The Next-generation Protocol Stack

----RAN Architecture, Protocol Stack and Function

Version 4.0

Executive Summary

The white paper firstly reviews the evolution of the communication protocol stack. Facing the future native AI characteristics of 6G network wisdom, analyze the potential application scenarios and the needs of the network development. On this basis, thinking about the protocol stack architecture, functions, and evolution of the smart endogenous network is proposed. The white paper hopes to arouse the continuous attention and thinking of the academic and industrial circles on the research of the 6G communication protocol stack architecture and function enhancement direction, so as to truly achieve the deep integration of DOICT and promote the comprehensive development of 6G.

摘 要

本白皮书首先回顾了通信协议栈发展演进历程。面向未来6G智慧内生的特征分析潜在的应用场景以及网络自身发展的需求。在此基础上提出关于智慧内生网络的协议栈架构、功能、演进的思考。本白皮书希望引起学术界和产业界对6G通信协议栈架构与功能增强方向研究的持续关注和思考，真正能够做到DOICT深度融合，促进6G全面发展。

Table of content

1. Introduction

The white paper proposes 6G protocol stack views and thinking for 2030+, based on the published versions such as "The Next-generation Protocol Stack over Air Interface 3.0" and "The Next-generation Protocol Stack over Air Interface 2.0". We hope to provide reference to study the 6G-oriented protocol stack architecture and functions for the industry.

With the opening of the 5G commercial prelude in 2020, universities, research institutions, industries and other parties are gradually turning to the research of a new generation of mobile communication systems, and 6G has become a new focus. The 6G Flagship of the University of Oulu in Finland took the lead in publishing the world's first 6G white paper, and held the world's first 6G summit in March 2019. China is also actively developing the layout of 6G key technology. The universities, research institutes and various industries such as operators, manufacturer, cloud computing and Internet companies are cooperating closely to conduct research on 6G-related topics. In addition, the companies such as Samsung, NTT DoCoMo, LG have also begun 6G research and exploration.

There are some white papers have studied the 6G vision and requirements, network architecture, and potential key technologies, which are of reference significance for the initial 6G research [1-5]. At present, opinions on various industries of 6G are letting a hundred schools of thought, but there are also some common opinions and trends in application scenarios, network characteristics, and key technologies.

In terms of vision and requirements, 6G will build a new network of intelligent and efficient interconnection of humans, machines and things. 6G will create a ubiquitous, refined, real-time digital world. The credible and organically integrated digital world accurately reflects and predicts the real state of the physical world in real time, and finally realize the vision of " digital twin, ubiquitous intelligence ". In the future, artificial intelligence (AI) technology will be born in the mobile communication system and become the cornerstone of the current digital twin world, which means AI technology is native to 6G networks. In general, there are two perspectives about the native AI 6G networks, the native-AI air interface and the native-AI network architecture.

In terms of key technologies, the native-AI air interface, that is, the deep integration of AI, machine learning (ML) and other technologies, will breaks the existing modular design framework of wireless air interface, realizes deep perception and efficient communication. Thereby the performance of native-AI networks will significantly improve, such as efficiency, reliability, real-time and security, and realize the self-operation and self-evolution network

In terms of network architecture, the Native-AI network architecture uses the integrated capabilities of communication, computing, and perception of network nodes to enable 6G networks to natively support various AI applications and build a new network ecosystem. The 6G network architecture should have the characteristics of native intelligence, native security, multi-domain integration, and integration of computing and network. Among them, the endogenous intelligence is the embedded AI capability of the 6G network, which realizes the endogenous intelligence of the architecture level. It can realize automated network operations, the combination of network and AI to provide communication and computing services, as well as intelligent perception, intelligent connection, discovery, services, management and orchestration of DOICT integration, laying the foundation for the intelligent connection of all things.

In summary, 6G and AI will gradually move towards full integration from the application scenarios, key technologies, network architecture, which requires all-round innovation in theory and design. Native AI will be a key development concept for the evolution from 5G to 6G, and the direction for promoting the sustainable development of 6G network

2. Situation and challenge

2.1 Review of the evolution of the protocol stack

The mobile communication system has been developed from the first-generation mobile communication system (1G), and the fifth-generation mobile communication system (5G) has also been commercialized in 2020.

The network structure in the 3G era basically uses a three-layer architecture. In the 4G LTE era, the network adopts an all-IP architecture, and revolutionary changes have taken place in network wireless transmission technology, air interface protocols, and system structure. The all-IP EPC (Evolved Packet Core, mobile core network evolution) supports unified access of various technologies, and voice, text, and video are all realized through the IP domain. Compared with the UTRAN system in the 3G era, the 4G E-UTRAN system integrates NodeB and RNC into one network element, i.e. eNodeB. The eNodeBs use IP transmission and are logically connected to each other through the X2 interface. Specifically, eNodeB refers to the elimination of the centralized control of the RNC and the addition of the physical layer, MAC layer, RRC layer, and scheduling, access control, bearer control, mobility management, and the RRM measurement, based on the original functions of the NodeB of the UMTS system. In other words, the eNodeB realizes all the functions of the radio access network. Such a network structure design greatly reduces the number of network elements, so that the network deployment and the network maintenance are easier, and the network is flatter.

The air interface protocol stack of the LTE system is divided into the user plane protocol stack and the control plane protocol stack, and the user plane protocol stack of the LTE system is decomposed into different protocol sublayers. The user plane protocol stack is similar to the UMTS system, including the physical (PHY) layer, the medium access control (MAC) layer, the radio link control (RLC) layer, and the packet data convergence (PDCP) layer. These sublayers are all terminated at the eNodeB in the network side. In downlink, the data is transmitted in the form of IP packets, and the IP packets are processed by multiple protocol layers before transmitting via the air interface, as shown in Figure 1.

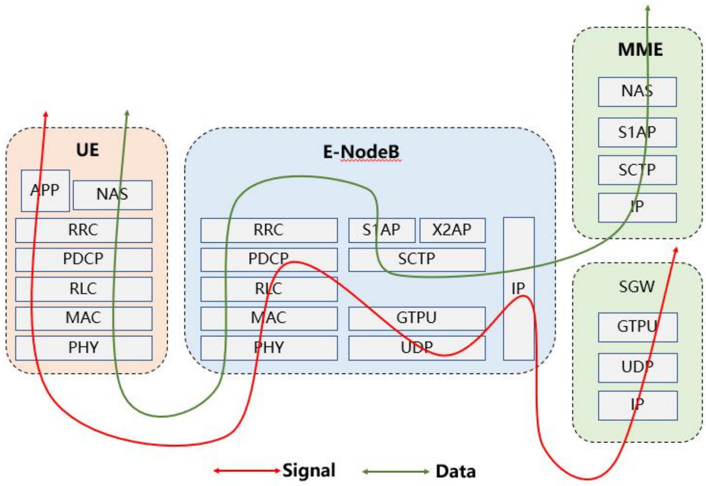

Figure 1 4G LTE overall protocol stack

The control plane protocol stack is shown in the following figure:：

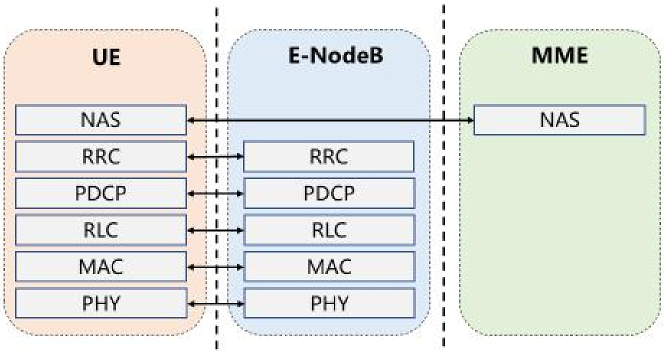

Figure 2 Control plane protocol stack of 4G LTE [6]

The control plane protocol stack mainly includes the non-access layer (NAS), RRC, PDCP, RLC, MAC, and PHY layers. Among them, the PDCP layer provides encryption and integrity protection functions, the RLC and MAC layers provide the same functions as in user plane. The RRC layer protocol terminates at the eNodeB in the network side, and mainly provides functions such as broadcasting, paging, RRC connection management, radio bearer (RB) control, mobility management, and UE measurement reporting and control. The NAS layer terminates at the MME in network side and mainly provides the functions including EPS bearer management, authentication, idle state mobility management, paging, security control, etc..

In the 5G era, the SBA architecture of the 5G core network is constructed. The SBA architecture includes modularization of network capabilities, the service-oriented interfaces, and the C-plane and U-plane separation. The mobile communication network has a closer relationship with business, and supports network function virtualization and slicing technology. The introduction of the CU/DU separation in wireless side can better realize flexible network deployment. The mobile network is no longer a solidification, but can be dynamically created and reconfigured so as to quickly support a variety of new business applications.

The 5G air interface protocol stack is mainly divided into three layers and two sides. The three layers are the network layer (L3), the data link layer (L2), and the physical layer (L1). The network layer is the user of air interface services, namely RRC signaling and user plane data; the data link layer (L2) distinguishes and labels different L3 data and provides different services; the physical layer (L1) provides the service of the the wireless resource and physical layer processing to the upper layer data transmission.

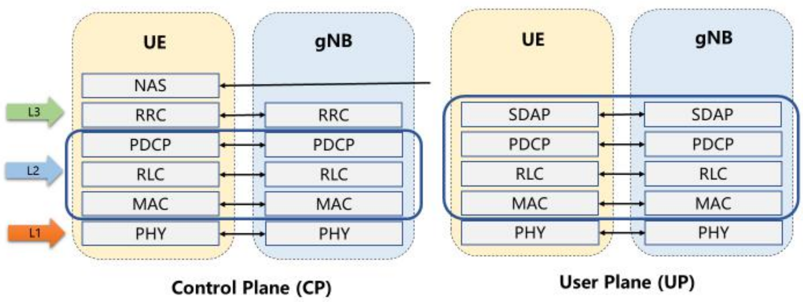

Figure 3 NR overall protocol stack [7]

From the control plane perspective, the structure is exact same in 4G and 5G. From the user perspective, 5G has the same structure as 4G except for the addition of a new SDAP protocol layer. In 5G QoS framework, in order to provide the refined granularity of the QoS requirement, the basic transmission granularity in the core network side is refined from the E-RAB in 4G to the QoS Flow; but in RAN side, 5G still uses the radio bearer concept in 4G. Therefore, in RAN side, the SDAP layer is introduced to realize the mapping between DRB and QoS Flow.

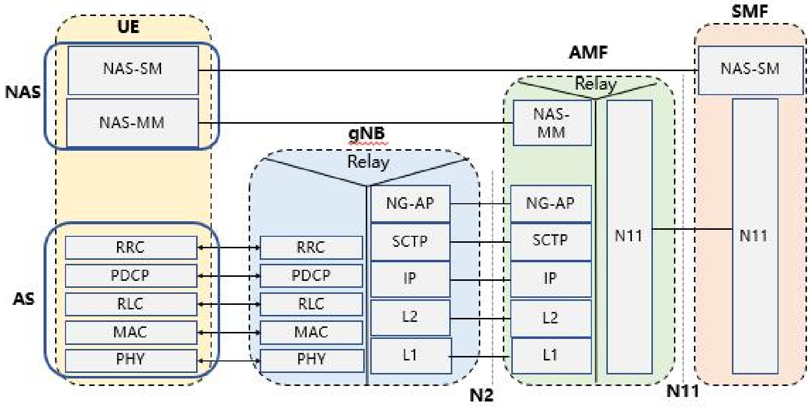

Figure 4 NR Control Plane Protocol Stack

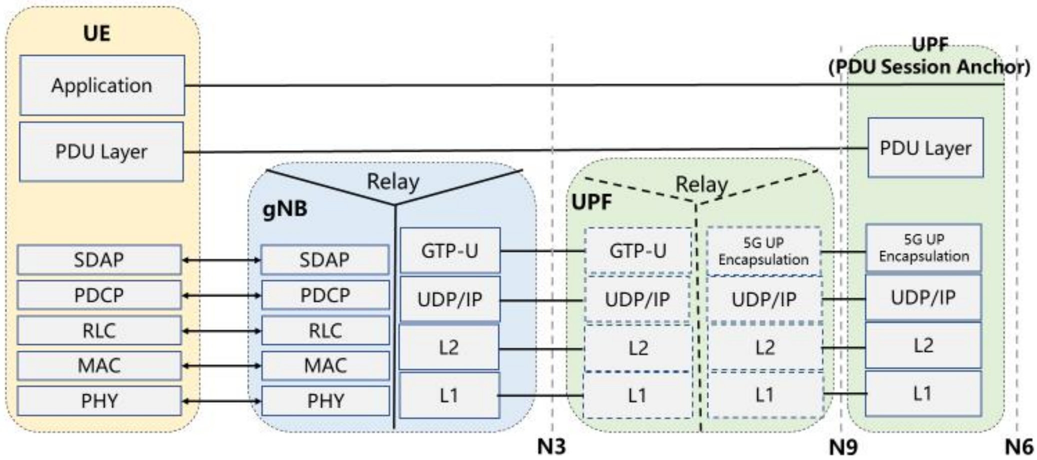

Figure 5 NR User Plane Protocol Stack [7][8]

2.2 Future scenario analysis

2.2.1 Space-Air-Ground-Sea Integrated Communication

The construction of terrestrial mobile communication networks is limited by constraints such as infrastructure cost and technical realization. Indeed, the network coverage accounts for only about 20% of the total land area at present. More than 95% of marine areas without mobile radio connectivity, which is far from reaching the construction of a global mobile communication network. Realizing the full coverage of Space-Air-Ground-Sea (SAGS) integration will be the goal of future 6G network communications, and its performance metrics e.g. network capacity, energy efficiency, delay and reliability will also be 10 to 100 times higher than 5G networks. The SAGS integrated communication system includes space-based (various types of satellites, etc.), air-based (unmanned aerial vehicles, airships, airplanes, and other high altitude platforms), ground-based (cellular, non-cellular network facilities), and sea-based (sea surface and deep-sea communication equipment). The system supports a variety of wireless access methods in different application scenarios, uses all available radio spectrum resources to achieve Internet access for many users worldwide, and provides fast and consistent communication services. It is envisaged that system will also integrate precise positioning, navigation, remote sensing, monitoring, and other functions to support regional economic and social development.

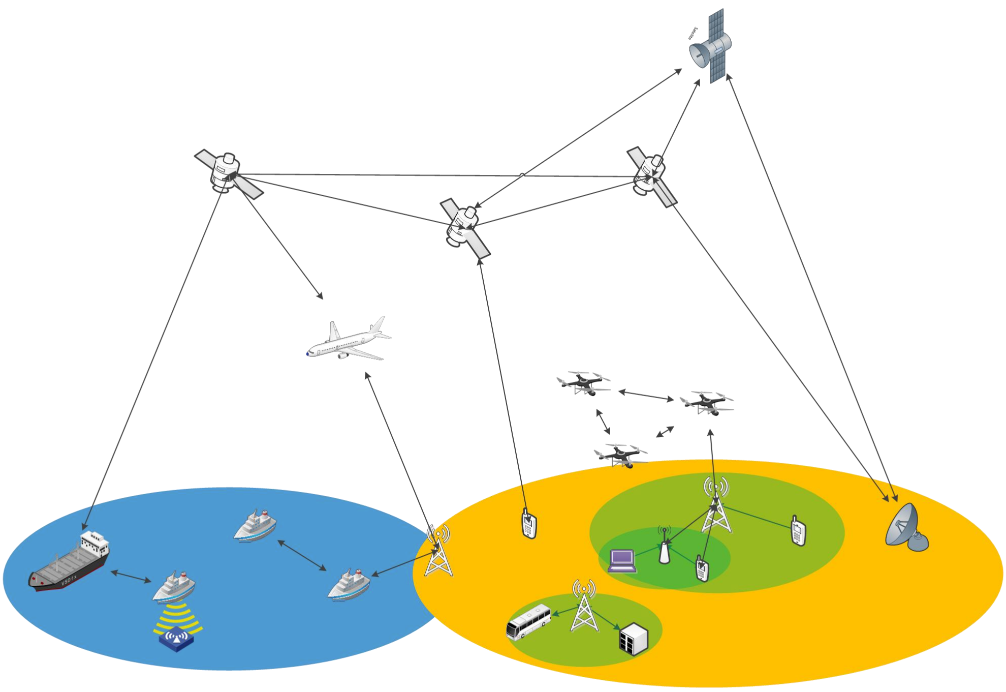

Figure 6. The architecture of the Space-Air-Ground-Sea integrated network

From the perspective of application scenarios, the inland areas are mainly covered by the terrestrial network, which gives full play to the advantages of high speed and capacity. UAV communication can improve the service capacity of densely populated urban areas with highly dynamic data traffic loads and help to reduce the network burden on the ground. Satellites or high-altitude platforms can provide effective communication solutions in remote areas, marine economy, infrastructure, aviation mobility, and post-disaster emergency support, which can give full play to the advantages of wide coverage, free from terrain constraints, flexible deployment. According to statistics, more than 70% of geographic space in the world involves more than 3 billion people without access to the Internet. The coordination of terrestrial and non-terrestrial network is an important means to solve the problem of the population digital divide. Moreover, there are 38 million flights worldwide each year with 4.4 billion passengers, and 80,000 freighters with 38 million voyages, all of which are target customers of SAGS connection services and have huge market space.

