2020·广州
世界5G大会
WORLD 5G CONVENTION

INTEGRATION OF SENSING, COMMUNICATION AND COMPUTING TOWARD 6G

WHITE PAPER V7.0 C
2020.11

Executive Summary

6G is expected to be an end-to-end (E2E) information processing system, whose core function will be extended from information delivery to information acquisition, information delivery and information computing, to support the emerging services, such as AI services, immersive services and digital twin services. The native sensing, enhanced communication and native computing will be the features of 6G, and the integration of sensing, communication and computing is the development trend of 6G.

The integration of sensing, communication and computing is an E2E information processing framework, where information delivery will be performed along with information acquisition and information computing. It can be divided into function cooperation, function fusion and function reconfiguration.

In traditional information service architecture, terminal is responsible for sensing. Network is responsible for communication, and cloud is responsible for computing. The integration of sensing, communication and computing is addressed in terms of cooperation of cloud, network, edge and terminal. Actually, cloud services, edge computing, network slicing, non-public network and internet of vehicles (IoV) have introduced such cooperation capabilities. Furthermore, the concept of multifunctional radio has been proposed according to the integration of sensing, communication and computing, where the air interface extends its core function from wireless access and transmission to wireless sensing, positioning, synchronization, and computing.

In practical design, the following key issues should be addressed for the integration of sensing, communication and computing.

1. Technology for integration of wireless sensing and transmission

The natural idea for the integration of wireless sensing and transmission is to design a unified waveform to enable them to work in same spectrum bands. They can share the radio resources and hardware resources, and realize wireless sensing and transmission functions through interference cancellation, interference isolation and other technical means effectively.

2. Deterministic information processing

Deterministic information processing is an E2E task-centric information processing technology including deterministic sensing, deterministic transmission and deterministic computing. It is expected to ensure service experience from system level by jointly reducing sensing delay, transmission delay and computing delay, and jointly improving sensing reliability, transmission reliability and computing reliability. The key issue is to solve the problem of ultra-high precision synchronization and joint resource scheduling.

3. Technologies for the multi-dimensional resource management and scheduling

Sensing resources refer to the radio and hardware resources such as radar, camera, various sensors and detectors. Communication resources refer to the radio and hardware resources of access, forwarding, routing and synchronization. Computing resources refer to special or general software and hardware resources such as computing and storage. These resources inside or outside the network can form a resource pool through virtualization for multi-party sharing. This can improve resource utilization and provide new capabilities such as cooperative sensing, cooperative communication and cooperative computing. Block chain is a better tool for resource sharing. The key problem is how to perform modeling, optimization, decomposition and allocation for a sensing task, a communication task and a computing task according to the service requirements.

4. Technologies for service continuity

Service continuity refers to the functional continuity of sensing, communication, computing and application. Sensing continuity refers to the ability of continuous information collection, positioning and tracking for the sensing objective, which requires sufficient sensing coverage of base station or mobile terminal. Communication continuity traditionally refers to the handover of access points, which can be achieved by multiple-connection over high and low frequency bands and increasing base station density. Computing continuity refers to the ability to decompose and migrate computing tasks in real time due to the dynamic changes of computing resources. Application continuity refers to the ability of real-time migration from local platform to another platform due to the change of service environment.

5. Capability exposure

The capability exposure of sensing, communication and computing can further enhance the information service ability of 6G. The capability exposure framework architecture can be divided into resource layer, capability layer and application layer. The resource layer includes all kinds of sensing resources, communication resources, computing resources and various data resources. The capability layer is the core layer, being responsible for capability modeling and encapsulation, capability arrangement and operation and maintenance, and capability announcement. The application layer is the underlying capability demander, which requirements are proposed to the capability layer through API interface.

6. The challenges

The primary challenge for the integration of sensing, communication and computing is to answer the coupling relationship between them and the integrated function gain boundary. Secondly, the integration of wireless sensing and transmission in the millimeter wave, visible light and terahertz bands faces the challenge of signal processing and devices challenges. Thirdly, the integration of sensing, communication and computing is facing the challenge of heterogeneous computing environment and chips. Fourthly, resource scheduling and collaborative machine learning meet the challenges of algorithms. Finally, trusted information processing should be addressed to create a trustable sensing, communication and computing environments.

With the breakthrough of the above key issues, the integration of sensing, communication and computing is expected to be applied in the following 6G scenarios:

- Inter-AT interaction: providing information and data exchange and sharing, supporting cooperative machine learning, focusing on solving the problems of 3D perception, model training, reasoning and decision-making required by machine learning (ML). Inter-AT interaction is the fundamental capability for the unmanned services.

- Human-machine interaction: supporting interactive XR and holographic communication, focusing on solving the problems of 3D perception, modeling and display. Human-machine interaction is the fundamental capability for the immersive services.
- Virtual-physical space interaction: supporting the key information exchange between the virtual space and physical space, focusing on solving the problems of multi-dimensional perception, modeling, reasoning and control. Virtual-physical space interaction is the fundamental capability for the digital twin services.

Content

1. Introduction

With the accelerated development of information technology, sensing, communication and computing have become the three essential functions to provide mobile services. The emerging services, such as artificial intelligence (AI) services, unmanned services, digital twin services and immersive services [1], are showing stronger coupled characteristics of sensing, communication and computing, which then bring the technical requirements on functional integration.

