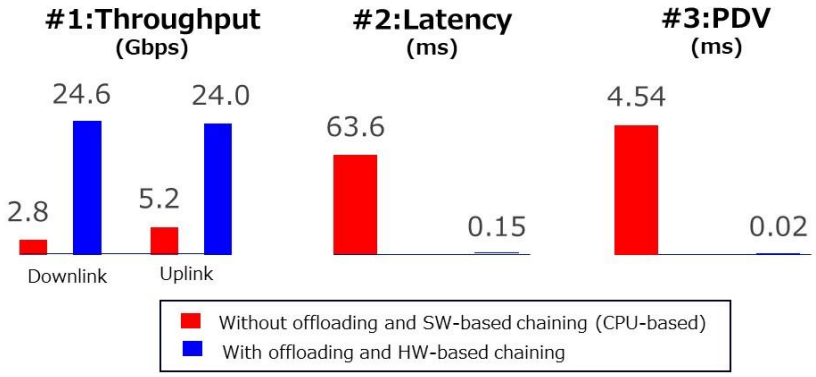


Fig. 6 Evaluation results compared to CPU-based model

6 Conclusion

This paper proposes the network architecture in the B5G era which achieves a computing service by mixing and closely linking network functions, and application functions on the computing platform that has a ubiquitous presence on such a network, and evaluates its feasibilities compared to the existing CPU-based computing one. Evaluation results are confirmed to be extremely improved on these performances as throughput, latency, and PDV on proposed network architecture.

REFERENCE

[1] Next G Alliance, “Next G Alliance Report: Digital World Experiences,” December 2022. https://www.nextgalliance.org/white_papers/digital-world-experiences

[2] Hexa-X, “Deliverable D1.1 6G Vision, use cases and key societal values,” February 2021. https://hexa-x.eu/wp-content/uploads/2021/02/Hexa-X_D1.1.pdf

[3] Beyond 5G Promotion Consortium, “Beyond 5G White Paper—Message to the 2030s—Version 2.0,” March 2023. https://b5g.jp/doc/whitepaper_en_2-0.pdf

[4] Next G Alliance, “Next G Alliance Report: 6G Technologies,” June 2022. https://www.nextgalliance.org/white_papers/6g-technologies

[5] Hexa-X, “Deliverable D5.2 Analysis of 6G architectural enablers’ applicability and initial technological solutions,” October 2022. https://hexa-x.eu/wp-content/uploads/2022/10/Hexa-X_D5.2_v1.0.pdf

[6] IOWN Global Forum, “Technical Outlook for Mobile Networks Using IOWN Technology,” Jan. 2022. https://iowngf.org/wp-content/uploads/formidable/21/IOWN-GF-RD-Technical-Outlook-for-Mobile-Networks-1.0-1.pdf

[7] https://www.verizon.com/business/solutions/5g/edge-computing/public-mec/

[8] https://www.rd.ntt/e/ns/inclusivecore/whitepaper_ver2.html

User-centric Network

Akio IKAMI, KDDI Research, Inc.
Masayuki KURATA, KDDI Research, Inc.
Masaki SUZUKI, KDDI Research, Inc.
Hiroyuki SHINBO, KDDI Research, Inc.
Atsushi TAGAMI, KDDI Research, Inc.
Yoji KISHI, KDDI Research, Inc.

***Abstract*—This paper provides an overview of a "User-centric network", an innovative approach designed to provide each user with a stable, fast, and continuous communication service anytime and anywhere. This network architecture adopts Cell-Free massive MIMO technology to realize personalized access from the radio access network (RAN) side. The principles underlying the concept of "User-centric" are further extended to the core network (CN) side, and the functionality of the RAN and CN are converged to improve the communication quality for each user. In addition, this section addresses several technological challenges towards 6G.**

1 Introduction

In the era of 6th Generation (6G) mobile networks, there is an imperative need to address increasingly diversified requirements. In addition, the necessity for sustainability is becoming more prominent. To meet these requirements, customizing mobile networks according to each user's intent is essential. In this paper, we propose a “User-centric network” that optimizes mobile networks for each user and converges the radio access network (RAN) and core network (CN). This user-centric and convergence approach aims to provide a more tailored and sustainable solution for the diverse needs of the 6G era, paving the way for a more efficient and user-focused mobile network infrastructure.

The target use cases have wide global recognition and are introduced in Hexa-X [1] and Next G Alliance [2], which derive the requirements envisioned in User-centric networks. These include robot/human cooperation, unmanned aerial vehicles (UAVs), artificial intelligence (AI) utilization, and eXtended reality(XR). The following are the extracted challenges.

(I) Continuous

Ensuring safety is one of the most essential requirements. For example, it needs mission critical applications in various locations over 6G such as constant monitoring and control of robots. This implies a strong need for consistent, high radio quality anywhere and anytime. In addition, since communication requirements vary for each communication, they need to

archive based on user intent. However, 5th Generation (5G) systems have cell-edge issues due to increasing path loss and inter-cell interference, leading to the degradation of radio quality at cell fringes.

(II) Fast

As the baseline concept in the present 5G core networks, user-plane (U-plane) network function (NF), e.g., U-plane function (UPF) instances are potentially deployed at the edge sites to reduce the transmission delay between terminals and applications (APPs), which provides a better quality of experience (QoE) and quality of service (QoS). Terminals include a smart phone, a robot, a UAV and many kinds of IoT devices, but for simplicity, we refer to all of these as user equipment (UE) in this paper. In this case, the other U-plane and control plane (C-plane) NF instances are deployed at the central site. However, signaling for procedures in the C-plane should also be completed quickly for further improvement in QoE and QoS.

(III) Stable

With the increase in the number of UEs simultaneously connected to mobile networks and the frequent utilization of AI in mobile networks, i.e., analytics generation by O-RAN RAN intelligent controller (RIC) and network data analytics function (NWDAF), the number and amount of signaling for C-plane procedures is predicted to increase explosively. However, the 5G system is not designed to perform load-balancing for each NF, and thus cannot process C-plane signaling efficiently. For example, to realize the interaction between RAN and CN in the 5G system, C-plane signaling goes through the central unit (CU) and access and mobility management function (AMF), which easily increases the load on the CU and AMF. In a worst-case scenario, the inappropriate handling of C-plane signaling can cause congestion in the C-plane, also known as a signaling storm [3]. Therefore, the mobile network needs to accommodate heterogeneous connections and an enormous amount of signaling to support stable communications.

2 User-centric network

2.1 Overview

We aim to realize the above continuous, fast, and stable communication to provide the user with a highly satisfying service anytime and anywhere to meet the requirements of the use cases and satisfy user intent. As shown in Fig. 1, a User-centric network is designed to be deployed across three sites, i.e., antenna, edge, and central sites. In the antenna site, multiple access points (APs) are installed for Cell-Free massive MIMO [4]. In the edge site and central site, each site incorporates a cloud infrastructure in central

and edge sites as computing resources, and network function instances are deployed across those clouds. AP and NF instances are physically isolated for each user or group of users to prevent failures in a network from spreading to other networks and to enhance reliability and resilience. Each network facilitates the dynamic placement of APPs and NFs between central and edge sites to achieve fast communications in both the U-plane and the C-plane. NFs handle not only CN functions but also RAN functions such as a virtual central processing unit (vCPU) as radio processing of Cell-Free massive MIMO. The rest of this section introduces the essential concepts for realization of a User-centric network.

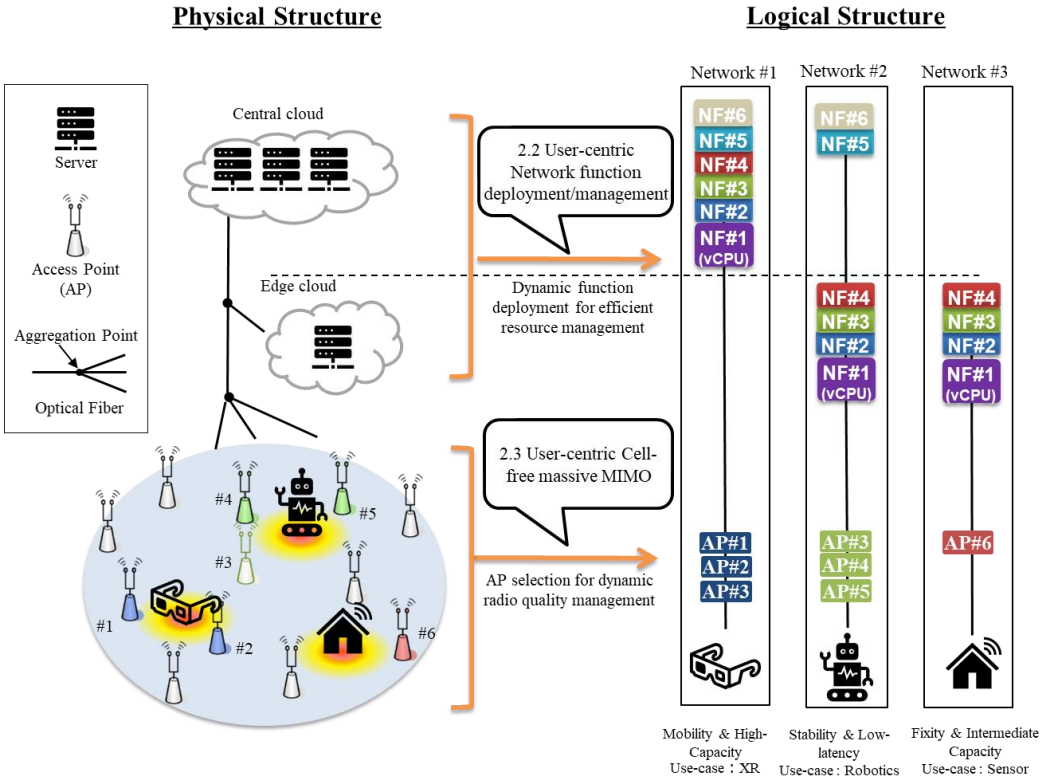


Fig. 1 Overview of User-centric network

2.2 User-centric network function deployment and management

Cloudification of RAN and core NFs will become a native feature in the 6G era, and all NFs are deployed in computing resources such as O-cloud of O-RAN. However, this deployment of NFs on physically distributed servers raises several challenges, e.g., mechanisms to resolve changes in demand so that the resources of the virtual infrastructure can be utilized effectively, and minimization of signaling between distributed clouds. To address these challenges, we have considered a user-centric network function deployment approach as shown in Fig. 1. In this approach, NFs are dynamically deployed based on the user's demand and resource status. Specifically, NFs