2.2.2 Holography

Holography is a technique that is used to display objects or scenes in three dimensions. Holographic telepresence is becoming a reality with the rapid development in the supporting technologies: high-resolution imaging and sensing, wearable displays, mobile robots and drones, specialized processors, and next-generation wireless networks. Holographic telepresence may be achieved through real-time capture, transmission, and rendering of a 3D holographic data on a local holographic display.

Current commercially available 3D display product is based on stereoscopic principle, which only utilizes human binocular depth perception to create 3D illusion. 3D holographic display has been considered as an ultimate glasses-free true-3D display technology because it can provide all depth cues and eliminate eye fatigue or visual discomfort. It has been considered as an alternative to current stereoscopic displays on the market.

As holographic display technology has made significant advances, holographic applications are becoming a reality. The real-time, 3D interactive applications will permeate future life: 1) tele conference: holographic telepresence will project remote participants as a hologram to local meeting; 2) Remote troubleshooting and repair applications: allow technicians to interact with holographic renderings of artefacts located in a remote location; 3) Training and education: provide users the ability to dynamically interact from remote with ultra-realistic holographic objects for teaching purposes; 4) tele-surgery: hologram feedback audio-visual and instruction to surgeon or robot; 5) immersive entertainment, gaming, sports, and much more

In a hologram, the same image is captured from different viewpoints, tilts, and angles. Depending on the position of the viewer relative to the image, a different "field" in an array of images is seen, with each image depicting the same "object" or "scene" from a slightly different viewpoint. Streaming holographic videos over network to the display systems has become an important research topic.

Holographic-type communication (HTC) will enable the ability to transmit and stream holographic data from remote locations across a network.

Besides the requirements of low latency and high reliability, HTC service has its specific KPIs

Ultra-high bandwidth: Required bandwidth may start from roughly 1 Gbps and increase up to 1 Tbps but depends heavily on encoding and trade-offs regarding bandwidth and compute for optimization schemes.

Strict synchronization: At 60 frames/second, latency variation across channels should not exceed 7 ms (duration for half a frame).

Support for concurrent flows. Depending on point cloud and image array dimensions, on the order of 1000 concurrent flows may need to be supported. And the network should be capable of prioritizing streams based on dynamic and varying criteria (related to viewing position and user focus)

Streaming a hologram over the network means to capture, render, and stream a target object. Capturing is performed by a camera array outputting images of the target object from multiple angles and views. The camera feeds need to be merged, rendered into a hologram, and encoded. The hologram is streamed across the network. On the receiver side, a client receives the stream, decodes it, and renders it for a holographic display or beamer.

Volumetric media is highly compressed, and the volume is independent of number of angles or tilts. Even with compression, holograms will require massive bandwidth. To cope with it, the current tendency is to come up with clever techniques aiming to reduce data that needs to be transmitted by means of eliminating portions of the content that will go unnoticed by a user. For example, some areas may be obstructed from the user's viewpoint, or some angles may not come into view based on the user's position. Schemes that take advantage of these aspects can dynamically adapt which parts of the contents to stream, and at what quality, at any given time. The effectiveness depends on the ability to predict the user's movement, thus rapidly adapting the data as needed. This is referred to as the "user interactivity challenge", and will enforce not only ultra-high bandwidth but also ultra-low latency (to ensure interactivity with the content). In addition, perfect synchronization of concurrent flows will be needed. Contrary to other types of multimedia services (e.g., UHD streaming), the interactivity challenge of immersive HTC will require ultra-low latency even if dealing with prerecorded content that does not involve real-time interaction with a remote party, as the user still interacts with the content simply by virtue of changing viewing angle and position.

In summary, handling HTC traffic, the network needs to manage a massive number of synchronized streams originating from either different sensors, an object at different angles, or a processed volumetric fusion. Traditional network is not competent to the specific requirements on ultra-high bandwidth (up to 1Tbps), ultra-low delay (1ms round-trip latency over long distance), stream bundle synchronization. HTC-enabled network will design for distributed control intelligence, network slicing, stream prediction and prioritization, etc. From protocol stack aspect, cross-layer optimization, flexible protocol layer deployment for multi-stream can be a basic objective.

2.2.3 Digital twin network

Digital twin is the real-time mirror image of physical entities in the digital world, with the characteristics of real-time interaction, real-time iterative operation and optimization. Based on the digital twin technology, the network can preprocess the data and predict the operation direction of the future network in the network virtual space in advance. That is to say, the network fault can be checked and solved before the network fault occurs, avoiding the occurrence of the fault, so as to realize the automatic network optimization and achieve the effect of "preventive cure and zero maintenance". In the digital twin network, some new functions, new services and even optimization means can be verified in advance, and the implementation scheme can be iteratively optimized. However, traditional network optimization and innovation often need to be tried directly on the real network, which may have an impact and interference on the existing network.

Digital twin network is a network system with physical network entity and virtual twin, which can be interactively mapped in real time. In this system, various network management and applications can use the virtual twin constructed by digital twin technology to efficiently analyze, diagnose, simulate and control the physical network based on data and models.

The 6G digital twin system provides a basic operating environment for 6G native intelligence, which provides basic support for AI-related processing and calculations, and simplifies the operating load and complexity of physical network. In other words, the 6G digital twin system and the native intelligence system together form a series of online operations for the operation, maintenance, application-oriented control calculation of the physical network, and become the brain of the physical network, directing each part of the physical network to complete the protocol or the service capability required by the operator.

As shown in Figure 7, the protocol stack scheme of native intelligence and digital twin can be implemented by introducing the intelligent function body and digital twin function body into different protocol stack functions of operation and maintenance system, core network, transmission network and access network.

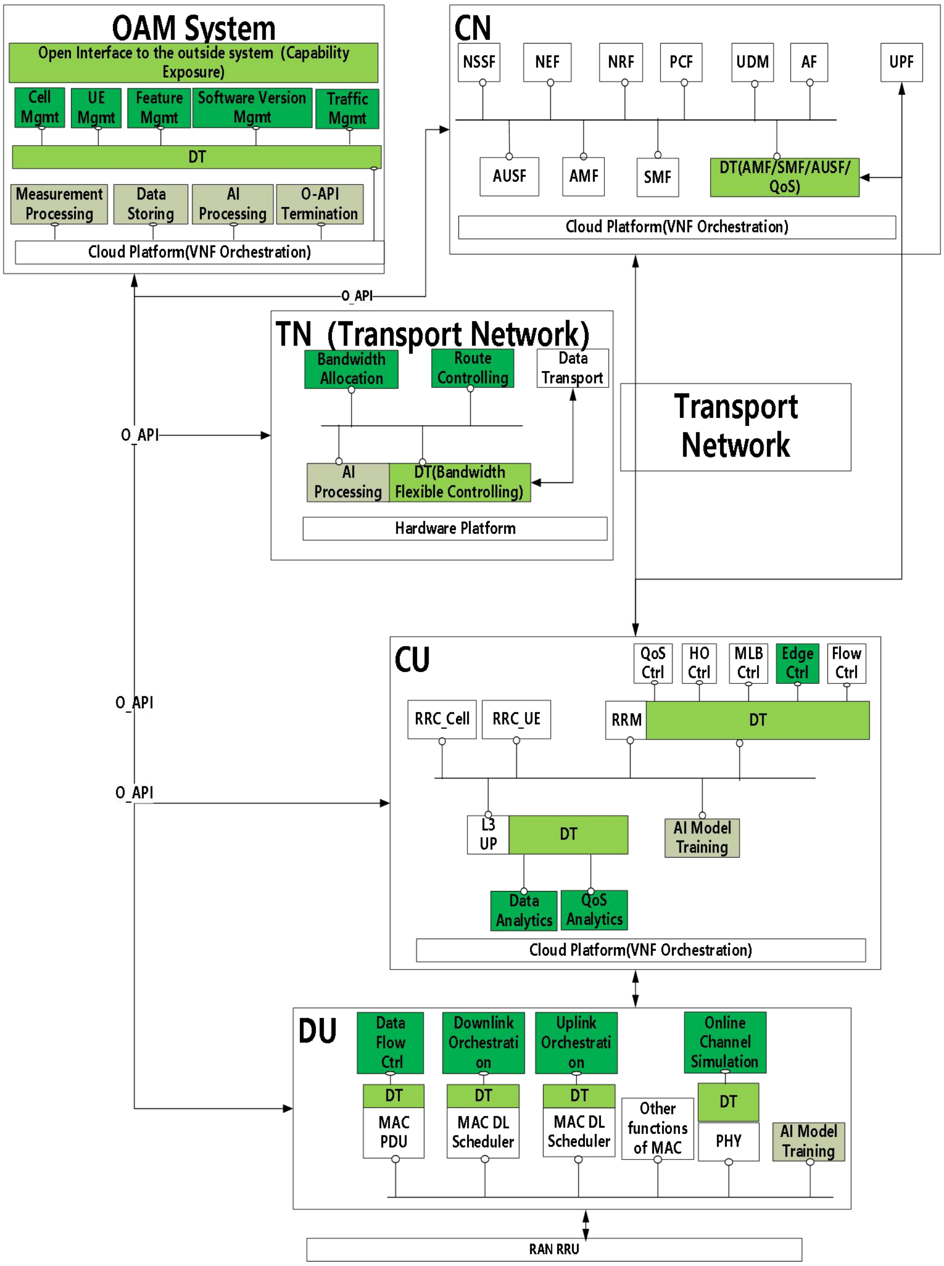

Figure 7 A protocol stack function of native intelligence and digital twin

The OAM system includes the following sub-functions: 1. Measurement processing function, which processes the measurement information reported by the network side and the terminal side, including the measurement information reported by the distributed intelligent function and the distributed digital twin based on the network side and the terminal side, as well as the large-time scale information directly reported by the network side and terminal side required by OAM. 2. Data storage function, which cleans, accumulates, calculates, smoothes and combines the received measurement information for storage; 3. AI processing function, which trains and runs AI model based on measurement information and stored data; 4. O-API termination function, responsible for the interface processing between OAM and each network element. The DT function system of the OAM system is based on the above sub-functions, and generates the overall digital twin mirror image of each network element and terminal. This digital twin mirror image is a large-scale mirror image, which mainly reflects the digital mirror images of different network element levels and the entire terminal, and depicts the overall characteristics of the network element or terminal for management through OAM system. Based on the DT function, OAM performs cell management, UE management, feature management, software version management, traffic management, etc. The above management functions are based on the digital twin of the network element or terminal provided by the DT system to carry out large-scale or large-granularity management such as coordination, control, and policy transmission at the cell level, terminal level, or service level.

Digital twin function mirroring is performed for L3UP, RRM, MAC and PHY in the access network. L3 DT in RAN is responsible for online simulation of the processing of data received and sent by the access network, and generates air interface-oriented transmission strategies (data analytics and QoS analytics) through the packet header information, payload length, cache status, etc. of each packet. DT in RRM is responsible for the online simulation of radio resource management, and generates control related to handover, flow control, QoS, etc. The edge control unction realizes the function configuration of DUs driven by AI, selection and management of different DUs, etc. DT in MAC layer is responsible for generating corresponding online simulation for different MAC functions, including uplink orchestration corresponding to MAC UL scheduler, downlink orchestration corresponding to MAC DL scheduler, and air interface flow control (data flow CTRL) corresponding to data packet transmission and reception (MAC PDU). DT in PHY layer is responsible for online simulation of physical channel processing, including channel estimation, channel model adaptation and selection, etc.

OAM system communicates with the interfaces (open API) of various systems for information exchange. Under the unified coordination of OAM, the upstream and downstream data processing is completed between various functional entities. For example, when a UE applies for a new service, the OAM system generates interfaces for the overall requirements of the core network, transmission network and access network, as well as the division of work of each network element according to the service characteristics applied by the UE and the characteristics of this type of service accumulated during the system operation, letting the transmission network provide proper bandwidth and corresponding transmission QoS guarantee and the core network and wireless network consider the QoS guarantee characteristics when establishing end-to-end bearer. Based on this requirement, the AI and DT systems of CN, TN and RAN respectively produce the required resource support in combination with their actual operation. If they cannot be satisfied, the information need to be reported to OAM in time.

The intelligent control process of network internal integrated service control is as follows:

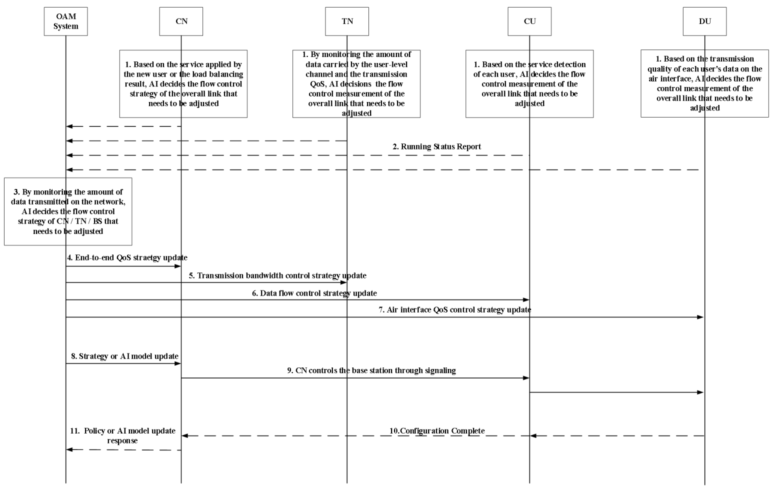

Figure 8 service control interaction flow chart of native intelligence and digital twin

Figure 8 shows the interaction process of native intelligence and digital twin:

1. CN, TN, CU and DU respectively perform service control at their respective levels.

CN generates a demand for service control when a new user performs service application or load balancing processing.

TN monitors the quality of data packets received or transmitted on each transmission node (such as router) in TN network in the process of receiving and transmitting the carried service data packets (such as packet loss rate, data cache occupancy rate, packet detention delay, etc.), resulting in the demand for service control.

The CU monitors the quality of service data packets received and transmitted by each user on the wireless link, including the data buffer occupancy rate of L3, the ratio of data packets to be retransmitted, etc., resulting in the demand for service control.

DU monitors the reception and transmission quality of each data packet on the air interface, including HARQ retransmission, block error rate (BLER), the ratio of large data packets sent by segmentation, the ratio of small data packets sent by concatenation into large data packets, the times of data packets moved between internal buffers, the applicability of dynamic QoS indicators in air interface reception and transmission, etc.

If it needs to be initiated through the network operation and maintenance pipeline, an application shall be initiated to the OAM system.

CN, TN, CU and DU send running status report to OAM system, including the monitoring information received and transmitted by the whole end-to-end service (application layer to application layer) on CN/TN/CU/DU, the running status of each user in the network, the transmission quality assurance of transmission network data receiving and transmitting, the running status information of CN/TN/CU/DU itself or AI/DT software body, etc. The information about the user or service operation status can be the information data directly reported by the relevant function body, or the reported information data after being processed by the relevant AI function body or DT function body on CN/TN/CU/DU according to the requirements defined with the OAM system interface.

3. The OAM system monitors the data volume of the whole network service. If it is necessary to initiate end-to-end transmission channel control, it will initiate the control policy update process.

4/5/6/7. OAM sends service control policy update to CN/TN/CU/DU respectively.

For CN, send end-to-end QoS policy update (4. End-to-end QoS control policy update), and CN can adjust the corresponding QoS parameters for the specified service according to the policy, and generate a new QoS indication.