This paper attempts to find how sensing, communication and computing work integratively to bring 6G new capabilities and services.

2. Information Technology Trends

2.1 Information technology trends

The key driving force for the development of information technology is to deeply understand and transform the physical world. The existing mainstream commercial technologies, such as 5G, Internet of Things, big data, cloud computing, are promoting the comprehensive upgrading of traditional industries and the digital development of social and economic. A large number of technology trend analysis results [2] show that digital virtualization and intelligence have become the emerging information technology trends. Such trends are extending the human society from the physical world to the virtual world, and will spawn a large number of autonomous things (ATs) and virtual spaces.

The AT refers to an entity that can carry out human-like activities through the ability of environment perception and interaction, such as robot, unmanned Aerial Vehicle (UAV), autonomous vehicle, virtual assistant, etc. Virtual space is a digital reconstruction, simulation or image of physical space from space-time dimension, such as virtual/augmented/mixed reality (XR), holographic image, digital twin, intelligent space, etc. The interaction and cooperation among ATs/virtual spaces defines the mobile service trends.

From 4G to 5G, the mobile communication technology features can be identified from the field of communication technology (CT) to the field of information and communication technology (ICT) by introducing cloud/edge computing. 5G design fully demonstrates the technical characteristics of ICT integration. With the accelerated ICT convergence and AI penetration, the mobile communication network develops towards the integration of IT, CT and big data technologies (ICDT), as shown in Fig.1. Then, the primary expected 6G feature is native AI. Similar to AI technology, digital twin technology has shown stronger vertical application ability in both the expected 6G capabilities and services.

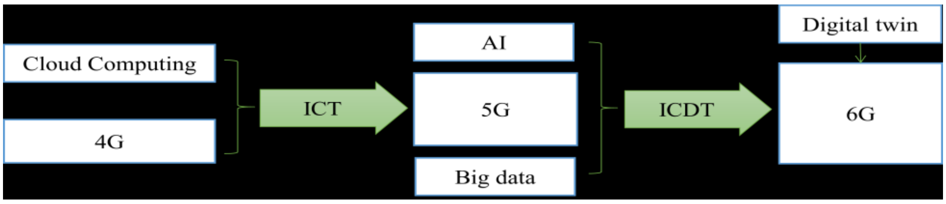

Fig.1: The information technology trends toward 6G

2.2 Requirements on information processing

The fundamental function of a mobile communication network is to deliver information. However, the upcoming explosion of interaction and cooperation among ATs/virtual spaces will bring about a clear shift in a mobile network from human-centric interconnection to AT-centric interconnection. Due to AI, sensing has become a fundamental interconnection means, as important as communication. Such kind of interconnection requires a mobile network to support more functions in terms of sensing, communication, computing and ML. This means a mobile network should support an E2E information processing, rather than information delivery. An E2E information processing covers information acquisition, delivery, computing, storage, utilization, and security, and involves cloud, network and terminals.

In fact, the 5G NR performance is continuously improved in terms of higher reliability, lower latency, higher data rate, deterministic transmission, and high-accuracy positioning. However, the NR framework does not support native sensing, native ML and native computing due to the functional limitation. Recently, preliminary viewpoints and thoughts on 6G possible architecture and potential technologies have been reported. However, there is little concern on what function 6G radio actually needs.

Integration framework of sensing, communication, and computing is believed to introduce the essential functions for E2E information processing for 6G. This framework should work under both function cooperation and function convergence model, support cloud-network-terminal cooperation and convergence, and bring the basic radio capabilities of cooperative sensing, cooperative ML, deterministic transmission with higher reliability and data rate to meet the tighter requirements of the future services.

3. Integration of sensing, communication and computing

3.1 Essential elements of information processing

3.1.1 Information acquisition: Sensing

Information acquisition is an essential prerequisite to E2E information processing, for the purpose of deeply understanding and transforming the physical world. Generally, it refers to the behavior of information collection from sensing objects by using sensing devices, data interface, and/or the ability of data analysis.

Information acquisition usually works in three typical scenarios. First, it is a component of information service framework to provide service content (such as IoT services). Second, it is an underlying function to support the information delivery and information computing, especially to provide the operation parameters. Third, it is one of key enablers for AI to provide big data and data sets for model training and reasoning.

Information acquisition mainly collects the service attribute and state, user attribute and state, network attribute and state, terminal attribute and state, environment attribute and state.

The service attribute and status includes service delay, service reliability, service transmission rate, service throughput, service application ID, service type, IP address and port of transmission data, space-time distribution of service, time to establish service, and performance requirements, etc.

The user attribute and status includes user personal information and user behavior information, etc.

The network attribute and status includes throughput, spectrum efficiency, energy efficiency, network average delay, network capacity, network reliability, transmission rate, uplink transmission rate, downlink transmission rate, computing power of network, network coverage, wireless resource allocation, cache resource usage, cell handover success rate and network interference relationship, etc.

The terminal attribute and status includes power consumption, location information, terminal's international mobile equipment identity, sending rate, receiving rate, signal-to-interference and noise ratio, bit error rate, battery capacity, number of terminals connected to network, number and type of applications installed on the terminal, etc.