are placed according to a physical hierarchical structure. This approach is expected to provide several benefits: reducing of signaling between sites by aggregating NFs, saving power consumption by launching only the minimum necessary NFs for each service, and placing them close to users, i.e., in the edge cloud, to reduce latency. In the 5G system, while the applications and U-plane NFs, e.g., UPF, are supported to be located at edge sites [5], C-plane NFs are still deployed at the central site. When the C-plane procedures, e.g., service requests and handover, are performed, the latency requirements can be violated due to transmission delay from/to the central site. Therefore, C-plane NFs should also be located at the edge site. Additionally, the upstream signaling can be reduced by locating the C-plane NFs at the edge site. For example, Hexa-X [1] introduces the use case where UPF instances are increased/decreased adaptively and predictably to avoid edge site congestion caused by excessive access to applications.

In addition to these strategies, network isolation per user or group of users helps to prevent the spread of failure. If the mobile core network is physically isolated, a failure that occurs in one NF instance in one network does not affect other NF instances on other networks. Network isolation also enables prompt recovery from the failure. In a legacy system, the cause analysis and restoration work are likely to be time-consuming. Our approach assumes that the isolated networks can be easily refreshed and immediately restarted, even without any root-cause analysis. The root-cause analysis can be performed independently of the recovery process.

2.3 User-centric RAN with Cell-Free massive MIMO

In the 6G era, there is a growing need for stable quality assurance in areas such as mission critical applications. Providing the necessary radio quality anytime, anywhere is also becoming increasingly important. However, in 5G, radio quality is highly dependent on the user's location due to cell-edge issues. This results in low radio quality at cell-edge areas, decreased signal power due to increased path loss, and increased interference from neighboring base stations. This cell-edge issue makes it difficult to maintain good radio quality due to movement, which is a bottleneck for QoS and QoE per user. To assure radio quality anytime, anywhere, User-centric RAN with Cell-Free massive MIMO technology is being studied in [6] and [7]. This technology aims to provide appropriate radio quality for the whole area according to user mobility and services, such as by spotlighting each user. This user-centric approach can balance radio quality assurance and power consumption in RAN, achieving sustainability. Cell-Free massive MIMO technology is gaining attention, as it has been demonstrated to enable user-specific wireless area construction through AP clustering and interference removal technology through vCPU cooperation [6]. To manage these new air interface technologies, distributed RICs and other management techniques are necessary.

2.4 Isolated RAN/CN which exists per user/user-group

Based on the motivation of creating a User-centric network, CN should also be enhanced toward for each user or for each user-group architecture. To realize the architecture, four essential ideas are introduced in this section: (a) Service-based architecture (SBA) extension to RAN, (b) Lightweight NF, (c) NF componentization, and (d) stateless NF. An overview of the essential concepts is shown in Fig. 2.

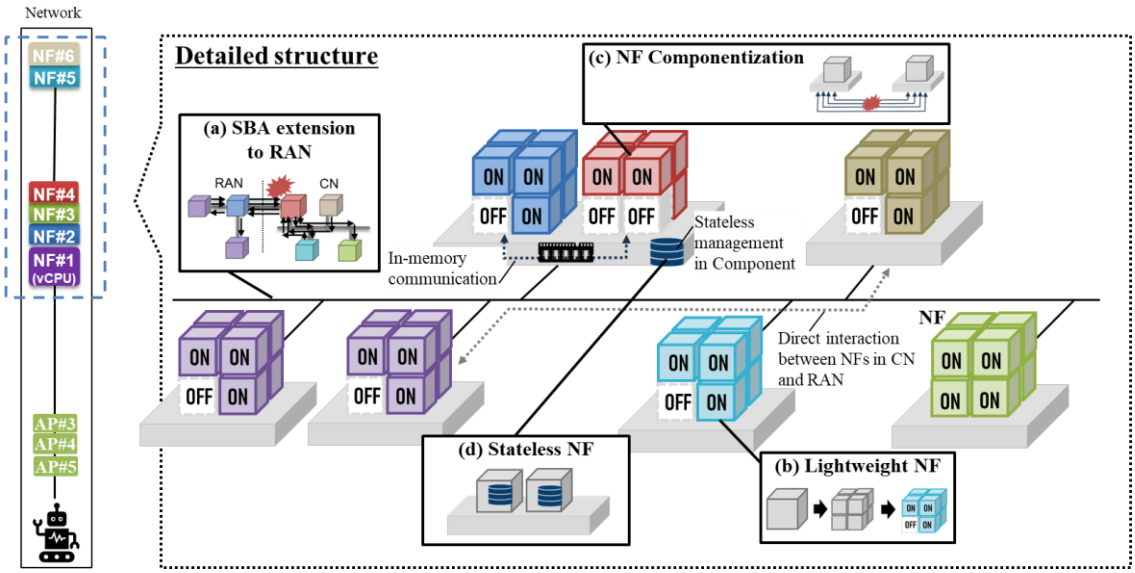


Fig. 2 Overview of each isolated network which exists for each user or user-group

(a) SBA extension to RAN is an essential enabler for efficient interaction between the RAN and CN [8]. In the current architecture, the access is likely to concentrate at AMF, resulting in a bottleneck. If both the RAN and CN are converted to SBA, instances of network functions can directly interact with each other without proxy functionality. Additionally, full-SBA eliminates specific interfaces and protocols between the RAN and CN, which means that the next generation application protocol (NGAP) over the stream control transmission protocol (SCTP), i.e., N2 session, is not needed anymore. Reducing the variation of protocols contributes to easier deployment and operational verification in cloud-native environments, especially in the case where generic equipment is used.

(b) Lightweight NF is a technology for reducing the image size of an NF instance. For smoother lifecycle management, in an isolated network, the functions of NF instances are carefully selected according to the requirements of users and services. For example, if a group of UEs do not need mobility management, e.g., sensors, etc., the NF instances turn off the mobility management function. By dropping subfunctions from the NF, the image size of an NF instance is reduced for fast instantiation and termination. However,

the choice of subfunction from NF leads to the multiple generation of NF images. The mechanism to handle different versions is a matter for future study.

(c) NF componentization reduces HTTP signaling between NF instances. Excessive C-plane signaling can lead to overloading or congestion, called a signaling storm [3], which degrades user QoE and QoS. "Component" is introduced to drastically mitigate the signaling load. First, it groups NFs that frequently exchange signaling messages with each other and (re)locates them on the same node. Then, HTTP messages between NFs on the same node are replaced with interaction via shared memory. In addition to reducing the number of messages, in-memory communication eliminates the HTTP parsing and de-parsing processes, reducing network header overhead and decreasing kernel processing latency [9].

(d) Stateless NF is the removal of context from an NF instance. In the 5G system, AMF and session management function (SMF) are stateful and capable of managing the contexts of UE and session. In the context of virtualization, NFs are supported to be scaled out/in for low power consumption and load balancing. However, it is difficult to scale out/in stateful NFs flexibly and quickly because the contexts need to be properly handled. Therefore, we remove the context from NF itself by storing context in the common storage, assuming componentized NF instances.

3 Research activities to realize a User-centric network

We describe the activities of KDDI Research, Inc. to realize a User-centric network. We have started evaluations of the RAN for a User-centric network (User-centric RAN) by conducting computer simulations and using a testbed with the element technologies: AP clustering technology [10] and CPU cooperation technology [7]. AP clustering is a method in Cell-Free massive MIMO that selects cooperating APs for each user to communicate, thereby ensuring wireless quality and reducing the computational load of wireless signal processing. CPU cooperation is a method that coordinates virtualized CPUs distributed across different buildings to achieve both interference suppression effects and reduction in computational load for wireless signal processing.

The simulation results for the element technologies are shown in Figure 3. Compared to traditional cellular networks, it is evident that a User-centric network can provide higher quality over a wider area. Furthermore, we discovered that better quality can be provided by combining AP clustering technology with CPU inter-cooperation. In addition, we have constructed an initial testbed with commercial smart phones as shown on the left side of Figure 4. In the testbed, we evaluated IP-level data communication between terminals and servers with AP clustering technology [11]. As shown on the right side of Figure 4, the throughput of the User-centric network is stable compared to the cellular network.