For TN, the transmission bandwidth control strategy is updated, including increasing or decreasing the (logical) bandwidth of the transmission channel, adjusting the data transmission service quality assurance parameters of each transmission node, etc.

For CU, the flow control strategy for data reception and transmission includes the flow state of received CN data, the flow state of received and transmitted data to DU, etc.

For DU, the air interface QoS control strategy is updated, including the MAC layer formulating targeted HARQ mechanism for the bearer of corresponding types of services, air interface transmission and feedback mechanism, and the arrangement of physical resources (physical channel, and corresponding coding/decoding, modulation / demodulation, power, control companion channel, etc.).

During the operation of the system, OAM system triggers the control flow between CN and base station according to the results of AI and DT inspection.

8. The OAM system sends the relevant updates of the AI model, strategy or DT system to the CN;

9. Triggered by the OAM system, CN generates NAS (non Access Stratum) signaling and AS (Access Stratum) configuration signaling to configure the base

station (CU/DU) respectively.

10/11. The base station sends a confirmation message to the CN, and the CN generates a corresponding completion confirmation message to the OAM system.

2.2.4 New Ultra-Dense Network

In 5G and before generations, Ultra-Dense Network (UDN) aimed to crowded scenarios, such as shopping mall, railway station, airport, stadium with big show, etc. Connection density is defined as the total number of devices fulfilling a specific quality of service (QoS) per unit area (per km^2) with 99% grade of service (GoS). In 5G, the requirement of connection density is 1 000 000 device/km^2 in urban environment.

The data traffic and the numbers of connected things will increase substantially for 6G. Device density may grow to hundred(s) of devices per cubic meter. This poses stringent requirements on area or spatial spectral efficiency and the required frequency bands for connectivity. The technical KPI enhancements may potentially bring deeper immersion and interaction for users (e.g. holographic interaction, tactile Internet etc.), even in hitherto underserved areas. Such capabilities may add value to remote service delivery (e.g. health, education etc.) in ways that far exceed legacy mobile technologies. Besides traditional use cases in UDN, use cases in new UDN can include Ultra-dense devices in moving vehicles, Multi-sensory extended reality, and Industrial automation and robotics, etc.

Ultra-dense devices in moving vehicles: the moving vehicles can be trains, buses and even civil aircrafts. Here the infrastructure, vehicles, passengers, and goods are seamlessly connected with high data rates. With the explosion of devices (including sensors and user application terminals) and diverse applications, the devices density and capacity requirements in the moving vehicles form a new UDN scenario.

Multi-sensory extended reality: Augmented, mixed, and virtual reality

(AR/MR/VR) applications, capturing multi-sensory inputs and providing real-time user interaction are considered under this use case. Extremely high per-user data rates in the Gbps range and exceptionally low latencies are required to deliver a fully immersive experience. Remote connectivity and interaction powered by holographic communications, along with all human sensory input information, will further push the data rate and latency targets. Multipleview cameras used for holographic communications will require data rates in the order of terabits per second. With the multiple sensors, the device density can be large even in the scenario with low population desity.

Industrial automation and robotics: Industry 4.0 envisions a digital transformation of manufacturing industries and processes through cyber-physical systems, internet-of-things (IoT) networks, cloud computing, and artificial intelligence. In order to achieve highprecision manufacturing, automatic control systems, and communication technologies are utilized in the industrial processes. The emerging need for the ultra-dense deployment of industrial IoT devices will require 3D connectivity supporting up to 10 connections per m^3.

In summary, in 6G, the number of devices connected to the wireless network will be much larger than the number of population, and the deployment of devices can be different to the population distribution. Some distinct characteristics of new UDN can be: 1) the new UDN scenarios can be extended to any areas, and traditional connection density on urban/rural may not be applicable; 2) the definition of connection density will be extended from area to spatial; 3) the coverage of new UDN can be dynamically.

2.2.5 Integrated sensing and communication

In wireless communication, electromagnetic waves propagate from the transmitter to the receiver through space to transfer information, while the received signal also carries environmental information. Therefore, the communication process has the sensing capability. There are a huge number of base stations in the world now, e.g. the number of base stations in China is on the order of one million. If the base stations and terminals can be used for sensing, the sensing range will be greatly expanded, as well as the sensing applications. The integration of sensing and communication is one of the most effective means to build a digital twin world.

The sensing and communication integrated system can provide location, direction, distance, and speed of the sensing target, or target detecting, tracking, ranging, imaging, etc. These sensing services can be used to improve the performance of wireless communications and the quality of life and production. According to the function and content of sensing, it can be divided into macro sensing and refined sensing. Macro sensing generally includes weather, air quality, traffic flow, people flow, environmental reconstruction, and target tracking. It can be used in scenarios such as meteorology, life services, intelligent transportation, smart cities, and network planning and network optimization. Refined sensing generally includes motion recognition, facial expression recognition, breathing monitoring, heartbeat monitoring and material detection, etc. It can be used in scenarios such as intelligent interaction, medical health, and security checks.

2.2.6 Native Determinacy

The deterministic network was originally proposed to solve the problem that the "best-effort" mechanism of the packet-switched network cannot guarantee the delay and jitter requirements of multiple services. Deterministic network technology includes IEEE TSN, IETF DetNet, and DIP, but it is limited to wired networks. With the continuous improvement of mobile network performance, vertical industry businesses have gradually begun to use wireless networks to achieve flexibility, large bandwidth, and low latency. 5G network achieve determinacy by integrating with TSN to guarantee the performance of industrial time-sensitive services. With the diversification and differentiation of services in 6G , there will be a higher demand for determinacy. The realization of native determinacy in the mobile network will become a very important feature.

In the future, the diversified scenarios and differentiated network requirements of 6G will have more stringent requirements for determinacy, not only including the existing deterministic delay and jitter, but also deterministic bandwidth and deterministic positioning, reliability, availability and certainty of clock synchronization. Typical services include cloud AR/VR, holographic communication, smart industry, etc. The existing deterministic mechanism for mobile networks only uses the entire network as a transparent bridge, which cannot truly sense service requirements and reuse the technical system of URLLC. In 6G mobile network, including the RAN side, in order to meet deterministic service needs, it needs to have native deterministic capabilities which can intelligently sense business needs, perform precise time synchronization and real-time and dynamic allocation of RAN resources. Besides, 6G will not only achieve native determinacy but also adapt to external vertical industry protocols to guarantee the needs of deterministic scenarios in every aspects.

2.2.7 Intelligent network

As the rapid development of future network on 2B/2C service, network design and maintenance will become very complicated, it's more and more difficult to use the traditional communication network optimization method to fulfill the requirements of future network. To break through the bottleneck for traditional communication network, the fusion between AI/ML technology and traditional communication network seems to be one of the trends of the future network development.

Compared with the traditional communication network, AI/ML enabled intelligent network can use diverse information including inputs from environment, the available resources from the network, the services provided by the network, the service object and so on. These inputs are valuable to improve the performance of communication network and the target is to create a self-organized and self-optimized communication network in the future.

Intelligent network not only emphasizes the intelligence empowered network optimization, but also pays attention on the role of terminals in the intelligent network. In traditional communication system, terminals usually passively accept network management. When it comes to intelligence empowered network, we believe terminals can actively participate in the management of the entire intelligent network through real-time/near-real-time interaction with the network, intelligent network can better provide personalized services for terminals.

2.2.8 User Native Networking

In the future 6G era, the user experiences in communication, Sensing, computing, and intelligence will be greatly boosted, and the degree of personalized services will be fully satisfied. The new 6G networking should not only support user-centric capabilities in terms of radio signals, resources, functions, and services, but also support the new architecture and networking mechanism to meet the requirements of users for self-development, self-initiative, customization, privacy, and personalization.

In the traditional IMT mobile unicast system, each user and terminal are considered as independent serving object entities. Each terminal establishes its independent communication connections with the network through random access, core network authentication, session establishment, and radio access network resource allocation processes etc. The logical functions of UE are obviously lower and less powerful than those of the network side, which seems extremely asymmetric and unfair, and UEs are relatively in passive and "served" roles. With the continuous status promotion of 6G users and the continuous improvement of UE capabilities, the logical function and status of UE in the new 6G system will be gradually improved from passive to active role, from simply being served to providing services.

In the user native networking scenario, several UEs can establish a subnetwork of UEs locally based on specific subscription requirements, and then any member UE in the subnetwork can access the 6G network through one or more primary UEs as the access interfaces. The UEs in the sub network can directly or indirectly communicate with the 6G network. Such application scenario includes: On a user, several wearable communication devices can be connected to the 6G network together at the same time, multiple user groups associated with the same services can be connected to the network at the same time, and several distributed communication devices collaborating in the same tasks can be connected to the network at the same time.

2.2.9 Multi-parties Native Networking

In the future 6G era, the new 6G network should be able to provide more powerful, comprehensive and intelligent services in the traditional ToC and ToB fields. In addition, it should be able to differentiate serving subject groups in more granularities and make full use of certain features of these serving subject groups, in order to provide services in more economical and customized manner. Each serving subject group can be identified and effectively served by new 6G networks due to its natural association e.g. in physical distribution, user profile, service requirements or social relations, thus forming more targeted classification and hierarchical optimization.

In the traditional IMT mobile systems, the network topology is relatively simple, focusing on the centralized cell topology with more emphasis on wireless coverage continuity. Network planning and optimization of mobile operators used to care less about different levels of serving subject groups. They could only customize and adapt networking by network slicing, parameter set optimization, and other lower-level means, which results in poor performances in terms of coverage and costs. With the introduction of more 6G networking modes, such as de-cellularizaiton, distributed subnet, multi-hop relay, and drone/satellite aiming for large coverage, 6G is expected to further optimize the overall wireless coverage solution.

In the multi-parties native networking scenarios, new 6G network can be divided into different levels of serving subject groups in accordance with their association characteristics of different granularities and serving objects. Taking advantages of those features, the new 6G networking can optimize the networks to provide more customized classified services. In addition to above ToC and ToB, such application scenario also includes: ToH which is oriented to family life, ToG which is oriented to government office, ToI which is oriented to industrial production and ToS which is oriented to social governance.

2.2.10 The energy and cost-prioritized networking mechanisms

In the future 6G era, new 6G network must not only evolve towards higher KPIs and ultimate system performances, but should also support energy and cost-prioritized architecture and networking mechanisms. The vision of a new 6G network architecture has two basic principles: Compatibility and simplicity. Therefore, the new 6G network should be compatible with some simplified, low-energy-consumption, and low-cost coverage wireless solutions to meet the basic service requirements in specific scenarios. This is very important to support China's "dual-carbon" and "green sustainable development" strategies in the future.

During the evolution of traditional IMT mobile systems, each generation of systems have been increasingly seeking better KPIs. The system becomes more and more complicated, and the system kernel is becoming much heavier. The protocol interface specifications turns to be more complicated, while system energy consumption and cost turns to be higher. Despite the reduction in energy consumption and costs by using AI and various narrowband technologies, the overall situation remains grim and unsustainable. If such trend would continue, the successful commercial usage of new 6G networks will face great non-technical bottlenecks and obstacles.

The energy and cost-prioritized networking means that, on the premise of satisfying the basic service requirements in specific scenarios, the new 6G network should be compatible with those simpler, lower energy consumption and lower cost wireless solutions. In addition, when the network traffic amount is zero, the system energy consumption and operation cost will tend to be zero as well, i.e. Scale-to-zero. Such application scenarios include special vertical industry coverage scenarios, massive concurrent transmission scenarios of small data packets in broadband systems, and narrowband services in narrowband systems.

2.2.11 The multi-band aggregated networking

In the future 6G era, new 6G networks will face more brand-new scenarios and services, and the technologies of integrating and utilizing higher frequency band, multi-band, and full-band resources (hereinafter referred to as "multi-band technologies") will be one of the important means to overcome the above challenges. The "multi-band" technology can achieve more efficient aggregation of a wider range or discrete wireless spectrum resources. In addition, the physical advantages of each sub-band can be utilized to achieve complementary performances. This can support new 6G services , for example, holographic communication, digital twin, to improve network resource utilization, to reduce cost and increase efficiency.

In the traditional IMT mobile systems, due to the limited cellular topology mode, the joint usage of multi-frequency resources is static and segmented. Although existing mainstream technologies such as CA and DC have been applied, their efficiency and flexibility have been proved to be not so satisfactory in practice. For example, CA regards each component carrier as an independent cell, and implements resource maintenance, system information message processing, and synchronization independently, which causes redundancy and waste of scheduler resources. The handover delay between PCell and SCells is long, and the benefits of carrier aggregation in idle mode need to be improved.

In a multi-band aggregated network, air interface carrier resources will be deeply virtualized and cloudified, and will no longer be used in traditional static and fixed manner. Radio carrier resources, corresponding to RF carriers, can be decoupled from and mapped to baseband carriers, and each physical channel can flexibly aggregate and utilize lower-layer multi-band resources. Such usage scenario includes: Operators with a large number of discrete spectrum resources, carrier resources shared by multiple operators or multiple RATs, and so on.

2.2.12 Technical challenges for 6G communication

From the law of network reform, there are three main driving forces. First, with the continuous emergence of new services, new applications and new requirements, the requirements for network capability are also more stringent. Typical application scenarios in the 6G era, such as digital twin, holographic communication, super-powered transportation, integration of communication and sensing, native intelligence, and integration of ground, air, space and sea, require the network to have better service capability. The requirements of performance indicators in the 6G era may include extreme user experience rate, ultra-low latency, ultra-high data rate, ultra-high security, integrated coverage of ground, air, space and sea, etc. Second, the future 6G network needs to face up to the problems and challenges faced by the current network, such as high energy consumption, high cost and low operation and maintenance efficiency. These problems should be fundamentally changed when designing the future network and protocol stack. Third, driven by new technologies, such as cloud computing, big data and AI technology, 6G network architecture should include and make rational use of new technologies, and new technologies should promote the development of 6G network architecture in a more efficient and lower-cost direction.

The difficulties and challenges that 6G network and protocol stack may face in the evolution process may come from the following aspects.

The first aspect is that the solidified protocol structure has led to a ceiling of network performance. For 5G wireless network, the protocol structure is hierarchical, and all services need to be processed by layers. However, the processing of each layer will introduce a specific delay, resulting in the delay bottleneck of service transmission. Moreover, a large number of signaling interactions are required to implement a specific service network, which also leads to a certain consumption of network resource. Therefore, the design of 6G network needs to consider the lightweight network architecture, so as to simplify the protocol stack structure and enhance the function as much as possible.

The second aspect is how to transform AI ability from plug-in design to native design. For example, there are already many schemes that use AI and big data to implement intelligent air interface, resource management, slicing and intelligent network operation and maintenance. However, since the initial network design did not fully consider the support for these functions, it is difficult to make major breakthroughs and innovations from the entire network structure. Patch-based AI function enhancement is difficult to meet various service requirements of 6G support for the whole society, the whole industry and the whole ecology. Instead, it leads to more complex network scale and functions.

The third aspect is that the single network structure leads to high cost and high power consumption. For example, the basic deployment of 5G is a full-function deployment based on the base station, that is, a full-function base station is configured wherever it needs to be covered, which will inevitably lead to higher network construction cost. In the 6G era, open and customized capabilities are essential. We should provide industry customers with more agile and friendly services through open interfaces and flexible network deployment, so as to better meet customers' on-demand network configuration and customized applications.

The fourth aspect is that the future network needs to have a unified access control management technology. The future 6G network is integrated with ground, air, space and sea access, with the ability to support various access modes. Therefore, 6G network should be a lightweight network, which can ensure reliable mobility management and fast service access through unified signaling coverage, ensure user's service experience and delay, and dynamically load data access through plug and play.

Patched and incremental function enhancements are difficult to meet the service requirements of 6G to support all scenarios, but instead lead to more rigid network functions and more complex network structure. In the early stage of 6G research, building a new protocol stack with powerful functions and native intelligence, designing a network architecture with deep integration of computing power, data and network, and building a multi-dimensional, full-scene access and multi-network symbiotic integration system will be the key to 6G research.

3. The next generation protocol stack architecture

3.1 Potential features of future protocol stacks

Throughout the history of the development of mobile communication, the network form has been constantly changing and evolving with the development of mobile communication technologies and service requirements. From 2G voice era, 3G text era, 4G data era to 5G Internet of Everything era, the network architecture and protocol stack function are also evolving as the requirements change. Considering the ubiquitous connection of all scenarios and the introduction of various new services, the diversification of 6G services and application scenarios further aggravates the complexity of the network. In order to solve this problem, 6G network may have the characteristics of on-demand fulfillment, lite, soft, native intelligence, native security and digital twin.