The environment attribute and status includes indoor environment, outdoor environment, high-speed environment, climate status, distribution status of trees and buildings, etc.

Sensing in context of 6G has multi-dimensional, multi-level and distributed characteristics.

3.1.2 Information delivery: Communication

Communication refers to the information delivery behavior through certain methods and media. Information delivery is the fundamental function of a mobile communication network, which has hardly changed from 2G to 5G.

A mobile communication network is a typical system composed of several network nodes with a certain network topology. Its development is a process of continuously expanding media (new spectrum resources and optical fiber resources), updating and upgrading transmission and access methods, reconstructing network structure and node functions, to improve communication ability in mobile environment. Network nodes are the basic elements for communication, which have independent functions, such as access function, transmission function, forwarding function, control function and computing function.

3.1.3 Information computing

In a broad sense, information computing refers to the calculation behavior for a specific task through special or general hardware and software. In addition to the computing work related to information acquisition and delivery, information computing mainly includes application layer processing, data storage, big data analysis, ML with modeling, reasoning and decision-making, etc.

Cloud computing, edge computing, terminal computing, and fog computing are common computing paradigms. Distributed cooperative computing is a key technology for the E2E information processing with the challenge of heterogeneous computing.

3.2 Basic concept and category

3.2.1 Definition

The integration of sensing, communication and computing is information processing framework to implement information acquisition, information delivery and information computing simultaneously in the E2E information processing. It includes the sensing behavior and corresponding resources to enhance the performance of communication and computing, the communication behavior and corresponding resources to enhance the performance of sensing and computing, and the computing behavior and corresponding resources to enhance the performance of sensing and communication.

The integration of sensing, communication and computing includes three modes: function cooperation, function convergence and function reconfiguration

3.2.2 Function collaboration

Function collaboration refers to the working mode, in which sensing, communication and computing functional entities exchange or share information through interfaces in order to enhance their performance. It also includes collaborative computing, collaborative sensing, collaborative positioning and cooperative communication between different entities.

3.2.3 Function convergence

Function convergence refers to the integration of sensing, communication and computing functions in the same process through unified signal and hardware design.

3.2.4 Function reconfiguration

Function reconfiguration refers to the function virtualization of sensing, communication and computing supported by software-defined hardware architecture. It is the highest level of integration of sensing, communication and computing, where the hardware resources are interconnected to form a relatively fixed functional path, so that the objective function can be carried out in a way close to "dedicated circuit". It is expected to be realized by software-defined chip with both software flexibility and hardware efficiency.

3.3 Integration of cloud, network, edge and terminal

For the purpose of E2E information processing, integration of sensing, communication and computing implies the integration of cloud, network, edge and terminal. In fact, to realize the 5G features, the integration of cloud, network, edge and terminal has been partially addressed. For example, E2E network slicing and cloud applications need the cooperation between cloud and network. Edge service continuity needs cooperation of edge and terminal. 5G non-public network needs cooperation of cloud, network, edge and terminal. IoV needs the cooperation of cloud, network, edge, terminal and even road. In practical design, such cooperation can be decomposed into multi-dimensional cooperation, such as multi-access cooperation, resource scheduling cooperation, data cooperation, service traffic cooperation, capability exposure cooperation, security cooperation, orchestration cooperation, etc.

3.3.1 Integration of cloud and network

Integration of cloud and network refers to the cooperation mechanism where network is able to acquire the service attribute and status to optimize resource allocation, while application layer is able to learn about network attribute and status to optimize service process. It embeds computing into the network and selects the optimal computing node in context of network to ensure service QoS, such as cloud XR, cloud games and cloud video monitoring.

3.3.2 Integration of cloud, network and edge

Edge computing expands the ability and scope of cloud computing. However, the efficient collaboration of cloud/edge computing requires the network to be flexible and reconfigurable. In order to solve the coordination of cloud computing and edge computing, better meet the requirements of application scenarios, it’s necessary to realize the integration of cloud, network and edge.

Cloud-network-edge collaboration, which refers to the deep integration of cloud computing, edge computing and other network architectures, can provide integrated cloud-network-edge services for users, whose architecture is shown in Fig. 2. Among them, three-dimensional network slicing technology can support the deep decoupling and flexible reconfiguration of the network, serve the efficient collaboration of cloud/ edge computing, and provide diversified services for users. And cloud computing schedules network resources according to the requirements of applications. Edge computing provides computing services on the edge of the network and relieves the pressure of cloud computing and network bandwidth costs.

Relying on software-defined networking (SDN), network function virtualization (NFV), E2E network slicing and AI, cloud-network-edge collaboration provides an E2E, flexible and extensible solution for cloud-edge collaboration to achieve flexible deployment of their facilities on demand while improving resource utilization and business reliability. It also provides three-dimensional decoupling of resource, network function and virtual network based on virtualization technology. The SDN control plane and virtual machine manager (Hypervisor) deployed in domains and hierarchies can fulfill the orchestration and provide fine control granularity, and improve the resource utilization and maintenance efficiency of the system.