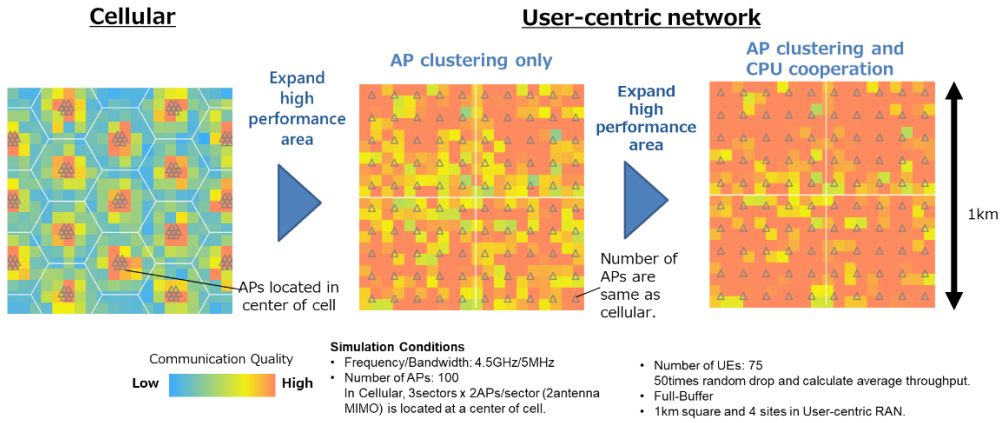


Fig. 3 Results of simulation comparison with elemental technologies

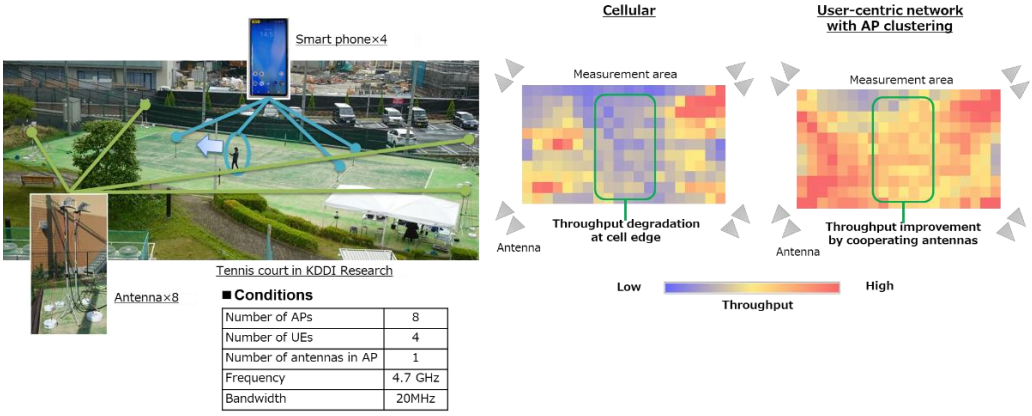


Fig. 4 Experimental results in initial User-centric RAN testbed

4 Conclusion

In this paper, we proposed a "User-centric network" that offers stable, fast, and continuous communication anytime and anywhere by optimizing mobile networks for each user and integrating the RAN and CN. Our approach ensures that the mobile network can accommodate heterogeneous connections and an enormous amount of signaling, supporting stable communications. In addition, KDDI Research has investigated User-centric RAN with Cell-Free massive MIMO in testbeds utilizing element technologies such as AP clustering technology and CPU cooperation technology. Moving forward, KDDI Research will continue research and technology development towards 6G by combining various element technologies (Section 2) of the User-centric network.

ACKNOWLEDGMENTS

The research results of User-centric RAN are obtained from the commissioned research (JPJ012368C00401) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

[1] Hexa-X. Accessed: Aug. 29,2023. [Online]. Available: https://hexa-x.eu/

[2] Next G Alliance. Accessed: Aug. 29,2023. [Online]. Available: https://www.nextgalliance.org/

[3] Ahmad, Ijaz, et al. "Overview of 5G security challenges and solutions." IEEE Communications Standards Magazine 2.1 (2018): 36-43.

[4] H. Q. Ngo et al., "Cell-free massive MIMO: uniformly great service for everyone," in 2015 IEEE 16th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC), Jun. 2015, pp. 201–205.

[5] Filali, Abderrahime, et al. "Multi-access edge computing: A survey." IEEE Access 8 (2020): 197017-197046.

[6] T. Murakami et al., "Analysis of CPU Placement of Cell-Free Massive MIMO for User-centric RAN," NOMS 2022 IEEE/IFIP Network Operations and Management Symposium, 2022, pp. 1-7

[7] A. Ikami et al., "Interference suppression for distributed CPU deployments in Cell-Free massive MIMO", the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall).

[8] Choi, Jinho, et al. "RAN-CN Converged Control-Plane for 6G Cellular Networks." GLOBECOM 2022-2022 IEEE Global Communications Conference. IEEE, 2022.

[9] Jain, Vivek, et al. "L25gc: A low latency 5g core network based on high-performance nfv platforms." Proceedings of the ACM SIGCOMM 2022 Conference. 2022.

[10] Y. Tsukamoto, et al. "User-centric AP Clustering with Deep Reinforcement Learning for Cell-Free Massive MIMO," In Proceedings of the Int'l ACM Symposium on Mobility Management and Wireless Access (MobiWac 2023).

[11] KDDI research Inc. Accessed: Aug. 29,2023. [Online]. Available: https://www.kddi-research.jp/english/newsrelease/2023/052302.html

Intent-based operational plan generation for business utilization of autonomous networks

Takayuki Kuroda (NEC)

Abstract— **Intent-based autonomous network operation technologies are attracting attention toward advanced automation of network operations. However, there is concern that the lack of a means for users to check the behavior of automatic operations in advance in conventional technologies is a barrier to their practical application. Therefore, we propose a method for generating network operational plans from intent. This allows users to activate the operational plan after checking its validity. In this article, we describe the functionality of the proposed technique, followed by a brief overview of the method.**

1 Introduction

Increasingly complex networks are becoming difficult to provide quickly and stably by hand, requiring a high degree of automation of operations [1][2]. The intent-based approach is promising as a fundamental approach to automate network operations [3]. Intent is information that expresses requirements in an abstract and declarative manner. According to intent-based automation technology, a machine interprets the intent and performs the construction and operation of the network. This allows users to easily build the desired network by simply entering high-level requirements without having to enter detailed information. However, conventional technologies do not provide a means for users to confirm the automatically determined behavior of operation and management in advance, and there is concern that this may be a barrier to the practical application of autonomous networks. Therefore, we propose a method to automate operation by generating network operational plans from intents and applying them after checking their validity. In this paper, we describe the functionality of the proposed technique, followed by a brief overview of the method.

2 Current intent-based autonomous network architectures and challenges

In current intent-based network architectures, the construction and operation phases are in one cycle. Figure 1(a) shows the architecture described in [4]. In this approach, the intent is translated into the network configuration; when the automated operation function receives the intent from the user, it translates the intent and activates it immediately. When a problem is detected by assurance, a response is considered according to the intent.

There are two problems with this architecture. First, in this architecture, problematic events are observed in assurance and then addressed in translation. Preferably, at the

time of initial activation, configuration and policies should be planned in anticipation of changes that may occur during operation. Second, users are not provided with a means to know in advance the range of changes that the target network can handle during operation and how to deal with them. In actual operation, the decision to start operation is made only after clarifying the operational plan and confirming its safety and adequacy in advance before deployment. Otherwise, unexpected costs will be incurred and operations will be black boxed, making it impossible to respond to unexpected problems.

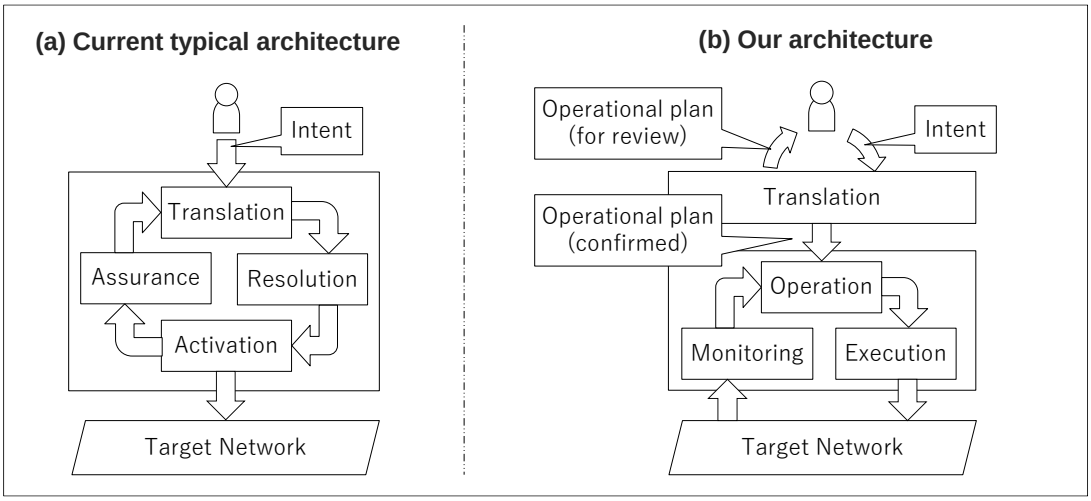


Fig. 1 Comparison between existing intent-based autonomous network architecture and ours.

3 Proposed Architecture

Figure 1(b) shows the architecture of our proposed intent-based autonomous networking. In our method, intent is translated into operational plan, which includes the configurations of the network to be built, the expected changes in the network and their scope, and the actions to be taken. The phases of using this method are divided into a planning phase and an operational phase. In the planning phase, the user repeatedly adjusts the intent and reviews the operational plan to create an appropriate one. Then, in the operation phase, the user instructs to deploy and start operation of the network based on the confirmed operational plan.