The future 6G society will face more new services, new scenarios and new user requirements, which will tend to be diversified and personalized. Therefore, 6G network should perform on-demand function deployment, parameter configuration and resource configuration, which requires the network to have dynamic fine-grained service capability supply to provide users with personalized on-demand fulfillment.

Lite

The 6G network will face the challenges of new differentiated scenarios such as ubiquitous network connection and integration of ground, air and space in the future, and the existing protocol architecture will inevitably lead to increased complexity. The 6G network needs a lite and unified protocol system to reduce the logical constraints when supporting various services. Through the integrated communication protocol and communication access technology, the unified protocol architecture, function design and process framework are adopted to realize the unified access of a variety of air interface technologies, so that new network functions and services can be introduced in a plug-and-play manner, achieving the purpose of unifying the network protocol system. The lite design greatly reduces the number of protocols and signaling interactions required for 6G network communication, thereby reducing the complexity of the network, while making it have the characteristics of powerful function, toughness, security and reliability.

Soft

The future network will be an end-to-end software-definable network that can realize rapid service deployment, automatic update and iteration of functional software versions, and self-evolution of network functions. With soft features, the network can realize end-to-end micro servitization functions, support independent network element functions, support elastic scaling and evolution of services, as well as efficient iteration and flexible deployment of the network, so as to truly realize network automation and intelligence.

Native AI

The AI capability of the network in the 6G era will be an native capability that truly realizes internal and external coordination. Internal AI can realize the on-demand supply of network capabilities and support distributed AI. And it can introduce the capabilities of external AI into the network through the intelligent platform to provide corresponding new services and support. It is also possible to share external data with the network, bringing further improvement of data efficiency and enrichment of data content. At the same time, the network's own data and capabilities can also be opened to external partners as a capability to provide services and corresponding support for the outside world.

Modular processing is one of the key features of traditional protocol stack, the functions among different protocol sub-layer are relatively independent. With the continuous development of intelligent network, the future protocol stack should have the characteristics of intelligence, compare to traditional protocol stack structure, intelligence empowered protocol stack in the future emphasis on the coordination between protocol stack modules and joint optimization can be easily achieved for any service and any procedure.

Digital twin

The fifth characteristic is by digital twin. A "dual world architecture" will be formed through digital twin, that is, a real physical world and a virtual world as an extension of the real world, corresponding to the needs of the real world, and realizing the mapping of the real world in the virtual world. Through digital twin, the network can monitor and predict the status of each network element, base station and even user service in real time, so as to reserve resources in advance or avoid accidents, to improve the efficiency of the whole network operation and service. At the same time, some new functions can be verified in advance to accelerate the introduction of new functions and realize the self-evolution of the network.

Native security

Native security is also a very important feature for 6G network. After the network has strong ability of native intelligence, AI will become an engine of 6G security in the future. Driven by AI, the network can efficiently realize the intelligent consensus of information, the intelligent defense against attacks, the self-immunity of the network and the self-evolution of network security strategies.

Reconfigurable unified wireless interface design

Design the protocol stack in the unified approach. By flexibly configuring the functions of the unified protocol stack, we can obtain various protocol stacks suitable for scenarios with different traffic characteristics. In this way, one unified protocol stack can satisfy different requirement for the wireless network, e.g. unified design for D2D air interface and the cellular air interface.

On the other hand, 5G only realized the virtualization of 5G core network, but it is possible to achieve the end-to-end network virtualization for 6G. The reconfigurable unified wireless interface design is beneficial to realize wireless access network virtualization, and enable the end-to-end network virtualization furthermore.

Native well-matched with vertical industry communication protocols

5G introduced additional enhancement to match the vertical industry communication protocols after the first version of protocol stack is completed, such as Ethernet header compression function for industrial Internet. 6G can take the adaptation for vertical industry communication protocols into account at the initial stage of wireless protocol stack design. Moreover, the maximum data length and the maximum number of radio bear supported by the protocol may need be extended.

Flexible network with higher frequency bands

6G will move to higher bands. Multiple-hop radio access network is a possible way to extend the network coverage at higher bands. The radio access network protocol and interface need support various data relaying requirements. Different kinds of relay nodes, e.g. IAB node, sidelink UE relay, smart repeater, and relay protocol architectures are designed in 5G, but still not deployed widely. Considering the costs of 6G deployment at higher bands, it is likely to have large-scale deployment of multiple-hop radio network.

On the other hand, in order to cope with the increased power consumption caused by the wider high-frequency bandwidth, 6G radio access network should be capable of flexible and dynamic adjustment to make the wireless access points denser or sparser according to traffic changes in the network.

Integrated sensing and communication

In the future, the protocol stack will expand from a single communication service to multi-dimensional services such as communication and sensing, etc. Correspondingly, the protocol function also needs to support sensing and communication integration. Sensing and communication integration can be comprehensively considered from the aspects of waveform design, parameter estimation, interference cancellation and architecture design.

3.2 The architecture of the next-generation protocol stack

3.2.1 Next generation cell model，Meta-cell

The traditional cell model cannot well meet the needs of customizable capability, elastically scalable capability, and grayscale evolution for future mobile communication networks, due to its tightly coupling between services and resources, and tightly coupling between resources and resources. Furthermore, it is constraint for the future network form and industry ecology. The future cell model needs on-demand orchestration of resources and services，and flexible networking.

The next-generation cell model, Meta-cell (meta-cell implies "source of the cell", "essence of the cell", "abstraction of the cell", or it can be understood as the future cell model) in which radio resources and services are decoupled and able to be flexibly orchestrated, supports flexible scaling and smooth evolving of network. Full-band converged networking, user and multi-parties native networking, energy and cost-prioritized networking, etc can be achieved by Meta-cell. The network form and networking with Meta-cell can adapt to different business scenarios, different service types, and achieve higher performance indicators (KPIs) and low-carbon sustainable development goals.

Meta-cell divides resources and services into different levels, and resources or services at the same level are decoupled. Resources between different levels, services between different levels, and resources and services from different levels can be flexible mapped. It can customize efficient network forms and networking modes, and even can give birth to a new business ecological models to orchestrate resources, and services in Meta-cell, according to business scenarios and requirements. The network based on Meta-cell not only is adapted to various business scenarios, but also attaches importance to efficiency and energy consumption of network and terminal. Figure 9（the right half of figure） shows the architecture of Meta-cell, which is divided from bottom to top into: carrier resource layer, channel resource layer, service layer, wherein channel layer can be further divided into transmission channel resource sublayer and physical channel resource sublayer.

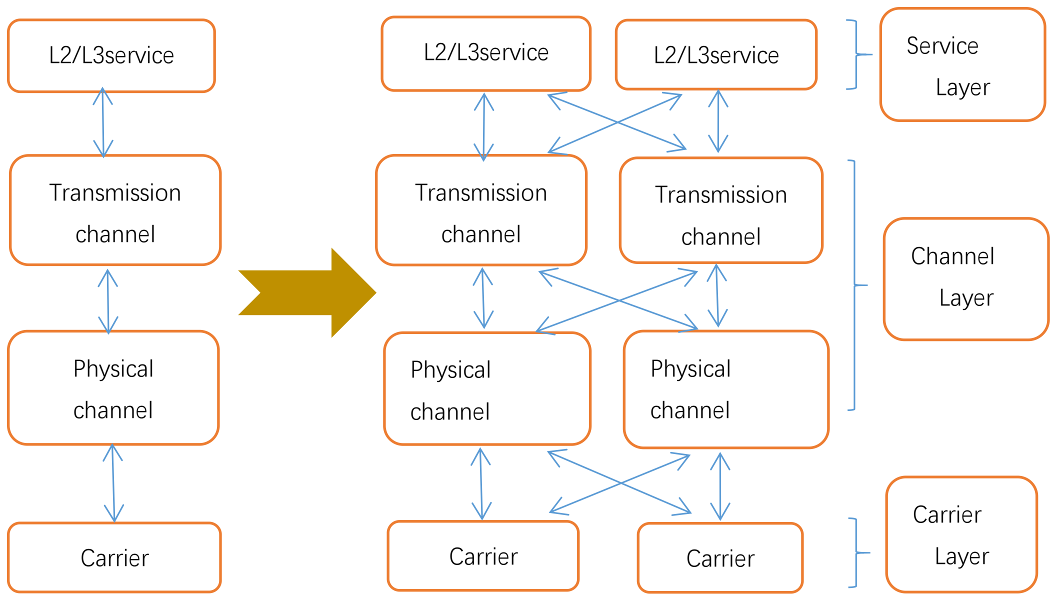

Figure 9 From traditional cell to meta-cell：on-demand multi-layer cell model

■ Meta-cell for multi-band converged networking

Meta-cell supports on-demand combination and orchestration of spectrum resources or carrier resources, which are in the same frequency band, or cross different frequency bands, or in the same or different Frequency Range (FR1/FR2), or cross different Frequency Ranges. The combination or orchestration includes: carrier concatenation to form a larger bandwidth spectrum resources for large throughput requirements; carrier selection according to different electromagnetic wave characteristics for large-capacity or wide-coverage. The above-mentioned full spectrum portfolio orchestration in Meta-cell, on the one hand, efficiently utilizes spectrum resources, and on the other hand, adapts to different business scenarios by flexible orchestration of spectrum resources.

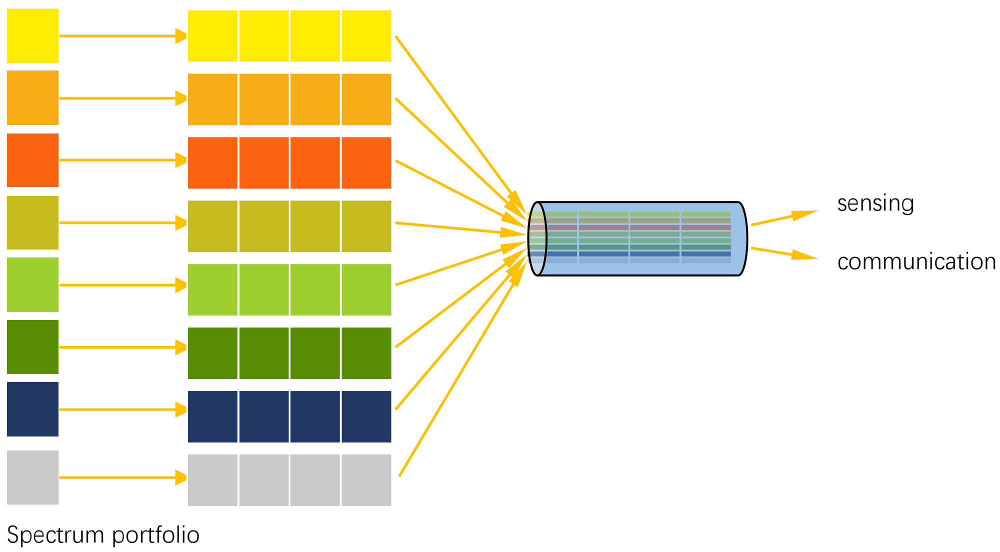

Figure 10 Meta-celll: Service-oriented full spectrum portfolio orchestration

To achieve the above goals, Meta-cell supports the decoupling of baseband processing and radio frequency processing, including that: multiple scattered radio frequency carriers are aggregated and correspond a continuous baseband carrier. Network with baseband processing in Meta-cell based on a continuous baseband carrier will achieve low complexity and workload of network management, low workload of network management, network planning and network optimization. Carrier concatenation (multiple radio frequency carriers are aggregated and correspond a baseband carrier) in Meta-cell improves the utilization efficiency of scattered spectrum (such as re-farming FDD spectrum or TDD spectrum), and expands bandwidth for physical channel (such as PDSCH/PUSCH) transmission, which increases the flow rate and reduces the overhead of configuration, scheduling, and feedback.

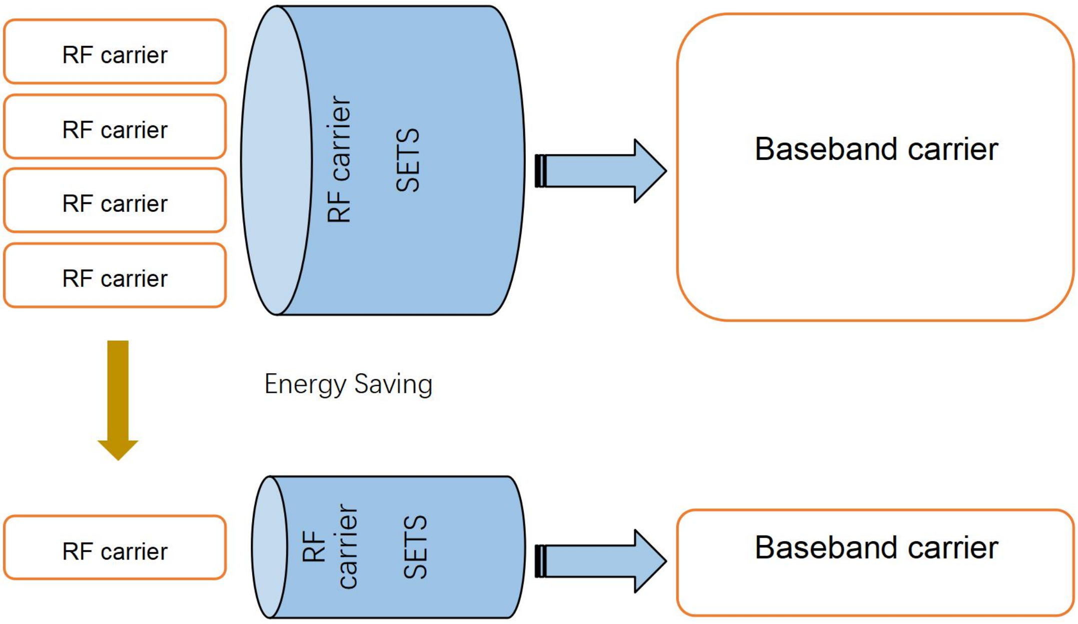

Figure 11 Meta-cell: Multiple RF carriers are aggregated into one baseband carrier

For full-band converged networking, Meta-cell supports the decoupling and aggregation of uplink and downlink to achieve on-demand orchestration and combination of uplink carriers and downlink carriers, which improves coverage and performance of network. For example, large bandwidth uplink is urgently needed in To B scenario. Meta-cell can support the aggregation of multiple uplink carriers or the aggregation of more uplink carriers than downlink carriers for To B scenario.

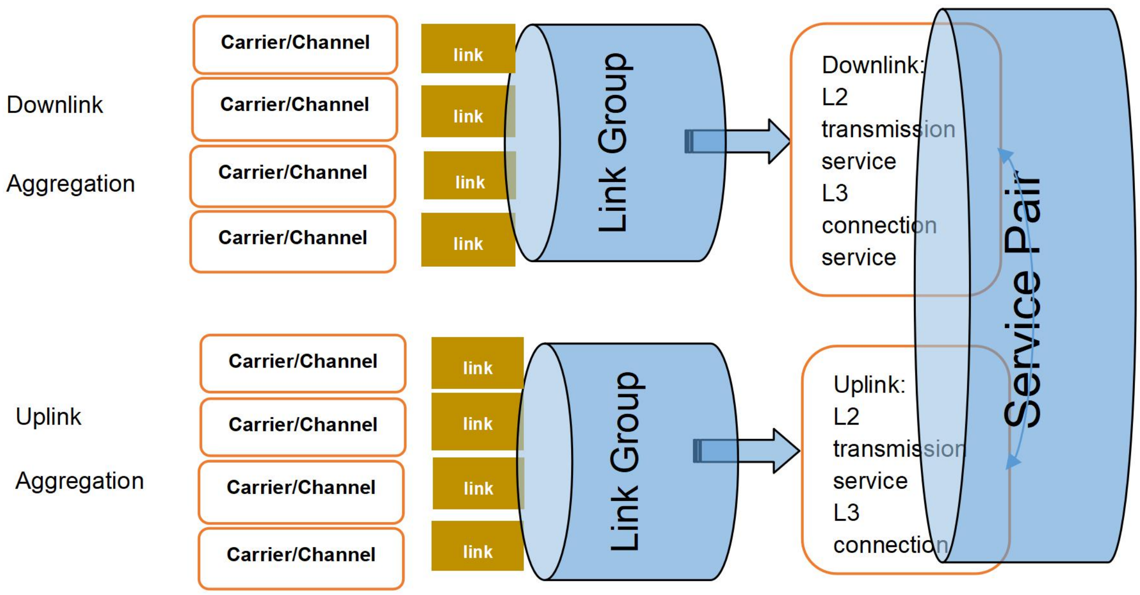

Figure 12 Uplink and downlink decoupling and service-oriented orchestration and

combination

Meta-cell supports decoupling of transmission channels to physical channels and on-demand mapping from transmission channels to physical channels. In the traditional cell (such as NR cell), large XR data packets are put in a large TB (transport block) containing multiple CBs (coding blocks), and any CB occuring error will cause that the entire TB cannot be delivered, resulting in lager latency for data transmission. In the Meta-cell, the original TB is split into multiple TBs (multiple transmission channels) so that error in any TB will have no affect on delivery of other TBs. Thereby it will greatly improve the performance of large-bandwidth and low-latency services， such as XR.