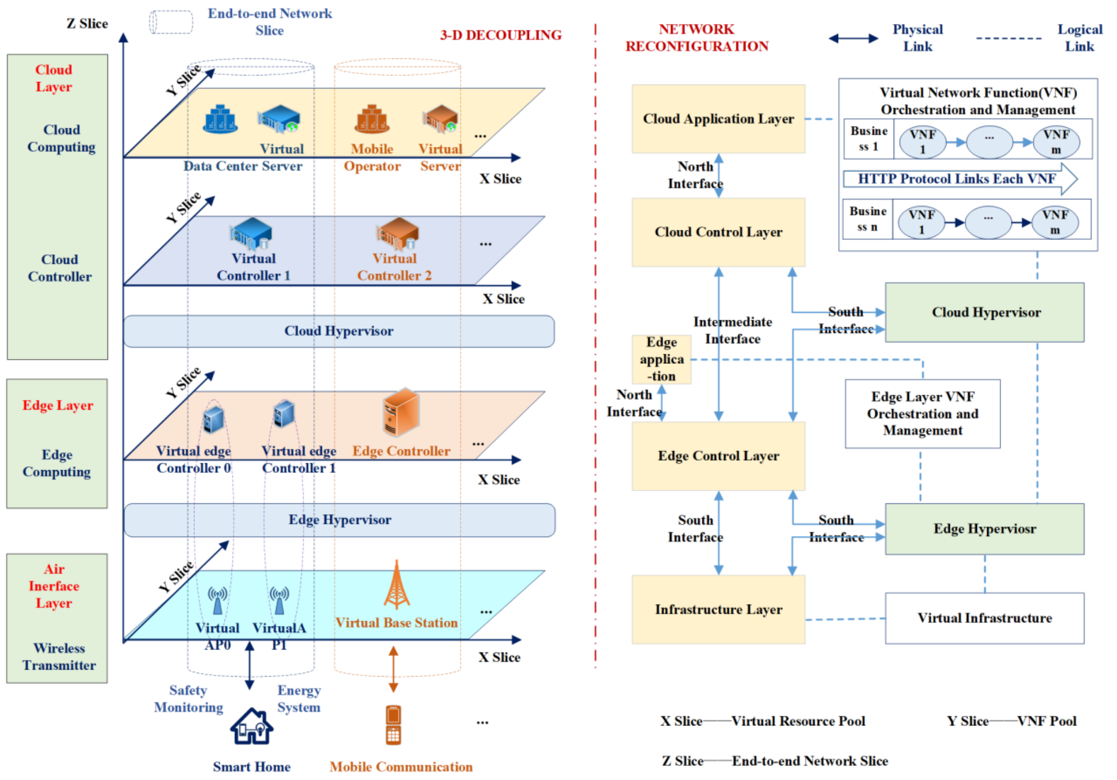

Fig. 2: Integration architecture of cloud-network-edge

Cloud-network-edge collaboration includes the docking of edge computing infrastructure and cloud infrastructure to realize the virtualized collaboration of network physical resources. It includes the docking of edge computing platform and cloud platform to realize network data collaboration, intelligent collaboration, application orchestration collaboration and traffic management collaboration. It also includes the docking of edge computing software and software running on the cloud platform to realize the collaboration of services in the network. In this way, it can provide optimal resource scheduling strategies according to the traffic requirements, to obtain a flexible and scalable dynamic resource management solution. Furthermore, cloud-network-edge collaboration can enhance the openness and security of services by means of collaboration and sharing, and it can also realize convenient and safe connections among cloud, network and edge, and solve the problems of low resource utilization of multiple service sources and low service access efficiency under large amounts of data. Cloud-network-edge collaboration can finally achieve multiple-cloud server, edge server, multiple-user and multiple-terminal in multiple network slicing optimized collaboration and efficient connection in the network.

3.3.3 Integration of cloud, edge and terminal

Integration of cloud, edge and terminal mainly refers to the distributed cooperative computing among cloud, edge and terminal. The interaction and cooperation among ATs/virtual spaces require multifold computing power, and the computing task should be decomposed and assigned among the cloud, edge and terminal.

3.4 AI/ML Splitting between AI/ML endpoints

3.4.1 Requirements for AI/ML Splitting

AI/ML Splitting is expected to enable the AI/ML applications with conflicting requirements which are computation-intensive, energy-intensive, privacy-sensitive and delay-sensitive.

The scheme of split AI/ML inference is depicted in Fig.3. The AI/ML operation/model is split into multiple parts according to the current task and environment. The intention is to offload the computation-intensive and energy-intensive parts to network endpoints, whereas leave the privacy-sensitive and delay-sensitive parts at the end device. The device executes the operation/model up to a specific part/layer and sends the intermediate data to the network endpoint. The network endpoint executes the remaining parts/layers and feeds the inference results back to the device.