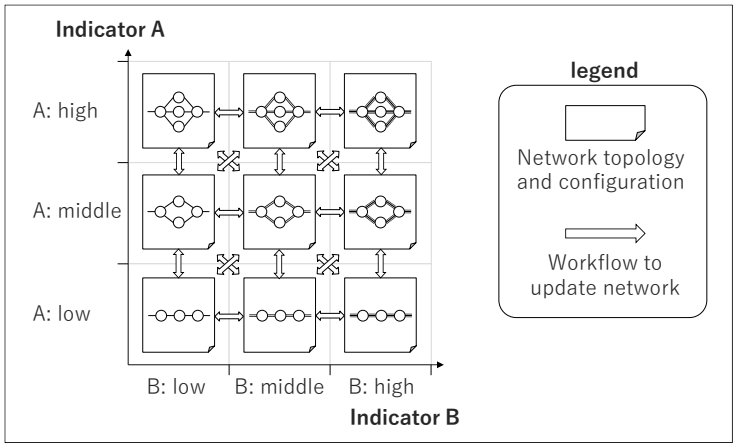


Fig. 2 Concept of operational plan.

Figure 2 shows the concept of operational plan. The operational plan describes the indicators specified by the user, the appropriate network configuration for each interval of the value of the indicator (hereafter, this interval is called the state), and the workflow for changing from each network configuration to the other network configuration. For example, the ideal configuration when both indicator A and indicator B are "low" is shown in the lower left state. The ideal configuration when the value of indicator A is "middle" and the value of indicator B is "low" is shown in the middle left state. By referring to the operational plan, the user can check the operational behavior in detail.

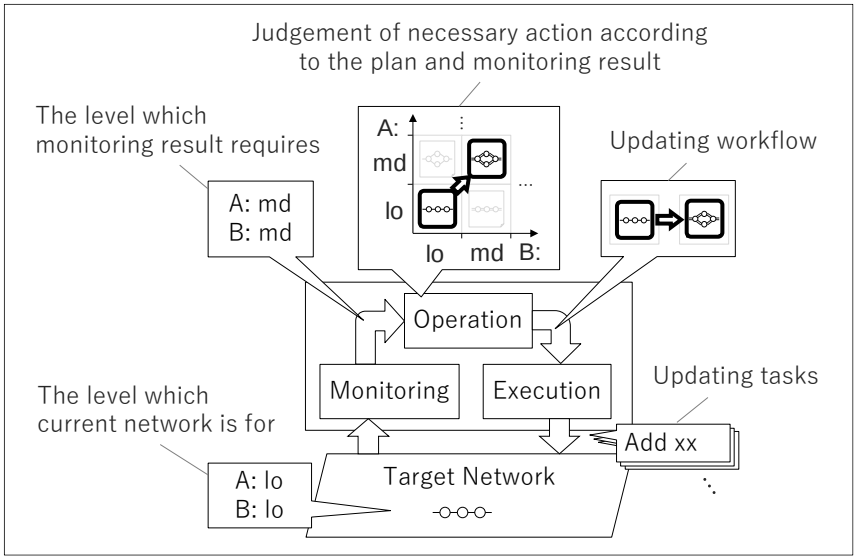


Fig. 3 Automated operation flow with using operational plan.

Figure 3 shows the automated operation flow with using operational plan. The automated operation function consists of three modules: monitoring, operation, and execution. First, monitoring reveals the new ideal state, while operation refers to the

operational plan and extracts workflows to change the current configuration to the ideal configuration if the ideal state is different from the current state. In Execution, the target network is updated to the desired configuration by executing the extracted workflows.

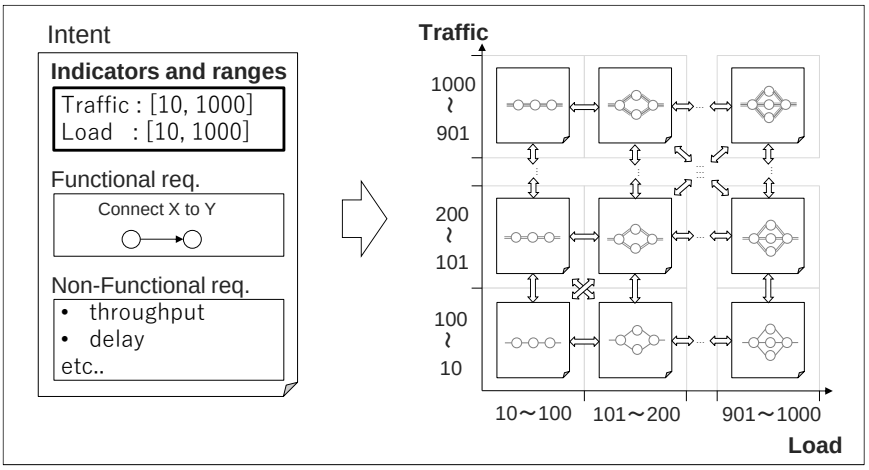


Fig. 4 Generation of operational plan from intent.

In the planning phase, the translation function of the intent generates an operational plan from the intent. This is shown in Figure 4. In addition to functional and non-functional requirements, the intent in this method specifies the indicator to be monitored and the range of values to be supported. Using the aforementioned intent, the translation function generates an operational plan as follows. First, by dividing the range of values of an indicator into multiple intervals, a state consisting of multiple indicators is generated. Next, the network configuration is automatically designed for each state. In the automatic design, the value of the indicator in the state is added to the non-functional requirements, and the function to monitor the value of the indicator is added to the functional requirements. Finally, for each generated configuration pair, a workflow is generated to transition between configurations. Automatic network configuration and workflow generation based on intent can be achieved by other results of the author's research group [5][6].

4 Conclusion

This paper introduces intent-based operational plan generation technology for business utilization of autonomous networks. This technology allows users to safely execute operations by knowing the contents of operations in advance. In the future, we will continue to refine and put the technology to practical use, as well as develop a method for explaining the contents of an operational plan more concisely.

Acknowledgements

These research results were obtained from the commissioned research(No.JPJ012368C04801) by National Institute of Information and Communications Technology (NICT) , Japan

REFERENCE

[1] TM Forum. Autonomous Networks: Empowering Digital Transformation For Smart Societies and Industries. TMForum White Paper, 2020.
[2] ETSI, “Intent driven management services for mobile networks”, TS 128 312 V17.0.1 (3GPP TS 28.312 version 17.0.1 Release 17), Jul. 2022.
[3] A. Clemm, L. Ciavaglia, L. Granville, and J. Tantsura, Intent-Based Networking Concepts and Definitions, ITU, Geneva, Switzerland, Feb. 2021.
[4] A. Leivadeas and M. Falkner, "A Survey on Intent-Based Networking," in IEEE Communications Surveys & Tutorials, vol. 25, no. 1, pp. 625-655, Firstquarter 2023, doi: 10.1109/COMST.2022.3215919.
[5] T. Kuroda, Y. Yakuwa, T. Maruyama, T. Kuwahara, K. Satoda, “Automation of Intent-based Service Operation with Models and AI/ML”, The Intelligence Provisioning for Network and Service Management in Softwarized Networks workshop (IPSN) 2022
[6] T. Kuroda, M. Nakanoya, A. Kitano, A. Gokhale, “The Configuration-Oriented Planning for Fully Declarative IT System Provisioning Automation”, IEEE/IFIP Network Operations and Management Symposium (NOMS) 2016

Task-Oriented 6G Native-AI Network Architecture

Peng Chenghui, Huawei Technologies
Wang Jun, Huawei Technologies
Yang Yang, The Hong Kong University of Science and Technology
Koshimizu Takashi, Huawei Technologies Japan

Abstract— **The vision for 6G networks is to offer pervasive intelligence and internet of intelligence, in which the networks natively support artificial intelligence (AI), empower smart applications and scenarios in various fields, and create a "ubiquitous-intelligence" world. In this vision, the traditional session-oriented architecture cannot achieve flexible per-user customization, ultimate performance, security and reliability required by future AI services. In addition, users' requirements for personalized AI services may become a key feature in the near future. By integrating AI in the network, the network AI has more advantages than cloud/MEC AI, such as better QoS assurance, lower latency, less transmission and computing overhead, and stronger security and privacy. Therefore, this article proposes the task-oriented native-AI network architecture (TONA), to natively support the network AI. By introducing task control and quality of AI services (QoAIS) assurance mechanisms at the control layer of 6G [1].**

1 Introduction

This explains the needs of Native-AI based 6G Wireless Network Architecture and lists of reason that requires to shift to task-oriented system mechanism. The proposed NW architecture called Task-Oriented native-AI network architecture (TONA), to natively support the network AI that create a "ubiquitous-intelligence" world. Reflecting the proceeding transformation, this article further proposes TONA to meet personalized AI service demand and requirements. This article mainly:

(1) Introduces three-layer logical architecture of task management and control system, and designs the task lifecycle management procedures, which include the collaboration of multi-dimension heterogeneous resources (communication, computing, data, and algorithm) and multi-node at the control layer.

(2) Defines task-specific QoAIS indicators for the mapping from Service Level Agreement (SLA) indicators — e.g., service requirement zone (SRZ) and user satisfaction ratio (USR) — to QoAIS indicators, and discusses task-level QoS assurance to meet individual requirements of different users.

(3) Compares the network AI and cloud/mobile edge computing (MEC) in terms of QoAIS indicators. Thanks to providing the AI executing environments closer to UE, TONA

is anticipated to have some advantages, such as better data privacy protection, lower latency, and lower energy consumption.

2 Network Paradigm Change

The TONA, as shown in Figure 1, introduces the orchestration and control functions as well as the resource layer in network AI. The control function uses control layer signaling to control multi-nodes (UEs, base stations, and CN NEs) and heterogeneous resources in real-time. We believe that the 6G network architecture requires the following changes in the design paradigm:

(1) Change 1: The object to be managed and controlled in network are changed from sessions to tasks.