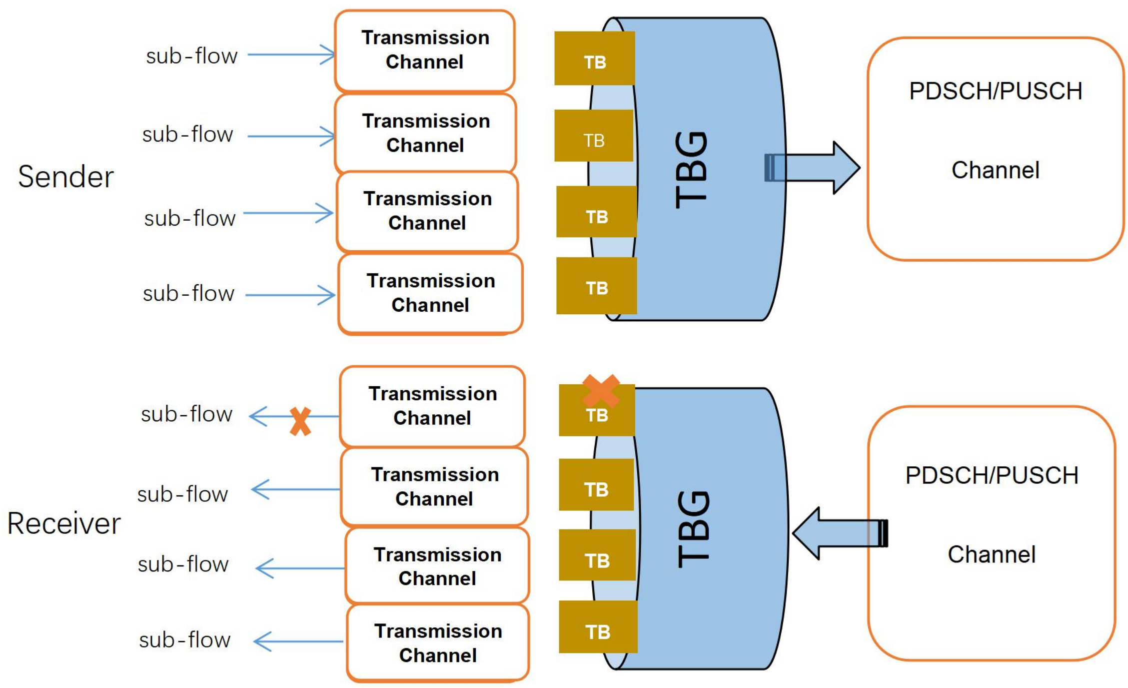

Figure 13 Meta-cell: Decoupling and aggregation between transmission channels and physical channels

■ Meta-cell for user and multi-parties native networking

In the future 6G era, in order to meet the requirements of user initiative and customization, it is necessary to consider user-centric networking, and it should be able to differentiate serving subject groups in more granularities, by using their similar features to implement optimal stereoscopic network between users and users,

users and terminals, users and multiple parties, which is called user and terminal native networking and multi-parties native networking.

User and terminal native networking and multi-parties native networking has the following forms: single-center mode (advanced and low-cost terminals collocation) and multi-center mode (peer terminals). Networking between terminals are based on D2D-like technologies. The network provides the cloud native DevOps environment for users and third parties. And users and third parties can deploy their applications in the above environment to select and use the corresponding functions as required. The applications could be application layer application(e.g. WeChat, SMS or cloud desktop) or a communication layer application. for example, an application on the RRC layer. The application of the RRC layer can arrange the networking between terminals and collaborate with the RRC application of the Mobile Operator (MO) to construct a cubic network. The advantage of the stereoscopic networking is to greatly improve the degrees of freedom: supporting mesh connection, UC-MIMO (user-coordinated MIMO), improving the air interface KPIs (coverage, throughput, deterministic and positioning accuracy) and solving the problem of high-frequency coverage.

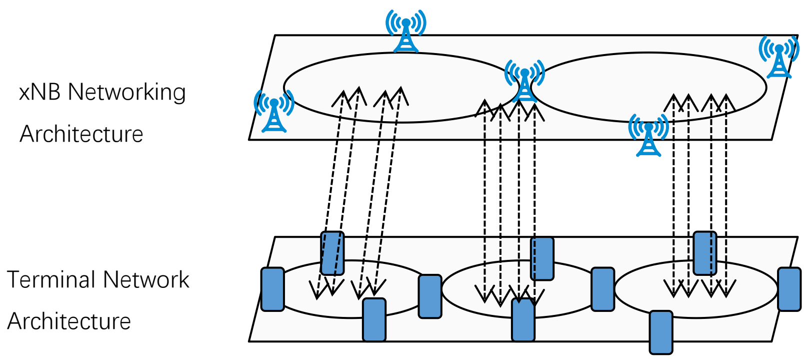

Figure 14 Meta-cell: multi-parties native and user native networking

■ Meta-cell for energy and cost-prioritized networking

Meta-Cell supports the construction network form and networking mode by micro-kernel approach, which can minimize resource overhead and energy consumption. The micro-kernel approach can be expressed as SSB resembling "connectionless" on the air interface. It can be used as a micro-kernel-style function/service to only transmit and receive SSB. By superimposing more functions/services, it can be extended successively to the system message SI function/service, access to the AC (Access) function/service, paging function/ Services, and CA/DC functions/services, etc. Since the micro-kernel approach can minimize in greatest degree the functions/services that the terminal needs to support, the extremely low-cost and extremely low-energy terminals can participate in networking in the the micro-kernel approach.

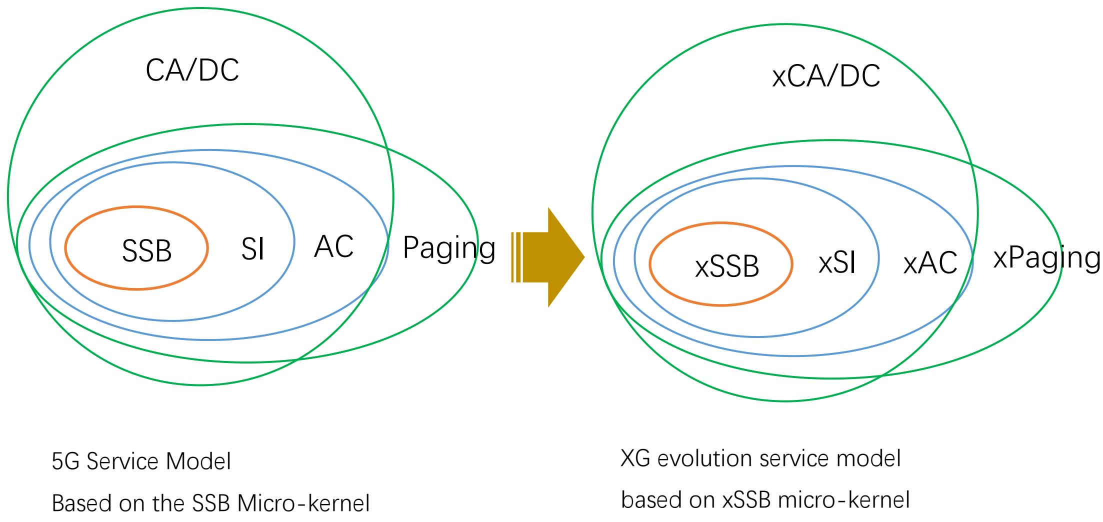

Figure 15 Meta-cell: micro-kernel to support networking with energy priority and cost priority

Meta-cell supports system message (SI) aggregation, that is, SIBs for multiple carriers can are aggregated and transmitted on one carrier, so that other carriers need not transmit SI and have a longer time “sleep”. Thereby it reduce the energy consumption and the workload of network for SI transmitting, configuration and scheduling, and simplify network operation and maintenance.

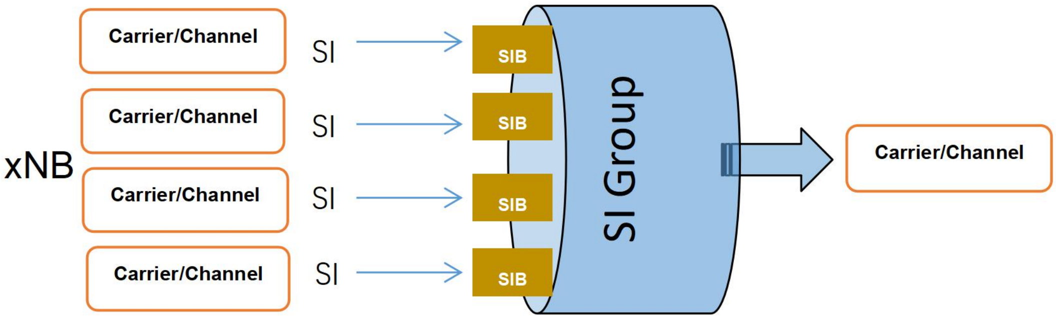

Figure 16 Meta-cell：Aggregation of system information

After receiving the aggregated system information (SI) by Meta-cell, the terminal will receive the multiple access channel (RACH) resource configurations. The access channels will correspond to different spectrum resources. The terminal can select the best spectrum for access according to coverage, frequency, and air interface quality. The optimal selection of RACH can improve the certainty (low-latency and high-reliability) of access, and quickly build connections in a frequency band with the best coverage/energy efficiency.

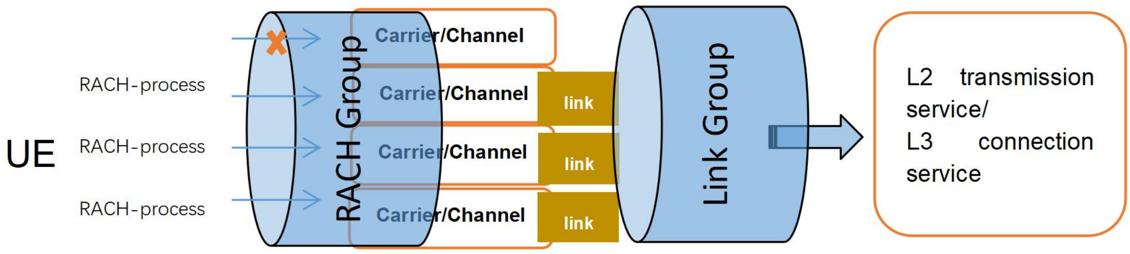

Figure 17 Meta-cell: Aggregation of system information SI, aggregation or optimized selection of RACH

3.2.2 Orchestratable Stack-Free Component-based packet forwarding protocol

The current wireless communication protocol stack architecture is designed based on OSI seven layer model. The functions contained in each layer are fixed and the order is unchanged. This tightly coupled protocol stack greatly limits the development of 6G network architecture. Facing the new requirements, new scenarios and new services of 6G, the forwarding plane function needs to meet the requirements of flexible orchestration, customization, generalization and processing ability improvement. Therefore, 6G needs to introduce Orchestratable Stack-Free Component-based packet forwarding protocol(OSFC).

Facing the future user native and multi-parties native networking, and meeting the requirements of different scenarios, different service types and the pursuit of higher performance, the forwarding plane related protocols will develop in the direction of componentization and parallelization of basic functions, including basic packaging (concatenation/segmentation), reliability assurance, security assurance, flow splitting and flow control, routing and mapping, etc. Based on the componentization of the basic functions of forwarding plane processing, an orchestrable stack-free component-based packet forwarding protocol（OSFC） is formed by flexibly orchestrating and configuring relevant components. The packet forwarding protocol has the characteristics of componentization, vectorization and parallelization.

- Function componentization: The architecture of the forwarding plane of the traditional communication system is very rigid. Introducing new functions or changing a function will bring a chain reaction to all parts of the architecture, and the same functions of different layer cannot reuse the same process. In order to reduce the coupling and redundancy between various functional modules of the forwarding plane, the component-based design principle of independence, no redundancy and free combination is very applicable.By componentizing the functions of the forwarding plane, the inherent function order of the original protocolt can be broken, the order of functional components can be changed, the components can be on/off or added/deleted dynamically, and the customization can be carried out according to personalized requirements. At the same time, the componentized function is extracted common components and common processes, so as to reduce the complexity of the protocol. In the new forwarding plane, independent components converge to form a component database. Through the configuration, orchestration and management of components in the component database, the forwarding plane can adapt to build a data transmission chain according to various personalized needs. The component-based design provides the possibility for the flexible arrangement of serial mode, parallel mode and serial-parallel mixing mode. This ensures that the components can be orchestrated flexibly and scaled elastically on demand in 6G's complex and changeable scenarios.

- Vector Packet Processing(VPP): In the field of IT, in order to quickly build switching and routing functions, an extensible and high-performance packet processing framework vectorized packet processing (VPP) is introduced. The main idea of VPP is to vectorize the continuous packets with the same characteristics for batch process, which is much better than the traditional single packet quantization. VPP design is introduced into the forwarding plane to vectorize the data packets, and batch process the continuous data packets of performing the same operation, which can significantly improve the processing speed of data packets. Based on the componentization of forwarding plane, the granularity of packet vectorization can be changed flexibly. For example, packet vectorization can be performed on a component, or packet vectorization can be performed on one or several steps within a component

- Component parallelization(PC): On the basis of componentization, the forwarding plane can also realize multi-component parallel processing for a single packet. These parallel components have the characteristics of independence and statelessness, and the change of components order does not affect the processing of data packets. Components with these characteristics can be orchestrated in parallel to reduce data processing delay. The parallelization of components increases the dimension of component orchestration, enables components to form a variety of combinations such as serial, parallel and serial parallel hybrid, and increases the flexibility of forwarding plane. In general scenarios, it is more common to use the mixed orchestration of component parallel and component serial.

In OSFC, componentization decouples the forwarding plane functions to form a stack-free protocol structure. Flexible component orchestration and packet vector quantization realize the dynamic adaptive matching of traffic requirements in 6G scenarios. According to the specific traffic scenario characteristics, packet types, component functions and other needs, the protocol architecture can be flexibly orchestrated into many forms, such as multi-component serial, multi-component serial + VPP, multi-component parallel, multi-component parallel + VPP, multi-component serial parallel hybrid and multi-component serial parallel hybrid + VPP. The protocol also supports cross domain or multi-hop deployment, which is very friendly to the adaptability of multi-level deployment, mesh networking and other scenarios. The orchestrable stack-free component-based packet forwarding protocol OSFC fundamentally breaks the hierarchical concept of protocol stack and deepens the cooperation between components. It is key to improve the data forwarding and processing ability of 6G network, and plays an important role in building a service-oriented adaptive flexible network. Moreover, through the intelligent configuration, intelligent orchestration and intelligent management of the forwarding plane components, the forwarding plane can introduce big data and AI algorithms to further improve the intelligent level.