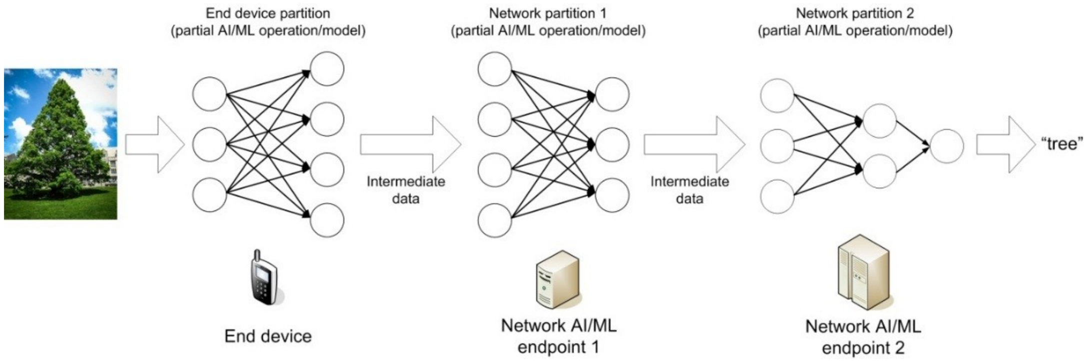

Fig.3: Split AI/ML inference scheme

3.4.2 Basic mode for AI/ML Splitting

The modes for AI/ML operations over device and network are illustrated in Fig.4.

The modes are in general applicable for AI/ML training as well as inference. In this section, we focus on the inference processing. Mode a) and b) are traditional schemes operating the AI/ML inference wholly on one endpoint. Mode c) - g) attempt to split the AI/ML inference or even the model into multiple parts according to the current task and environment, to alleviate the pressure of computation, memory/storage, power and required data rate on both device and NW endpoints, as well as to obtain a better model inference performance on latency, accuracy and privacy protection.

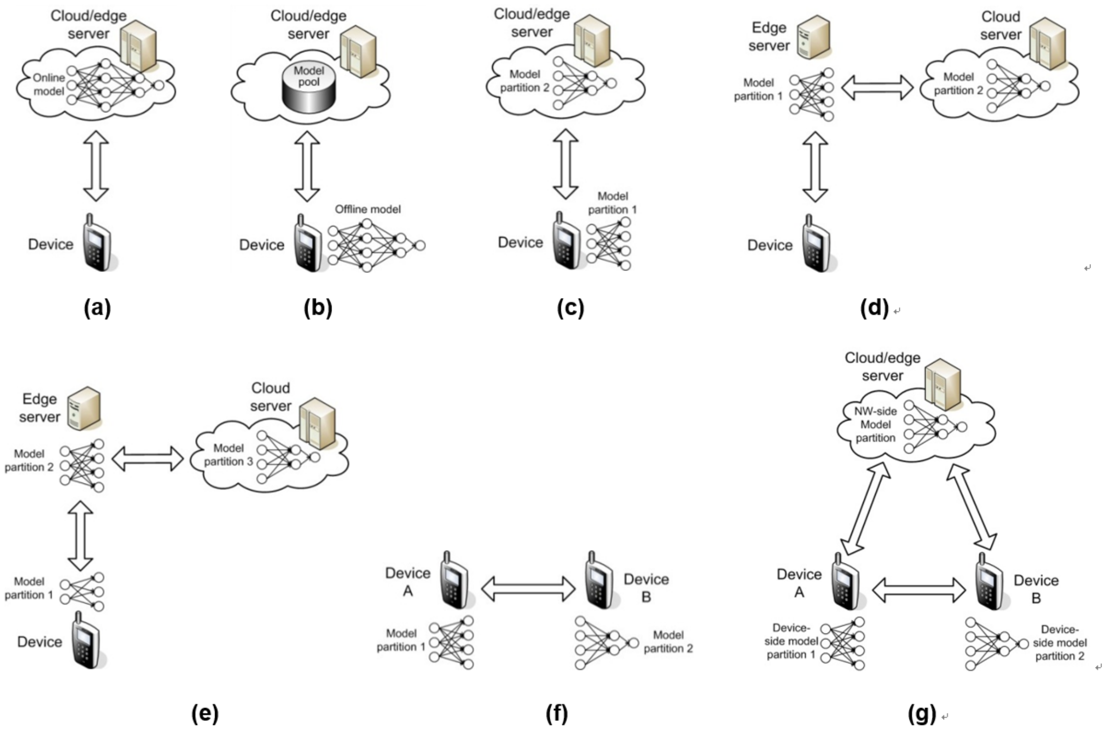

Fig.4: Basic modes for AI/ML Splitting

In mode a), the AI/ML model inference is only carried out in a cloud or edge server. The device only reports the sensing/perception data to the server, and does not need to support AI/ML inference operations. In mode b), the AI/ML model inference is performing locally at the mobile device. The device does not need to communicate with the server and can preserve the privacy at the data source. In mode c), an AI/ML inference operation or model is firstly split into two parts between the device and the cloud/edge server according to the current system environmental factors such as communications data rate, device resource, and server workload. Mode d) can be regarded as an extension of Mode a). The difference is that the DNN model is executed through edge-cloud synergy, rather than executed only on either cloud or edge server. Mode e) is the combination of Mode c) and d). An AI/ML inference operation or an AI/ML model is split over the mobile device, the edge server and the cloud server. Mode f) provides a de-centralized split inference. An AI/ML inference operation or model can be split over different mobile devices. Mode g) is device-device-cloud/edge split inference.