(2) Change 2: The resources of the object are changed from one dimension to multi-dimensions, from homogeneous to heterogeneous.

(3) Change 3: The object control mechanism are changed from session-control to task-control.

(4) Change 4: The performance indicators of the object are changed from session-QoS to task-QoS.

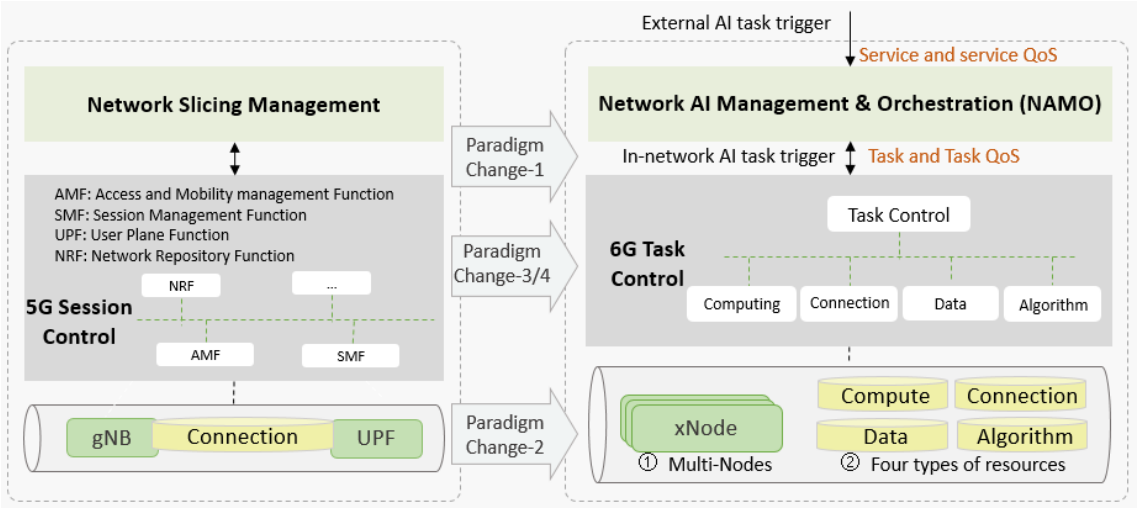


Fig.1 Network paradigm changes

2.1 Change 1: From Session to Task

AI tasks differ from traditional sessions in terms of technical objectives and methods. In terms of technical purposes, a traditional communications system provides session services, typically between terminals or between terminals and application servers, to transmit user data (including voice). Conversely, network AI (i.e., NE intelligence and network intelligence) aims to provide intelligent services for networks and improve

communication network efficiency. Service intelligence seeks to provide app-specific intelligent services for third parties. Thus, sessions and AI tasks have different purposes.

2.2 Change 2: From single-dimension to multi-dimension heterogeneous resources

The traditional wireless system establishes tunnels and allocates radio resources for data transmission. Conversely, TONA implements collaboration among heterogeneous resources of connection, computing, data and model/algorithm to execute AI tasks. Take an AI inference task as an example.

2.3 Change 3: From Session-control to Task- control

Unlike session control, task management and control in network AI includes the following functions: (1) Decomposing and mapping from external services to internal tasks, (2) Decomposing and mapping from service QoS to task QoS, and (3) Providing heterogeneous and multi-node collaboration mechanisms to orchestrate and control heterogeneous resources of multiple nodes at the infrastructure layer in real-time (to implement distributed serial or parallel processing of tasks and real-time QoS assurance).

3 Architecture and Key Technologies

This section describes the logical architecture and deployment options of TONA, and QoAIS details.

3.1 Logical Architecture of TONA

First, we introduce fundamental basic concepts in wireless network. A communications system consists of a management domain and a control domain. The Operations Administration and Maintenance (OAM) deployed in management domain is used to operate and manage NEs through non-real-time (usually within minutes) management plane signaling. The control domain is deployed on core network (CN) NEs, base stations, and terminals, and features with real-time controlling signaling (usually within milliseconds). For example, an E2E tunnel for a voice call can be established within dozens of milliseconds by control signaling.

Unlike the centralized, homogeneous, and stable AI environment provided by the cloud, the network AI faces the following technical challenges when embedded in the wireless networks: (1) AI needs to be distributed on numerous CN NEs, base stations, and UEs. Therefore, it is necessary to consider how to manage the massive number of nodes efficiently in the architecture design. (2) The computing, memory, data, and algorithm capabilities of different nodes vary significantly, requiring the architecture design to also consider how to efficiently manage these heterogeneous nodes efficiently. (3) The dynamic variation of the channel status and the computing load need to be factored into

the architecture design. To address the aforementioned challenges, TONA includes two logical functions, as shown in Figure 2: (1) AI orchestration and management, called Network AI Management & Orchestration (NAMO); and (2) task control. NAMO decomposes and maps AI services to tasks and orchestrates the AI service flows. It is not performed in real-time and is generally deployed in the management domain. Task control introduces the Task Anchor (TA), Task Scheduler (TS), and Task Executor (TE) functions in the control domain in three layers. This layered design strikes a balance between the task scope and real-time task scheduling, and effectively manages the numerous, heterogeneous nodes and aware of dynamic change of heterogeneous resources (e.g. channel status and computing load).

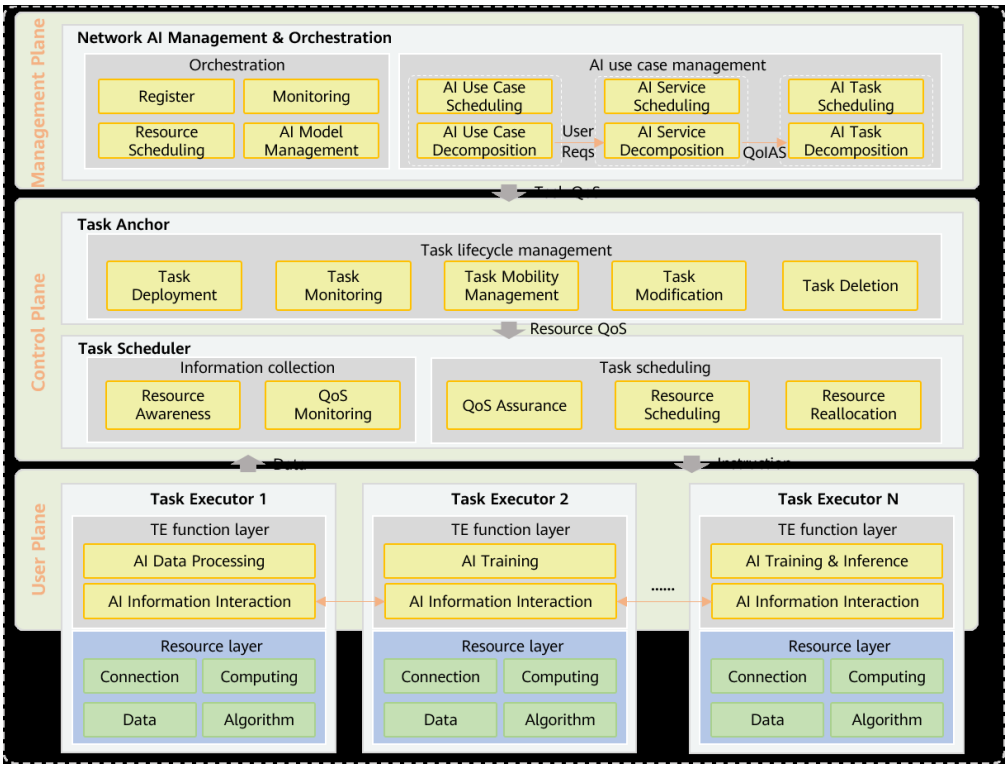


Fig.2 Logical architecture of TONA

3.2 Deployment Architectures

The statuses of TEs (e.g., the CPU load, memory, electricity, and UE channel status) change in real-time. As such, deploying TA and TS close to each other can reduce the management delay. According to the design logic of wireless networks, the CN and RAN need to be decoupled as much as possible. For example, the CN should be independent of RAN Radio Resource Management (RRM) and Radio Transmission Technology (RTT) algorithms. Therefore, this article recommends that TA/TS be deployed on RAN and CN, named RAN TA/TS and CN TA/TS, respectively. This way will allow TA to manage TEs in real-time flexibly. Four deployment scenarios of TONA are shown in Figure 3 to

describe the necessity and rationality of CN TA and RAN TA. These scenarios are only examples — there may be other deployment scenarios and architectures.

Scenario 1: gNodeB + UEs. In this scenario, the gNodeB serves as both TA and TS, and the UEs serve as TEs. Here, a UE is a computing provider and task executor, which accepts task assignment and scheduling from the gNodeB. The Uu interface and Radio Resource Control (RRC) layer between the gNodeB and the UE can be enhanced to support task controlling and scheduling purposes.

Scenario 2: CU + DUs. In this scenario, the CU serves as both TA and TS, and the DUs serve as TEs. Here, a DU is the computing provider and task executor. The F1 interface and F1-AP layer between the CU and the DU can be enhanced to support task controlling and scheduling purposes.