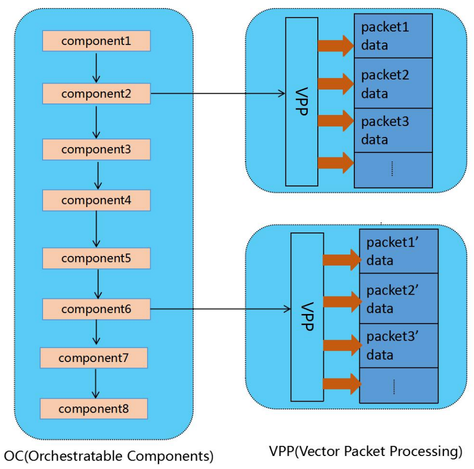

Figure 18 multi-component orchestration with VPP

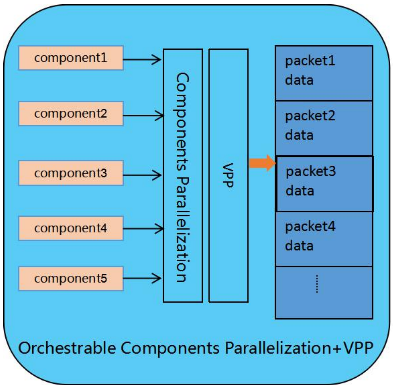

Figure 19 Single-hop OSFC with VPP

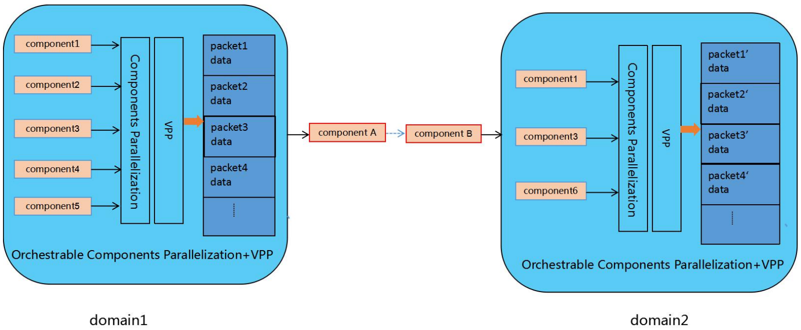

Figure 20 Multi-hop OSFC with VPP

3.2.3 Cloud native service based network

The design of the 5G core network architecture adopts the cloud native SBA idea in IT field, but the access network still retains the original layered architecture. These two different architectures make it difficult for the core network and access network to mutual invocation, co-deployment, and O&M. With the commercial use of 5G, The demand for 6G network flexibility and ICDT integration put cloud native 6G wireless access network on the agenda. And the cloud native wireless access network architecture also accelerates the integration of core network and access network. Cloud native concepts includes containerization, micro-service, DevOps, Conways law, etc. These concepts run through the whole process of cloud native architecture design of wireless access network. By further enhancing the ability of rapid integration, customization, orchestration, configuration, and extended iteration, the next generation network extends the service-based architecture for core network to End-to-End Service-based Architecture (E2E-SBA). For the purpose, the radio access network should be refactored with SBA to support SBA-RAN. It should be noted that while the UE and radio interface will also be redesigned in certain aspects with the basic guideline of SBA. From the efficiency perspective, it is not suitable to copy the service-based interface to the radio interface . Functions of real-time requirement still need to use original layered architecture).

Refer to the concept of cloud native, tool chains, and domain-driven design， the cloud native architecture decouples domains with different life-cycles and forms function sets based on independent life-cycles and service logic. Each domain has its own life-cycle management. So that the next generation network can carry out independent iterative update, grayscale evolution and cross domain CI/CD in different domains.

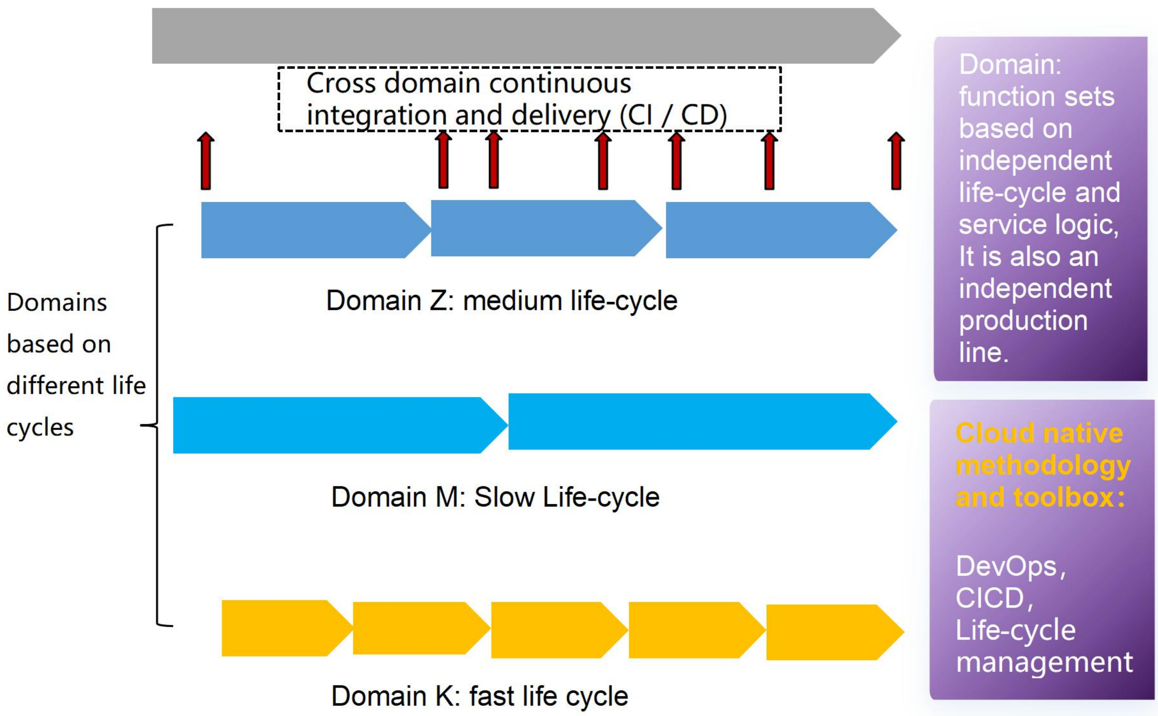

Figure 21 Domain-Driven Design(DDD)

Based on different service logic and life-cycles, cloud native service based network is divided into PHY domain, PKT (packet) domain and HUB (control hub) domain. Each domain can be partitioned according to functions attributes into enforcement plane, control plane, storage plane, intelligence plane and OAM plane. For example, the PHY field is divided into five functional planes: PHY-E, PHY-C, PHY-S, PHY-I and PHY-O. With such design, functions with different attributes can have different implementation platforms and deployment locations. They can be also orchestrated scalably on-demand. In addition, the control plane, data plane, and intelligent plane can be flexibly orchestrated. Driven by their evolved requirements for further enhancement and evolution. All functions can be continuously orchestrated, customized, configured, and integrated into a network system.

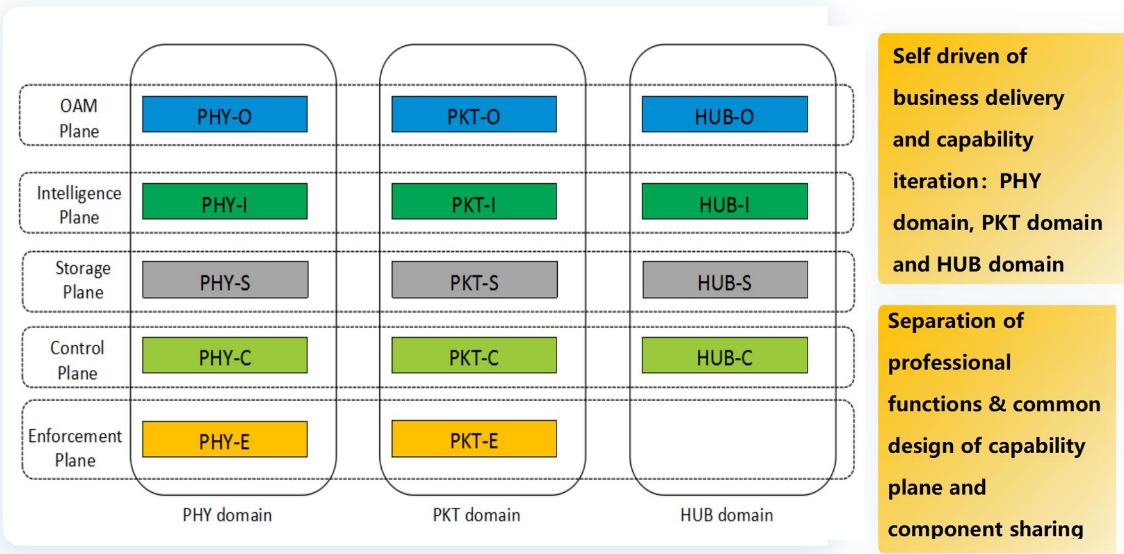

Figure 22 Two dimensional division of service network architecture

The cloud native service based network not only reshapes future infrastructure ecology, but also provides a deployable, multi-parties collaborative environment to realize domain level / function level interconnection. And the cloud native service based network can flexibly support the multi-parties native networking, user native networking and the integration networking of communication, sensing, computing in the future 6G network.

3.2.4 The generalized QoS framework

6G network supports new forms and new technical features in multi-parties native and user native networking, and the integration networking of communication, sesing, computing, intelligence and trust. The traditional QoS framework of communication system can not adapt to the new development of 6G, so it is very necessary to introduce a new generalized QoS framework. The new generalized QoS system adopts a generalized perception engine including intention, business, user, environment and scene perception, and uses the project-oriented principle for intelligent orchestration. The workflow formed by the orchestration is used to control the work decomposition and execution in different domains such as communication, sesing, computing, intelligence and security domain. The project-oriented framework can realize the entire closed-loop and whole process control of demand, delivery and settlement, and form an traffic framework of value guidance and deterministic management. The project oriented generalized QoS framework decomposes the project into services in various domains of communication, sesing, computing, intelligence and security, and carries out cross domain coordination of QoS and deterministic guarantee of QoS in each link. Each link has QoS control corresponding to this link, relating the QoS of all links to ensure the overall end-to-end QoS requirements. In the generalized QoS framework, it supports external input, QoS workflow orchestration, QoS configuration at all levels, and policy scheduling according to QoS requirements at all levels. This new project oriented generalized QoS framework forms a workflow through flexible orchestration on demand, realizes the whole process and value chain closed loop of information flow, decision flow, execution flow and feedback flow, and can meet various service needs and quality assurance.

The generalized QoS framework covers the domains of communication, computing power, sesing, AI and the big data. There are different in metrics for different domains, and the metrics can not be simply aggregated and stacked and should be comprehensively considered in generalized QoS. That is, the metrics of the generalized QoS framework need to consider the integration after communication, sesing, computing power, AI and data at the beginning. For example, in the communication traffic scheduling, in order to ensure the communication performance, the computing power needs to be guaranteed with deterministic; In AI and big data , the balance between computing power cost and power consumption needs to be considered. Based on multi-domain integration, generalized QoS needs to determine the weight of various metrics, establish a evaluation mechanism, and form a comprehensive metrics system that takes into account not only KPI (Key Performance Indicator) but also KVI (Key Value Indicator).

3.3 Review on 3.0 protocol stack function

The Next Generation Protocol Stack White Paper 2.0 analyzes the main problems faced by the network and important scenarios for future network deployment, and explores better protocol stack architecture and functions. One of the main issues discussed in the Next Generation Protocol Stack White Paper 2.0 is how to ensure the high reliability of data transmission and the ultra-low delay requirement of a single service while meeting the low transmission resource cost. A variety of solutions proposed in The Next Generation Protocol Stack White Paper 2.0 include multi-connection simultaneous transmission, single-link multi-transmission, MAC and PHY protocol layer solutions.

The Next Generation Protocol Stack White Paper 3.0 further studies the protocol stack architecture and protocol layer functions, and analyzes some existing problems in the network. For example, there are deficiencies in cloud CU/DU. The strong binding between CU/DU and protocol stack function hinders the flexible on-demand deployment of the protocol stack, resulting in limited network opening capacity and the inability to achieve accurate opening and cloudification of network capacity. In addition, the problems faced by core network slicing and access network slicing are also studied. According to the requirements of customers in the vertical industry, network slicing needs to refine the SLA requirements of end-to-end network slicing, decompose them into core network, wireless and transmission domains, and select proper templates for slicing instantiation. In order to effectively define SLA and carry out quality assurance, there is still a lack of end-to-end monitoring and dynamic adjustment mechanism for closed-loop operation of network slice.

The next generation protocol stack white paper 3.0 analyzes the next generation protocol stack from three directions: servitization, componentization, and intelligence. For servitization, with the continuous expansion of service scenarios and deployment scenarios, the requirements of network architecture for adaptability, including flexibility, soft, scalability, evolution and recoverability, in addition to the traditional efficiency requirements, continue to improve. In view of this, the service-oriented architecture with better adaptability naturally borrows from IT to CT: 5G core network innovatively adopts a service-based architecture that takes into account adaptability and efficiency. However, from a global perspective, an end-to-end service architecture can fully meet the requirements of overall network adaptability. For componentization, the function of the forwarding plane must first meet the basic requirements of security, efficiency, reliability, QoS control, etc. In order to support the rich requirements of MEC, vertical industries, integration of ground, air and space, integration of access and backhaul, dual connectivity, CA, URLLC, interoperability and other scenarios for the forwarding plane, the forwarding plane also needs to meet the advanced requirements of universality, independence, compatibility, scalability, portability, openness, etc. Therefore, the forwarding plane needs to adopt a component-based design idea, and comprehensively consider the basic and advanced requirements of the forwarding plane to split the protocol stack into orchestrable modular components, in order to achieve the balance between the basic requirements such as efficiency and reliability and the high-level requirements such as compatibility, scalability and openness.For intelligence, intelligence is one of the important driving forces and enablers of the next generation access network architecture. Its influence can provide micro empowerment in all functions of the access network architecture, or provide meso empowerment beside these functions, and can also surpass all functions of the current access network to provide macro empowerment. The existing mobile communication systems combined with AI are mostly cloud-based centralized learning methods, that is, a large number of training data are transmitted to the cloud, and the corresponding decision-making model is distributed after deep learning. This not only brings the problem of delay, but also has high requirements for transmission bandwidth, which can not meet the real-time requirements of service. In the future, with the significant increase of computing resources on the edge access side and the evolving network demand of differentiated services, AI and access network will continue to be deeply integrated.

It can be seen that although different types of services or user requirements emerge at different stages, the network may face similar problems at different stages, and some functions need to be continuously enhanced. The enhancement of network architecture and protocol stack functions is a spiraling and continuous process.

4. Next-generation protocol stack enhancement

4.1 Space-Air-Ground-Sea Integrated Network

The Space-Air-Ground-Sea integrated network is not a simple interconnection of satellites, various types of aircraft, and cellular network facilities on their original basis. It is necessary to fully integrate multiple access types in the system level, which achieve deep mergein terms of architecture, protocols, services, and terminals. And it can provide users with continuous, non-switch-aware, flexible ubiquitous communication services.

Satellite communication network: Considering both the network coverage and service quality at the same time, the satellite-ground integrated network architecture is the key research direction for the construction of the SAGS integrated communication system. The industry has proposed a variety of integration architectures including satellite as non-3GPP access and 3GPP RAT access. Among them, the former only connects satellites to the 6G core network, while the latter is a deep integration of satellite network and terrestrial cellular network. In Rel-17, 3GPP has started the standardization work related to satellite communication mechanism enabled by 5G. The air interface adopts the 3GPP enhanced protocol, however it only focuses on the deployment structure based on the transparent mode, i.e. the satellite is used as a forwarding relay. Only the most basic functions are supported in Rel-17, performance optimization will be considered in subsequent release, including support for dual connectivity and carrier aggregation, coverage enhancement, IoT terminals, etc. Meanwhile, new functions will be introduced, including network-based positioning, support for terminals without GNSS positioning capabilities, broadcast and multicast services, etc. With the advancement of related research and technology, 6G base station and even core network function can be deployed on the space-based network. This will require regenerative payloads on board the satellites and the inter-satellite links (ISLs) connecting them. In the future, it is very beneficial to comprehensively consider the architecture design of the satellite-ground integration network according to the different integration architectures of the satellite network and the terrestrial network.

Marine communication network: The marine economy, especially the communication involving marine machines, is receiving more and more attention from all parties. The sea surface and underwater communication coverage are expected to be solved in the 6G era. Marine communication networks mainly include shore-based mobile cellular communication systems, maritime wireless communication systems, satellite communication systems, and underwater communication systems. At present, each system is independent of the other, the interconnection and information transmission are not smooth, which cannot meet the needs of new marine applications. How to coordinate multiple platform resources and provide unified network architecture support is an urgent problem. We must ensure the universal connection from the coast to the ocean, data transmission and relay can be carried out between ships, ships and satellites, and various offshore platforms to ensure the continuity of maritime services. At the same time, it is necessary to integrate capabilities including navigation, data monitoring, and emergency communications to build an integrated ocean communication perception network and improve the comprehensive level of marine informatization.