3.4.3 Performance indicators for AI/ML Splitting mode

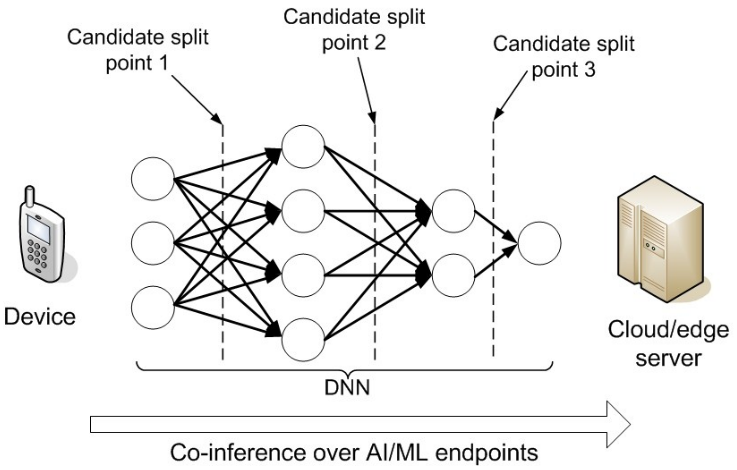

Fig.5: Modes for DNN model split

The key of split AI/ML operation is to select the optimum split mode and the split point to guarantee that the required resource is below the upper limit of the available resources on the mobile devices. And optimize the consumption of computation, storage/memory, power, communication resources on device as well as network side. As shown in Fig.5, a DNN model can be split at different points, resulting indifferent AI/ML performance. Different split points should be adopted for different AI/ML performance target. Some of the performance requirements for DNN splitting inference are listed below.

- Inference accuracy: Inference accuracy refers to the ratio of the number of the input samples that get the correct inference results to the total number of input samples, reflecting the performance of the AI/ML inference task.
- Inference latency: Inference accuracy refers to the time spent in the whole inference process, including the model inference in each involved endpoint

and data transmission.

- Inference power consumption: power consumption is affected by the size of the DNN model. Although split inference may introduce the communication power consumption, the overall power consumption is expected to be reduced.
- Memory footprint: Memory footprint is mainly impacted by the size of the DNN model and the way of loading the tremendous DNN parameters. Split inference enables the device to only load a part of the DNN model in memory, thus to reduce the memory footprint.
- Privacy: Split inference is an efficient way to preserve the privacy at the device meanwhile offloading the most of computation to network. The intermediate data brings a much smaller risk of disclosing the privacy than the raw data.

3.5 Multifunctional radio

In cooperative sensing and ML, a large amount of data and information need real-time or non-real-time transmission and computing. The 6G radio should support the following essential functions:

- Stronger wireless transmission: Support enhanced eMBB and URLLC on uplink, downlink and sidelink with the sensing ability and potential technologies.
- Deterministic transmission: Support integration of synchronization network and RAN, and coexistence with other TSNs. Support high-precision timing and positioning message delivery. Support integrated traffic and computing power scheduling.
- Cooperative computing: The computing resources at edge and terminal should be easily perceived and scheduled, even in the heterogeneous computing case.
- Multidimensional sensing and 3D positioning: Support multi-sensor operation and cooperative positioning with the help of wireless transmission. Support camera, Lidar, mmWave Radar, ultrasonic, and other sensors. Support wireless positioning, GNSS, and local positioning.

3.5.1 Multifunctional radio architecture

To address the challenge, we propose the multifunctional radio as a solution for 6G radio, as shown Fig.6. It is a conceptual framework aiming to expand radio function from wireless transmission to computing, sensing and positioning. It is expected to achieve the following objectives:

- Enhance radio resource management. The RAN state space and resource allocation space will be redefined by the perception and prediction ability of sensors. This benefits network capacity improvement.
- Reduce E2E transmission latency. The proposed radio covers HMI and MMI, as well as the traditional air interface (up/ down/ sidelink). This will reduce the latency interval between information acquisition and transmission. The transmission latency and computing time will be further reduced when data dimensionality reduction is performed over the radio.

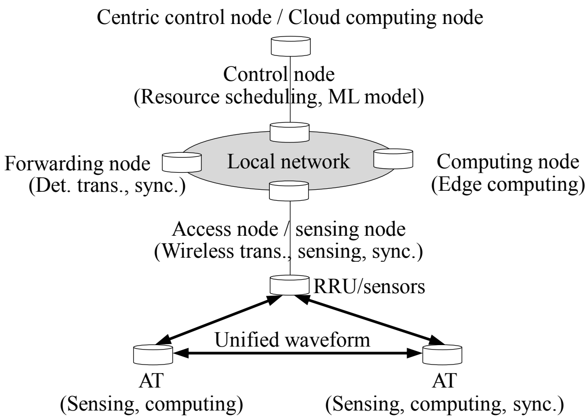

Fig.6: Multifunctional radio framework

From a functional point of view, 5G NR have four kinds of functions (nodes): access function, transport function, control function and computing function. The corresponding nodes form RAN, transport network and core network, respectively.

The proposed multifunctional radio framework is shown in Fig.6, where new functional nodes are introduced, namely sensing nodes, synchronization nodes, and computing nodes. These new nodes form synchronization network, deterministic transport network, and computing-aware network, respectively and then form a multifunctional RAN, rather than the traditional RAN supporting only wireless transmission

The sensing nodes support multidimensional sensing with multi-sensor operation, such as camera, ultrasonic, Lidar, Radar, other sensors, and other wireless detection devices. These sensing nodes also support positioning, such as GNSS, wireless positioning and cooperative positioning with ATs. They can be deployed independently, or be located with the access nodes. An access node equipped with both RRU and sensors can be considered as a super AT.