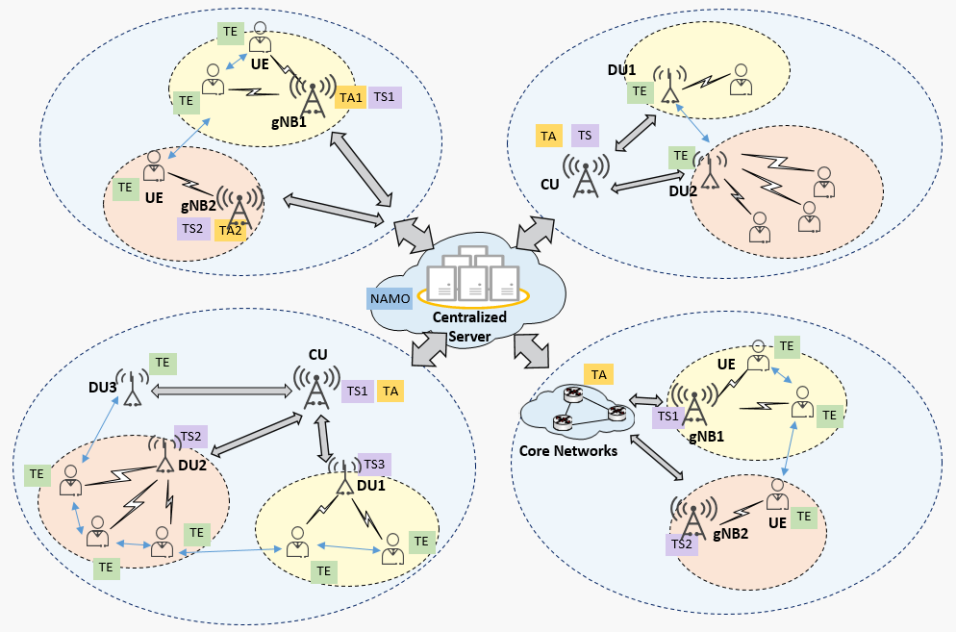


Fig.3 Four deployment scenarios of TONA

4 Advantage Analysis

Compared with cloud/MEC AI, the TONA and QoAIS have the following advantages (summarized in Table 2) in meeting users' customized AI service requirements:

(1) QoAIS assurance

Dynamic wireless environments require joint optimization of the heterogeneous resources (connection and three AI resources) to achieve precise QoAIS assurance.

(2) Latency

TONA computing is distributed on NEs closer to UEs or even directly on UEs to process data locally. This not only successfully achieves real-time and low-latency AI services, but also significantly reduces data transmission. In the cloud/MEC AI mode, a large

amount of data needs to be transmitted to the cloud/MEC for training, meaning that E2E data transmission takes longer to complete.

(3) Overhead

TONA can optimally allocate resources through the real-time collaboration mechanism of the heterogeneous resources, maximizing the overall resource utilization and reducing the transmission and computing overheads. Conversely, because the cloud/MEC AI cannot adapt to dynamic environments, it allocates resources based on only the maximum resource consumption to ensure QoAIS. As a result, the overall resource utilization is low, and the resource overhead is high.

(4) Security

TONA has native data security and privacy protection capabilities because it processes data inside the network. Unlike TONA, the cloud/MEC AI protects data privacy only at the application layer.

5 Conclusion

To meet the 6G vision of pervasive intelligence and internet of intelligence, TONA is proposed to support efficient collaboration of heterogeneous resources and multi-node in wireless networks, and to provide new services in the form of tasks at the network layer. By bringing new dimensions of resources to 6G networks (i.e., computing, data, and model/algorithm), this architecture enables the SLA assurance of computing related services such as AI services, further explores the application scenarios of 6G networks, and enriches the value of wireless networks. Furthermore, the task concept and TONA proposed in this article support not only AI tasks, but also sensing-, computing- and data processing-specific tasks.

REFERENCE

[1] IEEE Network, “Task-Oriented 6G Native-AI Network Architecture”, October 2023, https://ieeexplore.ieee.org/document/10273257

Envisioning Architectural Transformation towards 6G

Hideaki Takahashi, Nokia
Gerald Kunzmann, Nokia
Hannu Flinck, Nokia
Heiko Straulino, Nokia

Abstract - **This article covers system architecture migration options towards 6G, interworking aspects with the legacy generations, as well as a suggested design for the overall system architecture including RAN-CN functional split and logical RAN architecture. On the architecture migration, 6G system architecture is studied by leveraging the 5G system architecture principles, such as reusability, modularity, support of cloud native service-based architecture and resiliency, while adding principles to support 6G key values, sustainability, digital inclusion, and trustworthiness. The interworking design shall support seamless service continuity between legacy systems and 6G. as part of the overall system architecture, the necessity of RAN-CN separation and the functional split between those two domains is being analyzed. Finally, the logical 6G RAN architecture is analyzed covering both the emerging trend of cloud native deployments as well as the classical RAN deployments.**

1 Introduction

Approximately every ten years, the next generation of mobile networks is rolled out to enrich human life and society with evolutional innovations from the previous generation. Whilst the commercial 5G services have been successfully evolved, 6G research is in full swing together with strong momentum not only by individual organizations, but also from pre-competitive joint research initiatives in several geographical regions. Meanwhile, 6G standardization is gearing up towards the market roll-out around 2030. ITU-R WP5D has agreed on IMT-2030 framework covering usage scenarios and capabilities [1]. Following the ITU-R decision on IMT-2030, 3GPP has agreed on 6G standardization timeline on high level [2]. Technology innovations are supposed to be studied soon in 3GPP, encompassing overall system architecture. This article navigates architectural transformation and associated decompositions which are the tenets of architectural innovations towards 6G. Section 2 crystallizes design principles driving technology innovations across all the architectural transformations. Section 3 and onwards describes some of the key architectural innovations: migration, RAN-CN functional split and logical RAN architecture.

2 Architecture design principles

6G system architecture needs to enable a set of new use cases and unleash additional monetization capabilities, whilst still supporting existing use cases in an optimal manner [3]. To achieve this, the following principles are considered central pillars of the overall 6G architectural design.

A. Simplification

Architectural solutions should be lean in targeting performance requirements, whilst limiting complexity and avoiding over-optimization. Features and associated parameters should be designed for practical deployments and performances achieved by commercial devices. It is quintessential to make 6G consumable to its customers and limit costs for all phases from standardization, development, deployment to operation.

B. Energy Efficiency

6G per se should be sustainable to limit the carbon footprint of 6G systems as the number of UEs and traffic continues to grow exponentially. Besides that, the 6G system should maximize its handprint by enabling other industries to meet their sustainability goals [4, 5].

C. Native AI/ML in radio, system and management

Native AI/ML integration into the overall system is aimed at evolving the overall level of network intelligence and its perception of the environments. The AI-nativeness should cover management, control and user plane operation within the entire 6G system in a holistic way.

D. Multi-cloud nativeness

Embracing multi-cloud native technologies into 6G system can empower telecommunication providers to drive innovations, agility, and enhanced customer experiences. Multi-cloud native 6G system enables to integrate advanced technologies smoothly, e.g., AI/ML as well as enhancing resiliency and fault tolerance of the entire telecommunication system.

E. Programmability

6G should be able to offer a service innovation platform for MNOs and enterprises which embarks on multi-party value ecosystems. Exposing and leveraging the right open APIs at the proper abstraction levels is essential to leverage the capabilities of open hardware and software platforms.

F. Autonomy

Whilst the NW automation has been one of the principles since the legacy generations, e.g., SON, 6G should take it to the next level towards fully autonomous networks throughout their whole lifecycle from design, deployment, operation to decommissioning. It is even further indispensable in 6G era to manage complex NW and service offering in an economic way.

G. *Security and privacy*

6G as a system should be trustworthy and resilient. It needs also be protected against new attack surfaces emerging from distributed clouds, pervasive AI, usage/exposure of sensitive data and potential quantum computer attacks on cryptographic algorithms.

H. *Universal Access Convergence*

Seamless service experience should be supported across the growing number of radio technologies and access networks in terrestrial as well as non-terrestrial (satellites, HAPS, etc.).

3 System architecture migration towards 6G

3.1 Lessons learnt from 5G migration and deployments

When designing 6G system architecture and migration from 5G, it is worth to have in mind what was observed and learned from the commercial 5G migration and deployments. Some key learnings are outlined below.

Learning #1: A key learning from 4G to 5G migration was that many architecture options were standardized, including bearer type options in MR-DC, whereas only one Non-Standalone (NSA) option (i.e., EN-DC) and NR Standalone option have been implemented and deployed as of now. Furthermore, allowing a phased 5G deployment with NSA and SA has resulted in less migration toward NR SA, preventing to fully leverage 5G capabilities, such as URLLC, slicing, etc.

Learning #2: There is a solution to share a spectrum for 4G and 5G, a.k.a. Dynamic Spectrum Sharing (DSS) which was expected to contribute to migrate 4G to 5G. However, DSS has a considerable drawback to deteriorate spectrum efficiency. DSS enhancements to compensate the drawback was introduced in later 5G specifications, which are all optional for UE to support.

Learning #3: For 4G to 5G migration, NSA was chosen for the case where the existing spectra are used for 4G in conjunction with a new 5G spectrum, in spite of the drawback of compromising UL coverage due to semi-static UL Tx power split and the complexity of tight coordination between two gNBs. It was thought as challenging to re-farm all the existing 4G spectra to 5G since Day-1 5G deployment. The performance drawback of DSS was also the hurdle to migrate the existing spectra to 5G.

Learning #4: In addition to the RAN architecture with a single logical entity for base station (i.e., gNB), 5G RAN supported the disaggregated architecture for which RAN functionalities are split into 2 entities, Central Unit (CU) and Distributed Unite (DU). Whilst the CU-DU split architecture was thought as beneficial, enabling flexible RAN functional allocation and suitable to cloud based infrastructure, CU-DU split resulted in higher latency than E-UTRAN (LTE). Furthermore, not all of the standardized

disaggregation options are implemented and deployed, which infers that multiple standardized split options are counterproductive.