Space-Air-Ground-Sea Integrated Network: In essence, SAGS integrated network is a layered and heterogeneous structure. Different access modes in integrated network have their advantages and disadvantages in terms of coverage, transmission delay, throughput, mobility, reliability, etc. In order to achieve the real system fusion networking, it is necessary to adopt effective air interface and network management technologies, including dynamic spectrum sharing and interference coordination, intelligent access and mobility management, as well as intelligent management and scheduling of cross-domain resources, etc. Specifically, as an essential part of SAGS integrated network, satellite systems have a series of characteristics such as complex heterogeneity, time-varying topology, and limited node resources, which brings huge difficulties to network management and resource allocation. Based on the above discussion, an intelligent, simplified and universal protocol that enables cross-network data transmission, though highly desirable, becomes a major challenge. SRv6 based on IPv6 and segment routing technology simplifies network protocol types and has the advantages of strong scalability and programmability, as well as supporting large-scale networks. It has good application prospects and can be used as a candidate possible solution. Another point of interest is to simultaneous introduce new technologies such as artificial intelligence to help improve the efficiency of routing mobility management, complete rapid networking in a dynamic environment, and satisfy the needs of integrated slice management at the network layer level. Furthermore, the combination of SRv6 and Software Defined Network (SDN) technology, which can achieve the elastic reconfiguration of the network and carry out flexible and efficient resource scheduling. The SDN controller can maintain an integrated space-air-ground-sea network topology structure, and perform real-time network resource scheduling, routing strategy selection, as well as protocol stack segmentation as its dynamic changes.

4.2 Intelligence based RAN

5G raido network resource allocation, resource management and scheduling are network-centric. In the future, massive smart devices and sensors, greater connection density, more diversified user requirements, more frequent intelligent interactions and potentially high network energy consumption bring great challenges to 6G. 6G needs to study user demand centered and intelligence based dynamic self-organizing wireless access network technology. 6G will be aware of user demands, service requirements and spectrum usage state changes in real time, and self-adapt the real-time changing user demands, service requirements and spectrum demands, in order to achieve flexible usage of the end-to-end virtual resources, physical resources and spectrum resources for dynamically shaped wireless access network topology.

The 6G spectrum is rich and diverse. ‘Dynamic and intelligent spectrum selection’ can match the appropriate operation spectrum for specific users and specific services through intelligent perception of spectrum usage and analysis of frequency band propagation characteristics suitable for service requirements, make full use of the available spectrum resources and control the interference in the network. Intelligent spectrum sensing technology and intelligent spectrum sharing technology will be important supporting technologies. The full-duplex technology can enable ‘Dynamic and intelligent spectrum selection’ to utilize spectrum resources more flexibly.

6G radio access points are diverse and the wireless propagation environment is complex. ‘Dynamic and intelligent routing selection’ can match suitable access points, fixed relay nodes, mobile relay nodes, or even reconfigurable smart surface panels for specific users and services based on the intelligent perception and prediction of the dynamic wireless environment and node load, select the best transmission path, reduce the wireless network congestion and interference, improve network coverage and ensure user experience.

The 6G network topology is complex and the access points are densely distributed. With the smart sensing and prediction of user demands and service demands distribution in time and space, ‘Dynamic and intelligent topology selection’ can identify the access points that need to be activated or hibernated and dynamically adjust wireless network topology and network configuration with artificial intelligence technology. Real-time awareness of the quality of user experience and quality of service are also important input for the ‘Dynamic and intelligent topology selection’. At the same time, the sharp increase in the number of 6G access points and users makes the amount of data and signaling to be transmitted on the back link increase sharply. Thus, it is necessary to study the efficient wireless backhaul mechanism with high capacity.

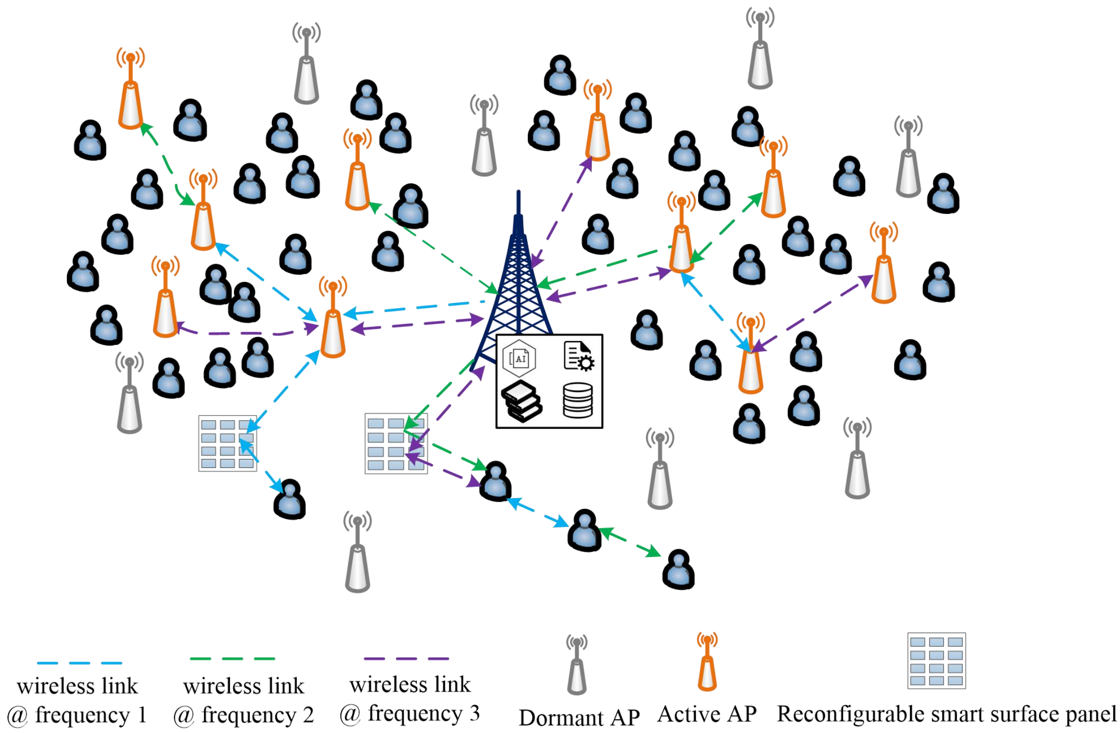

Figure 23 An example of intelligence based RAN scenario

User demand may change in time and space. Dynamic self-organizing wireless access network technology can integrate the network resources that are scattered in different geographical or logical locations to provide users with timely and effective services. In the pursuit of wider and better coverage, support the universal existence of artificial intelligence, at the same time it can control the growth of network energy consumption and improve network energy efficiency to achieve sustainable development.

4.3 Adaptive protocol stack for new architecture to support RAN nodes cooperation

4.3.1 New architecture to support RAN nodes cooperation

In “The next-generation protocol stack v3.0”, we proposed two types of new RAN architecture: HTC support architecture and separation of user and UEs. In HTC support architecture, an anchor UE can trigger multi-network node to transmit services cooperatively, according to the location information of the anchor UE. In separation of user and UEs architecture, multiple UEs serve one or a group of users. The two types of architecture can be combined as architecture to support RAN nodes cooperation.

With the introduction of new services, applications and scenarios, RAN nodes including gNBs and UEs should work together to satisfy the QoS requirement of a specific service, such as Holographic Communication service; or improve system performance in specific scenario, such as multiple devices in proximity belonging to one user.

For Holographic service, different profiles could be presented by different nodes, especially in large scale activity. The end points could be gNB(s) and/or UE(s). And different flows belonging to one holographic service should be transmitted to corresponding end points via separate interfaces including wired and wireless interface such as Uu and sidelink. In this architecture, the functions and relationships of RAN nodes should be redefined and remodelled, such as introducing L2 protocol architecture for multiple nodes cooperation. More detailed, RAN procedures such as access control containing participating node(s), system information, paging, and mobility should be studied considering multiple participated RAN nodes. In user plane, QoS satisfaction for coherent flows of specific service(s)/application(s) which could be transmitted/ received in multiple terminals should be studied, such as synchronization among the coherent flows. With the cooperation among gNBs and UEs, the new upcoming holographic service can be served in communication system practically.

For the scenario that a user owns multiple devices, UEs can be cooperated in RAN higher layer or physical layer. Service continuity should be guaranteed in different terminals. Thinner protocol stack and diversified controlling mechanism over air interface are applicable. Other technologies for UE cooperation/aggregation include security for a group of UEs, connection control for the UE group, capability coordination and etc. With UE cooperation/aggregation, adaptive network deployment can be realized by adding/deleting UEs in a UE group dynamically while keeping service continuity; uplink transmission can be enhanced to improve system performance.

4.3.2 Flexible RAN node split

In 5G R15 IAB SI, various proposals had been raised to split the RAN NW node, namely into one “Central Unit (CU)” and one or more “Distributed Units (DUs)”. These methods were captured in 3GPP TR 38.801, numbered according to the UP layer whence the RAN node is split. At that time the SDAP layer was not introduced yet and thus not mentioned during the discussion. Finally, option 2 is adopted that PDCP is located in CU and other L2 sublayers are located in DU.

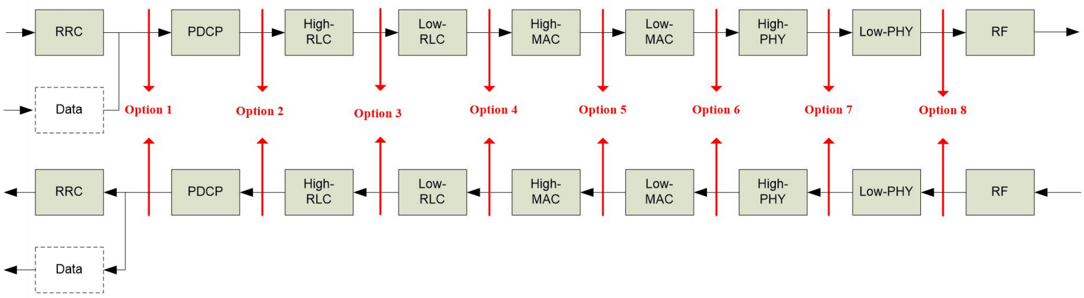

Figure 24 Candidates of CU-DU split in 5G SI

The motivation of flexible RAN node split for all of the splitting approaches studied in TR 38.801 still persists. It can be a good direction to consider when designing the 6G network.

The ideal way is to make every module flexible to deploy and configured, e.g. one RAN node can decide to use Option 2 for one radio bearer (i.e. utilising RLC/MAC/PHY of another node for this radio bearer), whereas to use Option 3 for another radio bearer (i.e. offloading the RLC/MAC/PHY but holding the ARQ function) and Option 6 for another more (i.e. offloading only the PHY toward another node). With the development of inherent AI and cloud computing in distributed RAN nodes, flexible RAN node split can be adapted dynamically with or without pre-deployment and applied with flexible wireless link selection (Uu, sidelink, or IAB, UE relay, etc.).

To apply flexible RAN node split, security of multi-CU-UP connection should be reconsidered. Taking gNB-CU-CP/UP split in 5G as an example (Figure 25), one gNB may contain one gNB-CU-CP and one or more gNB-CU-UPs. It is a common understanding that different gNB-CU-UPs which are designed to provide different services (typically of different slices) are deployed at different positions. However, the security architecture for 5G RAN still forces all the DRBs serving a given UE to use the same algorithm keys for encryption and integrity protection respectively. And on the other hand, there is another common understanding (i.e. security isolation) that one security key should not be used in multiple "security domains". As the result, it almost prevents utilising multiple types of gNB-CU-UPs to serve one UE, and further hinders the use of slicing in RAN.

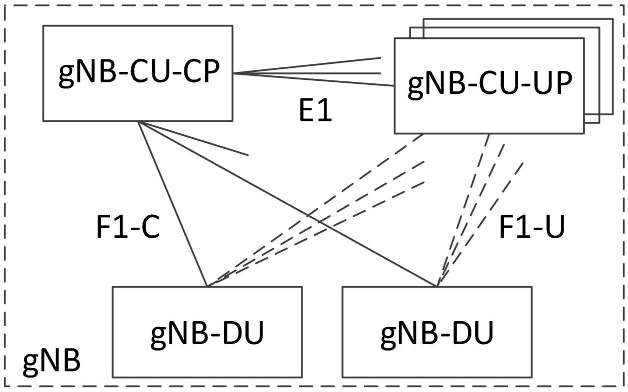

Figure 25 Overall architecture for separation of gNB-CU-CP and gNB-CU-UP

This problem is widely acknowledged but left without any solution, mainly due to its low cost-effectivity, i.e. it is not worthy to make any big change on RAN security structure once the 5G network is already taken in to use.

For 6G, as the entire security architecture will be designed from the beginning. It is very beneficial to take multiple-CU-UP connectivity into consideration when designing the 6G security architecture, so that CU-UPs are unaware of the security key used its peers even if they are serving one UE, thus facilitate the virtualisation and enhance deployment flexibility of RAN UP functions. In flexible RAN node split, even with different L2 protocol(s) adopted in different DUs, the security algorithm and/or key(s) will be applied.

4.3.3 L2 protocol split/aggregation

In the new architecture supporting RAN nodes cooperation, L2 protocol can be splitted or aggregated in both gNB side and UE side based on RAN node split and service flow characteristics (such as co-flow requirements).

Depends on flexible RAN node split in section 4.3.2, L2 protocol split/aggregation can be:

- SDAP split/aggregation;

- PDCP split/aggregation;
- RLC split/aggregation;
- L1 split/aggregation.

Moreover, considering the multi-stream transmission in HTC, L2 protocol split/aggregation should be applied in SDAP sublayer or PDCP sublayer. And PDCP split/aggregation or split/aggregation in other sublayers can be applied to the scenario that multiple UEs belonging to one or more users.

Some examples of L2 protocol split/aggregation are shown below:

- SDAP split/aggregation

Assuming different gNBs transmit part of sub-flow(s) of a HTC (or XR) service to one UE, SDAP split is performed by gNBs and SDAP aggregation is performed by UE.

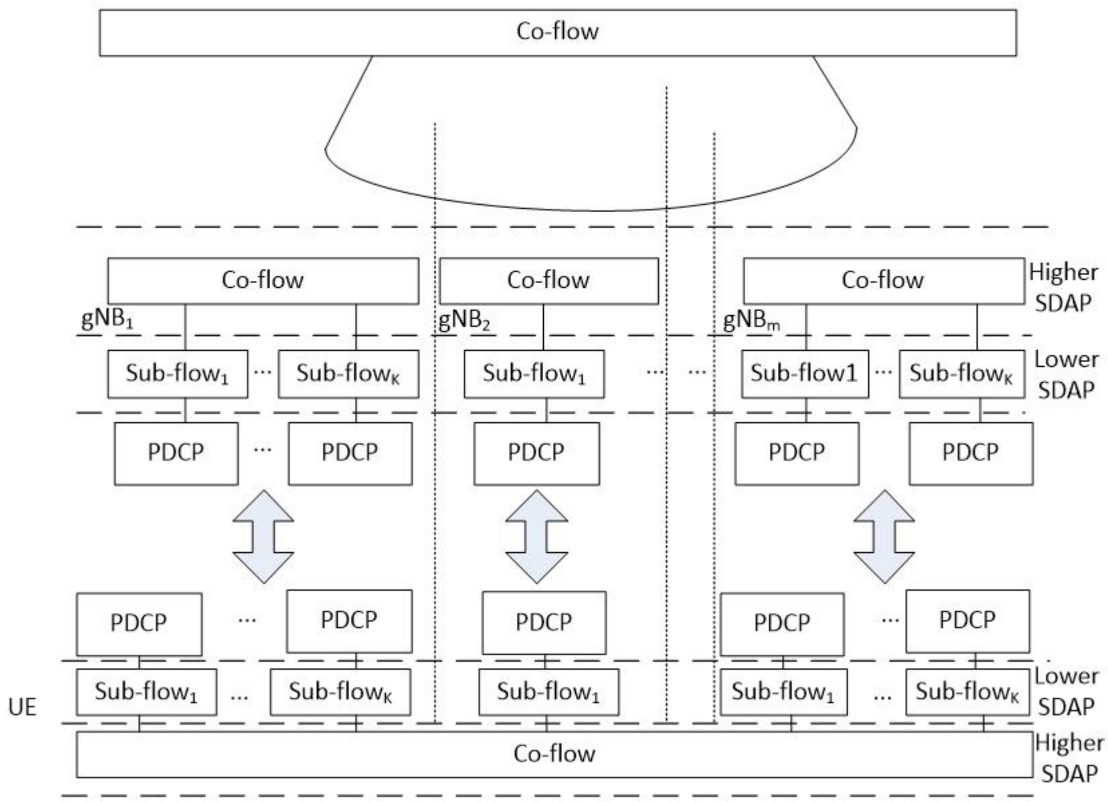

Figure 26 An example of SDAP split/aggregation

- PDCP aggregation

For the scenario that multiple UEs belonging to one user or HTC service associated to multiple RAN nodes, PDCP aggregation can be used. For example, multiple UEs can transmit same or different subflow(s) to gNB, and gNB can aggregate the transmissions by PDCP aggregation as in 29.