Similarly, a synchronization node may be an access node or a forwarding node. It delivers the crucial timing messages through the local network to enable local deterministic transmission. One of ATs associated with one task should be selected to receive the timing messages and forward them to other ATs.

The proposed framework adopts two-level multi-domain distributed architecture for the control function, where the centric control node and local ones cooperatively manage information acquisition, delivery and computing. These control nodes will take part in cooperative sensing and ML to do computing task scheduling. Such operation is defined as cloud-network-terminal cooperation. For example, the model training task will be assigned to cloud and edge, while the reasoning and decision making tasks may be handled at ATs.

Moreover, the proposed radio framework extends from air interface to MMI and even HMI. The design of unified waveform is needed to support the convergence of wireless communication and sensing, not only for uplink/downlink, but also for sidelink. This can help reduce the E2E latency, improve the spectrum utilization, enhance radio resource management, improve network capacity and simplify the hardware architecture.

3.5.2 Function separation architecture

With the rapidly development of the mobile communication technology, the space-time span of wireless network in the future is becoming larger and larger, and the number of users and network traffic are increasing exponentially. Besides, the business types tend to be diversified and emerging application scenarios are constantly emerging, which requires higher coverage and energy efficiency.

To meet the requirements of wireless communication network in the future, a new network perception, communication and computing separation architecture emerges. The architecture separates the perception from the general control signaling transmission on the basis of the traditional control and user plane separation network architecture to realize the separation of perception, communication and computing. The traditional base stations are divided into three groups: low frequency perception base stations (PBS), medium frequency control base stations (CBS) and high frequency traffic base stations (TBS). They work in different frequency bands and perform different tasks.

PBSs broadcast in low frequency band to sense the location of users/air base stations (UAV and airship), wireless environment, available resources of edge/cloud, and so on. By using intelligent beam, dynamic cooperative relay and other methods, the transmission energy efficiency of TBSs is improved, and the flexible extension of wireless coverage is realized, so that CBSs and TBSs can provide better control and data services for users.

CBSs work in the medium frequency band, and send low rate control coverage signals to users to ensure the users' access to the network. According to the users' service requesting status, TBSs are selectively turned on and off, and TBSs with the best performance is selected to provide services for users. The coverage of CBSs is large.

TBSs send high-frequency and high-speed service signals to users to provide data services. According to the requirements of different application scenarios, TBSs can provide differentiated services for users, such as high-throughput requirement, delay-sensitive requirement, and compute-intensive requirement. The coverage of TBSs is small.

In addition, through the dynamic deployment of unmanned devices, the traditional coverage is dynamically extended to realize the joint optimization of perception, communication and computing, and the blind spot coverage problem and mobile coverage problem of land-based network can also be solved.

Perception, communication and computing separation architecture is a multiple domain and full band system design and architecture with better energy efficiency and better delay. It comprehensively utilizes the different coverage characteristics of high, medium and low frequency band signals, adapts to the differentiated requirements of wireless coverage, and greatly improves the coverage ability of the network. The base stations working in different frequency bands perform their respective duties without interference with each other, which greatly reduces the network energy consumption and improves the network performance and transmission rate.

4. Key Technologies and Challenges

4.1 Unified design for wireless sensing and transmission

The integration of wireless sensing and transmission requires not only new system architecture and protocols, but also the development of devices and the support of key technologies, which are listed in detail as follows:

- Unified waveform design
- Communication demodulation algorithm and sensing algorithm
- Joint signal detection and interference cancellation
- Antenna array and precoding/ beamforming scheme
- Channel measurement and modeling for wireless signal and sensing signal
- Hardware architecture and scheme for device sharing

The key issue of unified waveform design is to minimize the interference between communication signal and sensing signal to guarantee communication and sensing functions, and improve spectrum efficiency on the premise of ensuring system performance. The unified waveform can be multiplexed with orthogonal resources, including time division multiplexing, frequency division multiplexing, space division multiplexing, code division multiplexing, and can also be designed with a fusion form. The latter design should find the performance balance point between sensing and communication. It is necessary to consider which function is primary.

Generally, the unified waveform is mainly divided into single-carrier waveform and multi-carrier waveform. Single-carrier waveform is usually combined with spread spectrum technology, such as direct sequence spread spectrum (DSSS) and chirped spread spectrum (CSS). DSSS is widely used in communication systems to obtain higher security and robustness, and provide stronger anti-jamming and anti-interception capabilities. It can also realize multi-user access. Some problems on single-carrier waveform design based on DSSS should be addressed [2]:

- Keep low sidelobe level when pulse shaping
- Select the sequences with good autocorrelation/ cross-correlation characteristics (e.g. m sequence, Gold sequence, Kasami sequence, Golay complementary sequence) and appropriate spread spectrum factor to meet the sensing accuracy and dynamic work range
- Design data frame structure reasonably to limit the length of correlation sequence and reduce the Doppler influence in case of the sensing system design based on echo signal

CSS technology is based on chirp signal, which is more used in radar sensing applications to help improve the balance between resolution and maximum search range. CSS signal is used to realize the orthogonality of sensing signal and communication signal. For example, chirp signal with the same carrier frequency but opposite polarity frequency modulation slope is used to represent communication signal and sensing signal respectively [3]. Besides chirp signal with opposite polarity frequency modulation slope, unified waveform can be designed by using chirp signal with different initial frequency of the same frequency modulation slope. There are also schemes based on the communication frame structure to realize the sensing function. Such schemes reuse the training sequence of the communication signal such as pilot to complete the sensing, without the spread spectrum method [4].