3.2 6G system architecture and migration enabler

Based on the experience learned from 5G as described above, the number of architecture options standardized for 6G should be limited to the one which is essential to migrate towards 6G smoothly for leveraging the full technology potential of 6G from its Day-1 deployment. The best approach is to standardize, develop and deploy only a single SA architecture for 6G as illustrated in Figure 3.2-1. 6G SA architecture should be designed to fulfill the design principles, e.g., energy efficient, implementation friendly, cloud native and trustworthy, as described in clause 2.

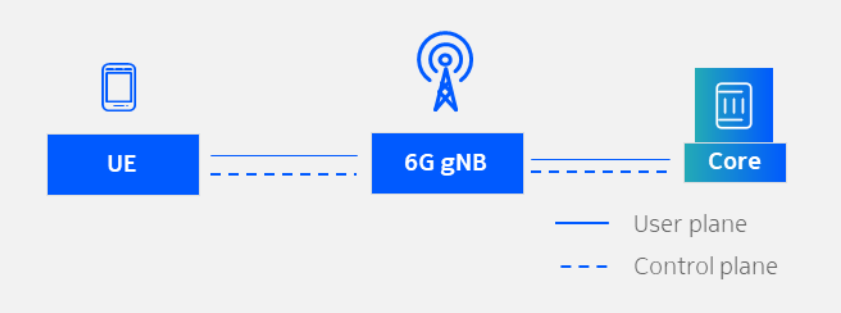


Figure 3.2-1: 6G standalone architecture

If 6G architecture supports only the standalone mode of operation, the similar concern on spectrum re-farming from 4G to 5G (Learning #2 in subsection 3.1) would be envisaged for 5G to 6G migration. Thus, necessity of the NSA mode of operation might be claimed in spite of the drawback observed in Learning #3 in subsection 3.1. Nonetheless, a revolutionary spectrum sharing innovation, Multi-Radio Spectrum Sharing (MRSS) can iron out the issue observed in 4G to 5G migration by achieving significantly higher efficiency than 4G-5G DSS. MRSS enables self-adjusting refarming and is desirable to be an integral part of the 6G radio design from the initial standards. MRSS also enables to aggregate a new 6G frequency band(s) and the existing frequency band, whereas the legacy 5G UE can be accommodated over the MRSS cell, as illustrated in Figure 3.2-2. Carrier Aggregation can be used to aggregate the existing frequency band(s) and the new frequency band(s) within 6G radio. UL CA as well as DL CA should be supported from Day-1 deployment to achieve high 6G capacity and performance over the new 6G spectrum as well as the existing spectrum. Thus, MRSS is a key enabler to smoothly migrate from 5G to 6G with a single SA architecture and diminish the necessity of NSA.

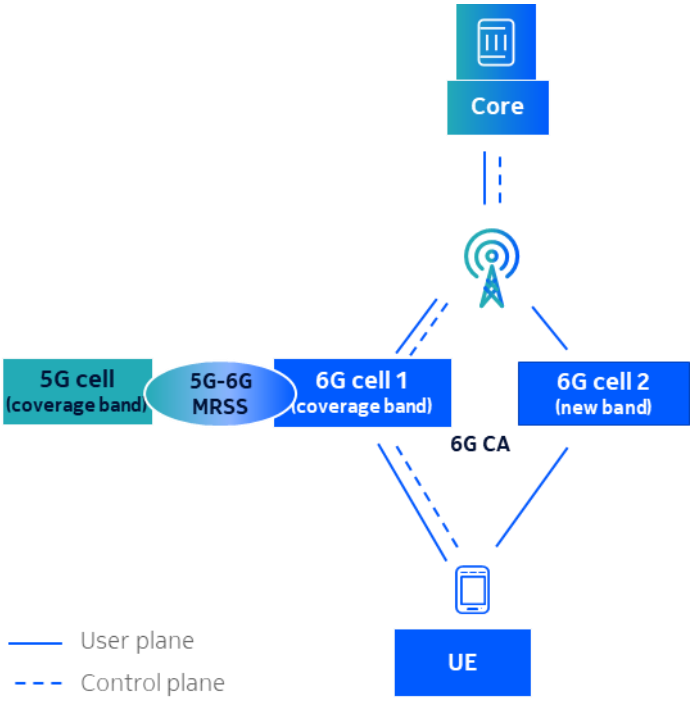


Figure 3.2-2: Migration from 5G to 6G SA with MRSS

3.3 Interworking with legacy systems

Apart from the architecture migration towards 6G, interworking between 6G and 5G is another quintessential enabler to support smooth migration. It is realistic to assume the scenario where 6G radio coverage is sparse compared to the existing 5G and 4G radio coverage. UEs need to move from 6G to 5G, and vice versa, for coverage reasons. Thus, the interworking between 6G and 5G needs to support UE mobility for all states to be supported (e.g., idle, inactive and connected as in 5G). For UE in the connected state, support of seamless mobility is deemed as necessary for essential services supported from legacy generations, e.g., IMS voice/video services. An open question is the necessity of interworking with 4G (i.e., E-UTRA connected to EPC). Albeit a desirable approach for functional simplification is to avoid the interworking with 4G, we have to consider the 4G to 5G migration situation predicted for the time of initial 6G deployments.

4 RAN - CN separation and interface

The mobile telecommunication systems standardized in 3GPP have been comprised of two network domains, i.e., Core Network (CN) and Radio Access Network (RAN) since 2G (GSM). Clear domain separation between RAN and CN is one of the basic tenets for the overall system architecture to facilitate independent technology evolution as well as different scalability and sharing needs of the two NW domains. Therefore, it is sensible to stick to the RAN-CN separation for 6G system, whereas a common solution across RAN and CN might be beneficial and should be considered for some areas, e.g., data

management and network analytics. With RAN-CN separation, an important aspect to be studied in a holistic way is how to design the interface between RAN and CN enabling independent evolution, different scalability and sharing needs.

Figure 4-1 shows the envisioned 6G system architecture built upon 5G system. 5G system leverages architecture principles to offer innovative services efficiently, e.g., reusability and modularity, supports cloud native service-based architecture and resiliency. Whilst 6G offers an opportunity for innovation, it is imperative to build 6G system upon 5G system evolution to sustain a balance between innovation and protection of existing investments. As illustrated in Figure 4-1, some Network Functions (NFs) in the proposed 6G system architecture are shared between 6G and 5G, whereas there would be some new NFs dedicated for 6G. The figure shows our current assumptions, and study is on-going which NFs to be shared with 5G and which new NFs are needed to support 6G RAN and new 6G services.

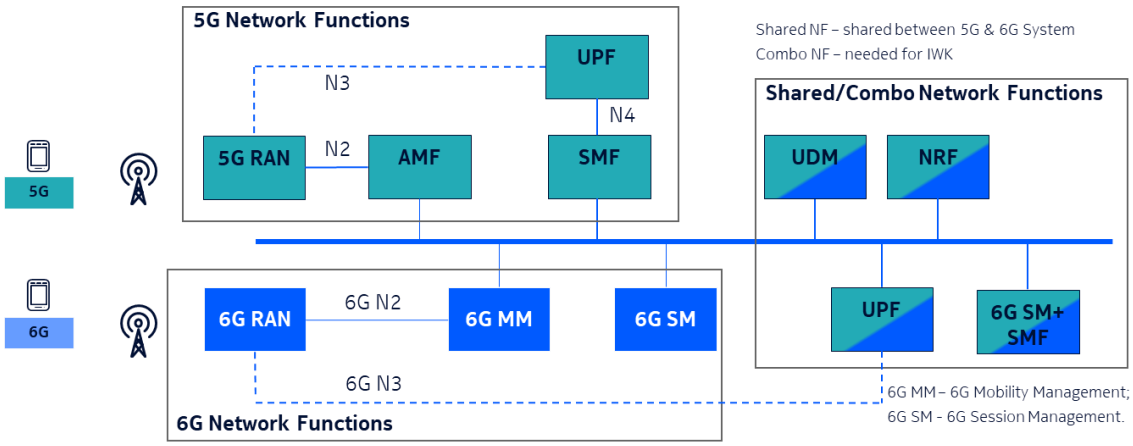


Figure 4-1: Envisioned 6G System Architecture

With respect to the RAN-CN interface, several options can be considered as listed below:

Option A. Point-to-point based on evolved NGAP

The N2 reference point between RAN-CN interface is kept for 6G system. The application protocol over N2, i.e., NGAP is evolved for new 6G services and features. This option sticks to the strict domain separation between RAN and CN as all communications between the two domains need to go through 6G N2 reference point between 6G MM and 6G RAN in Figure 4-1. ASN.1 encoding is continued to use NGAP information elements, which can be optimized for low latency.

Option B. Hybrid approach with SBI and evolved NGAP

The basic procedures like Mobility Management and Session Management continue using the N2 based reference point (i.e., NGAP over SCTP). In contrast, auxiliary services (like analytics, positioning, etc.) and/or new 6G services (e.g., sensing) are realized as SBI.

Option C. Full SBI

The 6G N2 is fully transitioned into a Service-Based Interface (SBI), i.e., using service operations (requests and responses) for the respective service APIs over HTTP2/3 as in the CN. 6G RAN and CN NFs expose services towards each other and communicate directly over SBI without relaying any NFs (like AMF in 5G Core).

All options are valid and expected to be studied further from the viewpoints of technical gain, complexity of implementation, etc.

5 Logical RAN architecture

In E-UTRAN (a.k.a., 4G LTE), a single logical entity is defined, i.e., eNB. It is up to NW implementation whether and how to disaggregate RAN functionalities. In 5G, standards support both a disaggregated and a non-disaggregated architecture. It is expected that both architectures will continue to be supported for 6G RAN, as illustrated in Figure 5-1. In this article, the classic RAN and the disaggregated RAN are defined as follows:

- Classic RAN: consisting of monolithic Node Bs (i.e., Base Station) without any disaggregation (like eNB/gNB in 4G/5G).
- Disaggregated RAN: built upon cloud-native environments where RAN protocol stack implementation is neatly divided into functional components and these functional components are packed into containers.