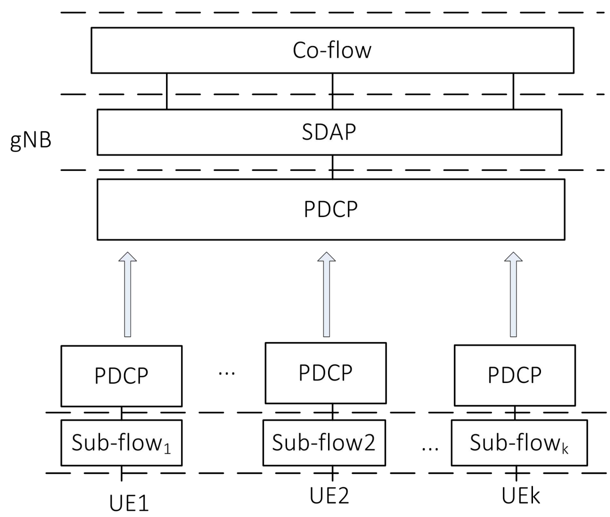

Figure 27 An example of PDCP aggregation

4.4 Integrated sensing and communication

In terms of mobile network standardization, UE-oriented positioning has been standardized as a generalized sensing service in the 3GPP Rel-16 [9], and will be further improved in subsequent releases. In terms of WiFi standardization, the IEEE 802.11 bf working group is defining standards based on WLAN to support high-value use cases and solve interoperability issues [10]. For the scenarios of integrated sensing and communication, the sensing target can be UE in the wireless network, non-communication device (e.g. obstacle) or geographic area. Therefore, it is necessary to further enhance the protocol stack design for different integrated sensing and communication scenarios in order to meet the requirements from both communication and sensing at the same time.

According to the sensing signal sending node and the receiving node, it is divided into the following 6 types of sensing links, as shown in the Figure 28. We just take one sending node and one receiving node as an example in the Figure. More than one sending nodes or receiving nodes can be used according to the sensing requirements. During the sensing process, we can also use single sensing links or more than one sensing links in different scenarios. According to the existing security and privacy rules, the sensing function in the core network is responsible for sensing results based on the sensing measurements obtained from the above sensing links.

1. Uplink sensing: the base station receives the sensing signal sent by the UE and calculates sensing measurements based on signal processing.

2. Air interface sensing between base stations: a base station receives the sensing signal sent by another base station, and calculates sensing measurements based on signal processing.

3. Base station echo sensing: the base station sends the sensing signal, and calculates sensing measurements by receiving the echo of the sensing signal.

4. Downlink sensing: the UE receives the sensing signal sent by the base station and calculates sensing measurements.

5. Sidelink sensing between UEs: the UE receives the sensing signal sent by another UE, and calculates sensing measurements.

6 UE echo sensing: the UE sends a sensing signal, and calculates sensing measurements by receiving the echo of the sensing signal.

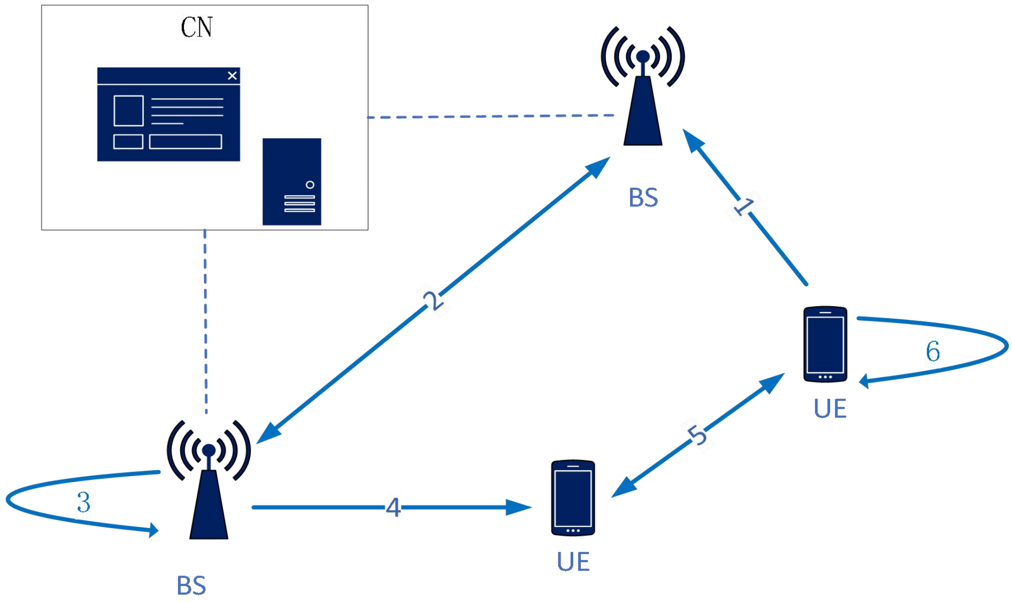

Figure 28 Overview of sensing links in integrated sensing and communication

For the 5G-A integrated sensing and communication protocol stack design, radio access network needs to support L1~L3 sensing-related processing according to sensing request from core network sensing function based on the end-to-end architecture. Sensing-related processing include sensing parameters configuration, sensing resource scheduling, sensing signal generation, sensing measuring, and sensing measurement reporting. A potential NG interface reusing solution is to introduce sensing functions and the NR sensing protocol (NRSPa) in the core network, considering the compatibility with the existing 5G protocol stack. NG-RAN nodes report sensing measurements to the sensing function through the newly introduced NRSPa. The sensing function is responsible for generating the final sensing result or information, and providing services to the sensing service requester. In some sensing scenarios, the volume of reported sensing measurements may be large. In order to improve the efficiency of the control plane-based solution, another potential solution is reporting sensing measurement over user plane protocol. If the sensing measurements are reported by the NG-RAN node, introducing a new interface between the NG-RAN node and the sensing function also can be an option in addition to reusing the NG interface. If the sensing measurements are reported by the UE, measurements in the user plane-based solution needs to be forwarded through the UPF, considering the compatibility of the existing 5G protocol stack.

When UE is the sensing signal sending node or receiving node, L1~L3 needs to be extended to support sensing. L3 RRC needs to support sensing parameter configuration, including subcarrier spacing, bandwidth, sensing signal duration, time domain spacing, sensing measurement, etc. L2 MAC needs to schedule time and frequency resource used for sensing, according to the priority of communication and sensing. At the same time, when NG-RAN node or UE is used as the transmitter of sensing signals, L1 needs to support sensing signal generation. When NG-RAN node or UE is used as the receiver of sensing signals, L1 needs to perform sensing on the received sensing signals according to the sensing configuration information. Then the mentioned receiver of sensing signals needs to report the calculated sensing measurements. When UE reports the sensing measurements, a potential solution is to add the NR sensing protocol (NRSP) function on the UE side. Then UE communicates with the sensing function via NRSP.

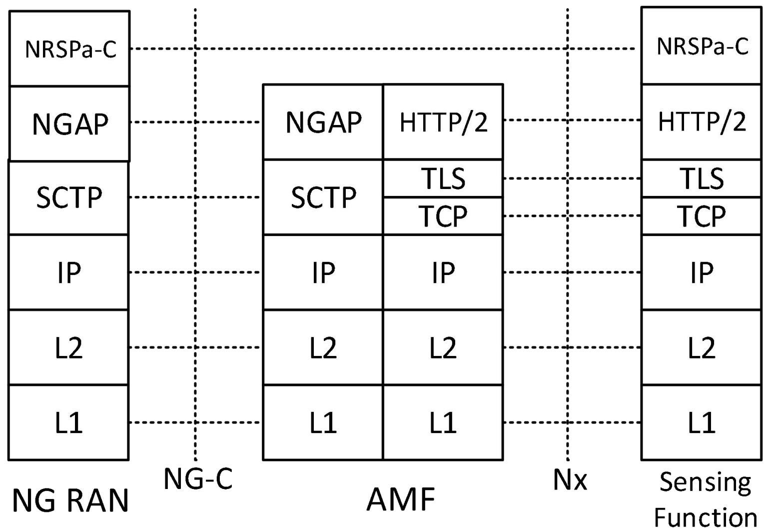

Figure 29 Integrated sensing and communication protocol stack based on control plane

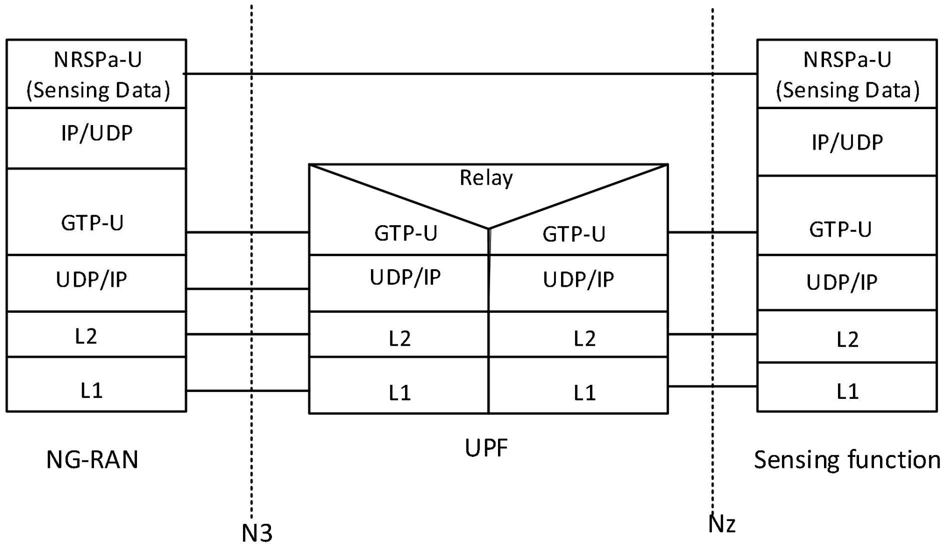

Figure 30 Integrated sensing and communication protocol stack based on user plane

5. Summary

The next-generation protocol stack 3.0 analyses and studies the next generation protocol stack from the perspective of servitization, componentization and intellectualization, based on the analysis of the current network shortcomings including the cloudification with CU/DU architecture, edge NWDAF, RAN slicing and network slicing. Besides, it elaborates the difficulty faced by the current network based on the scenarios and technologies focused by 3GPP, and raises the corresponding solutions and research directions.

The next-generation protocol stack 4.0 starts with elaborating the technical challenges brought by the various future scenarios including Space-Air-Ground-Sea integrated communication, holography, digital twin, new ultra-dense network, integrated sensing and communication, native determinacy, intelligent network, user native networking, multi-parties native networking, energy and cost prioritized networking, and multi-band aggregated networking. On this basis, it identifies the potential characteristics of the next-generation protocol stack and provides several possible next generation protocol architectures initially, such as Orchestratable, Stack-Free, Component-based packet forwarding protocol, cloud native service architecture, and generalized QoS framwork. Furthermore, 4.0 proposes the possible research ideas or solutions for potential key technologies such as Space-Air-Ground-Sea integrated communication, intelligence based RAN, adaptive protocol stack for new architecture to support RAN nodes cooperation, and integrated sensing and communication technology.

Besides the continuous evolution and enhancement to 3.0, the next-generation protocol stack 4.0 brings out potential protocol enhancements enabling new application scenarios, future protocol stack architectures and 6G key technologies. We hope 4.0 can provide some help for the friends interested in 6G and next-generation protocol stack architecture. Let’s work together and explore the better protocol stack architecture to better support the communication requirements raised by the vision of a better life in the future!

Reference

[1] 中国移动, 2030+愿景与需求白皮书.

[2] NGMN White Paper, 6G Drivers and Vision.

[3] 中国移动, 2030+技术趋势白皮书.

[4] IMT-2030, 6G 网络架构愿景与关键技术展望白皮书.

[5] IMT-2030, 6G 总体愿景与潜在关键技术白皮书

[6] 3GPP TS 36.300: " NR and NG-RAN Overall Description "

[7] 3GPP TS 38.300: " Evolved Universal Terrestrial Radio Access (E-UTRA)

[8] and Evolved Universal Terrestrial Radio Access Network "

[9] 3GPP TS 38.300: " System architecture for the 5G System (5GS) "

[10] TS 38.455, NG-RAN; NR Positioning Protocol A (NRPPa), 3GPP

[11] IEEE 802.11bf: Toward Ubiquitous Wi-Fi Sensing

[12] 《White paper on 6G vision and technology trends》, DT Mobile，2020.12

[13] WHITE PAPER ON 6G NETWORKING, 6G Flagship, University of Oulu

[14] New Services and Capabilities for Network 2030: Description, Technical Gap and Performance Target Analysis ITU-T FG NET-2030

[15] WHITE PAPER ON 6G NETWORKING, 6G Flagship, University of Oulu

[16] 3D Holographic Display and Its Data Transmission Requirement, X Xu, Y Pan, PPMY Lwin, X Liang, 10.1109/IPOC.2011.6122872，IEEEXplore

[17] Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions, Alexander Clemm, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck, IEEE Communications Magazine • January 2020

Abbreviation

- **5G** the Fifth-Generation Mobile Communications
- **6G** the Sixth-Generation Mobile Communications
- **AI** Artificial intelligence
- **BWP** Bandwidth Part
- **CU** Central Unit
- **GoS** Grade of servic
- **HTC** Holographic-type communication
- **IoT** Internet-of-things
- **IP** Internet Protocol
- **ISLs** Inter-satellite links
- **MAC** Medium Access Control
- **MCG** Master Cell Group
- **ML** Machine learning
- **NRSP** NR sensing protocol
- **NAS** Non-access layer
- **OSFC** Orchestratable Stack-Free Component-based packet forwarding protocol
- **PHY** Physical
- **PDCP** Packet Data Convergence Protocol
- **PDU** Protocol Data Unit
- **RB** Radio Bearer
- **RLC** Radio link control
- **SCG** Secondary Cell Group
- **PCell** Primary Cell
- **SDN** Soft Definition Network
- **UE** User Equipment
- **UDN** Ultra-Dense Network

Acknowledgement

Grateful thanks to the following contributors for their wonderful work on this whitepaper:

Editors:

China Mobile	Chih-Lin I, Junshuai Sun, Huimin Zhang

Contributors:

China Mobile	Na Li, Yingying Wang, Gang Li, Yun Zhao, Quan Zhao, Min Yan, Xuan Liu, Xin Sun, Siqi Chen, Zhou Tong, Zirui Wen
CICTM	Li Chen, Haichao Qin
ZTE	Feng Xie, Fei Wang, Li Yang, Yan Xue, Hanchao Liu
VIVO	Yannan Yuan, Dajie Jiang, Wei Bao
China Telecom	Jing Wang, Shuo Peng
OPPO	Cong Shi, Jiangsheng Fan
China Unicom	Shan Liu
Apple	Huarui Liang, Fangli Xu, Xiaoyu Qiao

FuTURE FORUM is committed to cutting edge technologies study and applications. Controversies on some technical road-maps and methodologies may arise from time to time. FuTURE FORUM encourages open discussion and exchange of ideas at all levels. The White Paper released by FuTURE FORUM represents the opinions which were agreed upon by all participating organizations and were supported by the majority of FuTURE FORUM members. The opinions contained in the White Paper does not necessarily represent a unanimous agreement of all FuTURE FORUM members. FuTURE FORUM welcomes all experts and scholars' active participation in follow-on working group meetings and workshops. we also highly appreciate your valuable contribution to the FuTURE White Paper series.