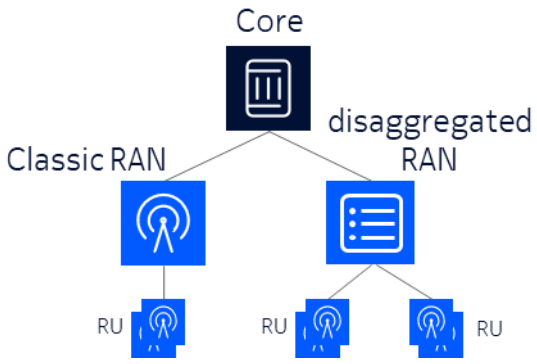


Figure 5-1: Logical 6G RAN architecture

Whilst classical RAN deployments are likely to be present even in 6G era, enhanced support of disaggregated RAN is worthwhile exploring to offer a wide range of 6G use cases and deployments. The disaggregated RAN architecture allows flexibility of RAN function placements aiming at a future-proof deployment. Nonetheless, the lesson learnt from 5G standardization and deployment, i.e., Learning #4 in subsection 3.1, should be reflected into 6G RAN disaggregation work and ironed out for 6G RAN by focusing on only essential open interfaces and optimizing to achieve lower latency comparable to classical RAN deployments.

6 Conclusion

This article presented design principles of architectural transformation towards 6G. The candidate system architecture, migration enablers, RAN-CN IF and logical RAN architecture were explored from the lessons learnt from 5G migration and deployments.

REFERENCE

[1] Recommendation ITU-R M.2160-0 (11/2023), "Framework and overall objectives of the future development of IMT for 2030 and beyond," https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2160-0-202311-I!!MSW-E.docx

[2] 3GPP SP-231693, "High-level Considerations for 6G Timeline," TSG Chairs. https://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGS_102_Edinburgh_2023-12/Docs/SP-231693.zip

[3] H. Takahashi, et. al., "Envisioning 6G Outlook and Technical Enablers," IEICE Trans. Commun., Vol.e106-B. No.9, September 2023. https://search.ieice.org/bin/pdf_link.php?category=B&lang=E&year=2023&fname=e106-b_9_724&abst=

[4] H. Holma and H. Viswanathan, "In the 6G era, we won't need to sacrifice sustainability for the sake of performance," Nokia Blog, November 2022. https://www.bell-labs.com/institute/blog/in-the-6g-era-we-wont-need-to-sacrifice-sustainability-for-the-sake-of-performance/

[5] S. Mukherjee, "Creating the future: maximizing our handprint and minimizing our footprint," Nokia Blog, June 2023. https://www.nokia.com/blog/creating-the-future-maximizing-our-handprint-and-minimizing-our-footprint/

Abbreviation List

Abbreviation	Explanation
6DoF	Six Degrees of Freedom
3GPP	Third Generation Partnership Project
ACO-OFDM	Asymmetrically Clipped Optical OFDM
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AIaaS	Artificial Intelligence as a Service
API	Application Programming Interface
APN	All Photonic Network
APs	Access Points
AR	Augmented Reality
ASIC	Application Specific Integrated Circuit
BaaS	Banking as a Service
BMI	Brain Machine Interface
BtoBtoC	Business to Business to Consumer
BtoC	Business to Customer
BW	Band Width
BWA	Broadband Wireless Access
CASE	Connected, Autonomous/Automated, Shared, Electric
CAPEX	Capital Expenditure
CMOS	Complementary Metal Oxide Semiconductor
CPRI	Common Public Radio Interface
CPS	Cyber Physical System
CPU	Central Processing Unit
CSI	Channel State Information
DaaS	Data as a Service
DCO-OFDM	Direct Current biased Optical OFDM
DDoS	Distributed Denial of Service
D-MIMO	Distributed MIMO

Abbreviation	Explanation
DMM	Distributed Mobility Management
DNA	Deoxyribonucleic Acid
DRAM	Dynamic Random Access Memory
DSRC	Dedicated Short Range Communication
DT	Digital Twins
DTC	Digital Twin Computing
DX	Digital Transformation
ECC	Elliptic Curve Cryptography
eCPRI	enhanced CPRI
EESS	Earth Exploration Satellite Service
eMBB	enhanced Mobile Broadband
EC	Electronic Commerce
ECC	Elliptic Curve Cryptography
ECU	Engine Control Unit
EHF	Extremely High Frequency
ETC	Electronic Toll Collection System
EV	Electric Vehicle
FA	Factory Automation
FDD	Frequency Division Duplex
FG-AN	Focus Group on Autonomous Networks
FH	Fronthaul
FPGA	Field Programmable Gate Array
FR	Frequency Range
FS	Fixed Service
FSPL	Free Space Path Loss
FW	Firewall
GDP	Gross Domestic Product
GEO	Geostationary Orbit
GPU	Graphics Processing Unit
GPS	Global Positioning System
GSO	Geostationary Earth Orbit

Abbreviation	Explanation
HAPS	High Altitude Platform Station
HARQ	Hybrid Automatic Repeat Request
HEMS	Home Energy Management System
HMD	Head Mounted Display
HMI	Human Machine Interface
HUD	Head-up Display
I/F	Interface
IC	Integrated Circuit
ICDT	Information, Communication, and Data Technology
ICT	Information and Communication Technology
IDS	Intrusion Detection System
IFFT	Inverse Fast Fourier Transform
IFoF	Intermediate Frequency over Fiber
IoT	Internet of Things
IP	Internet Protocol
IPS	Induced Pluripotent Stem
IPv6	Internet Protocol version 6
ISAC	Integrated Sensing and Communication
ISAC-OW	Integrated Sensing and Communication with Optical Wireless
ISRU	In-Situ Resource Utilization
ISS	International Space Station
IT	Information Technology
ITS	Intelligent Transport Systems
KEM	Key Encapsulation Mechanism
KPI	Key Performance Indicator
LCT	Laser Communication Terminals
LED	Light-Emitting Diode
LEO	Low Earth Orbit Satellite
LiDAR	Light Detection and Ranging
LLS	Lower Layer Split
LOS	Light of Sight

Abbreviation	Explanation
LTE	Long Tern Evolution
MaaS	Mobility as a service
MAC	Media Access Control
MCS	Modulation and Coding Scheme
MEC	Multi-access Edge Computing
MIMO	Multiple-Input and Multiple-Output
ML	Machine Learning
mMTC	massive Machine Type Communication
MR	Mixed Reality
MS	Mobile Service
MTP	Motion to Photon
MUP	Mobile User Plane
NBI	Northbound Interface
NGSO	Non-Geostationary Orbit
NIST	National Institute of Standards and Technology
NLOS	Non-Line of Sight
NOMA	Non-Orthogonal Multiple Access
NPU	Neural network Processing Unit
NR	New Radio
NTN	Non-Terrestrial Network
O&M	Operation and Maintenance
OAM	Orbital Angular Momentum
ODD	Operational Design Domain
OFDM	Orthogonal Frequency Division Multiplexing
ONAP	Open Network Automation Platform
OPEX	Operating Expense
O-RAN	Open Radio Access Network
OSINT	Open-Source INTelligence
OTA	Over The Air
OWC	Optical Wireless Communication
PA	Process Automation

Abbreviation	Explanation
PA	Power Amplifier
PCB	Printed Circuit Board
PCR	Polymerase Chain Reaction
PD-NOMA	Power Domain Non-Orthogonal Multiple Access
PDs	Photodetectors
PF	Platform
PHR	Personal Healthcare Record
PHV	Plug-in Hybrid Vehicle
PHY	Physical Layer
PII	Personally Identifiable Information
PMIPv6	Proxy Mobile IPv6
PPM	Privacy Policy Manager
PPP	Purchasing Power Parity
PUE	Power Usage Efficiency
QKD	Quantum Key Distribution
QoE	Quality of Experience
RaaS	Retail as a Service
RaaS	Robotics as a Service
RACH	Random Access Channel
RAN	Radio Access Network
RAS	Radio Astronomy Service
RAT	Radio Access Technology
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surface
RMSE	Root Mean Square Error
RoF	Radio over Fiber
RSA	Rivest-Shamir-Adleman
RTT	Round Trip Time
SDGs	Sustainable Development Goals
SDN	Software Defined Networking
SEAL	Service Enabler Architecture Layer for Verticals

Abbreviation	Explanation
SISO	Single-Input Single-Output
SNS	Social Networking Services
SPOF	Single Point of Failure
SR	Segment Routing
SRS	Sounding Reference Signal
SRv6	Segment Routing version 6
SRv6 MUP	Segment Routing IPv6 Mobile User Plane
TAT	Turn-around-Time
TCO	Total Cost of Ownership
TDD	Time Division Duplex
TEE	Trusted Execution Environment
ToF	Time-of-Flight
TRP	Transmission Reception Points
UAV	Unmanned Aerial Vehicle
UCN	User Centric Network
UE	User Equipment
UI	User Interface
UMTS	Universal Traffic Management System
URLLC	Ultra-Reliable and Low Latency Communications
UX	User Experience
V2X	Vehicle To Everything
VHTS	Very High Throughput Satellite
VR	Virtual Reality
vRAN	Virtual Radio Access Network
XaaS	X as a Service
X-OFDM	Mixed OFDM
XR	Extended Reality
ZSM	Zero-touch network and Service Management