A typical multi-carrier waveform is Orthogonal Frequency Division Multiplexing (OFDM) waveform, which has many advantages over single-carrier spread spectrum waveform, in terms of higher spectrum efficiency, flexible bandwidth resource allocation, no range-Doppler coupling effect, etc. To ensure sensing capability, the frame structure should be designed with careful setting of subcarrier spacing and cyclic prefix length. In addition, some new waveform technologies, such as Orthogonal Time Frequency Space (OTFS), can be combined with OFDM for unified waveform design.

In addition to waveform design, joint detection of communication signal and sensing signal is also a key issue for the unified system design. Due to the random access and different resource division, the communication signal presents fragmentation and discontinuity in time, frequency and space, which increases the difficulty of sensing signal detection [5]. For communication systems, typical channel estimation algorithms, such as LS, MMSE, only estimate composite channels with limited unknown parameters, which are usually combined with interpolation scheme, so the accuracy is limited. However, the sensing system, especially radar, usually uses optimized or unmodulated transmitting signals to obtain the specific channel state information in terms of amplitude, angle, Doppler, delay, etc. Joint detection scheme needs to choose appropriate algorithms to guarantee the function division of communication and sensing and reduce the mutual influence. The conventional detection algorithms such as sequence correlation, matched filtering (MF), MUSIC, ESPRIT can be considered.

For the integrated system architecture, better joint detection algorithm is further needed. In the practical channel, multipath helps to improve the freedom of communication, but brings about clutter interference for the echo-based sensing system, which needs to be suppressed at the input of the sensing detection module. Moreover, the mutual interference between the communication signal and the sensing signal needs to be addressed. Serial interference cancellation (SIC) is commonly used, where echo signal in the received signal will be first cancelled according to the prior information, and then the communication signal is detected (demodulated and decoded), reconstructed and eliminated from the original received signal to obtain the sensing signal, which is detected finally [6]. When detecting the sensing signal, the problem of perception ambiguity caused by time non-synchronization problem is necessary to consider.

4.2 Cooperation of wireless sensing and transmission

Electromagnetic waves, including light waves, can be used not only to transmit information, but also to sense and image. Compared with electromagnetic wave-based approach, visible light-based imaging technology appeared even earlier (Niepce in France made the world's first photo on a photosensitive material in 1822). Based on visible light, holographic imaging is available now. In addition to visible light waves, infrared waves and terahertz waves, etc. can also be used for imaging. Besides that, some of the features of the target object can be sensed by using the electromagnetic wave as well. Thus, radar, remote sensing and other technologies have been developed.

6G will continue to develop towards even higher frequency ranges. In the research of 6G, THz band and even visible light band have recently attracted great attention.

In the 6G, communication and connectivity are no longer the only missions. Sensing based on electromagnetic waves, also known as wireless sensing, can be used to acquire some target object’s information, such as location, shape, size, posture, color, material, environment information, etc. Such valuable information can be acquired by continuous sensing and be utilized by upper layer applications or wireless communication itself.

The application scenarios of convergent communication and sensing can be divided into three categories: transparent transmission scenario, open-capability scenario and air interface fusion scenario.

Transparent transmission scenario: mobile communication system only provides channel for sensing data transmission.

Open-capability scenario: the communication system provides an open interface for the applications, and thus the sensing capabilities of the system can be used for the applications to provide some new network services.

Air interface fusion scenario: deeper fusion of wireless sensing and wireless communication will be carried out on the air interface. The new capabilities after fusion can be provided to the communication system itself or to the applications by the capability opening. The air interface fusion scenario can be further divided into passive sensing, active sensing, cooperative sensing, and alternative communication.

Passive sensing: the perceiver, network nodes or terminals, receives the electromagnetic wave (such as infrared) emitted or reflected by the target object.

Active sensing: the perceiver, network nodes or terminals, sends electromagnetic wave, and receives the echo for sensing after being reflected by the target object.

Interactive sensing: the perceiver (network nodes or terminals) interact with the target object (network nodes or terminals) after agreement on the subject, time, frequency, format and other parameters of electromagnetic wave transmission. And the perceiver senses the received electromagnetic wave.

Alternative communication: by wireless sensing, the perceiver (network nodes or terminals) uses image acquisition and recognition technology to acquire the original image or its corresponding information instead of transmitting the information via wireless network. In some certain scenarios, it's expected that the demand for air interface communication can be reduced.

4.2.1 Wireless communication assisted sensing

Sensing capability can be enhanced with the help of wireless communication system, especially the communication mechanism.

There are multiple approaches to improve sensing performance. In terms of the overall system architecture, the network can sense through terminals and sensors, or directly through network nodes such as base stations, where the communication signal can be used for sensing. The communication information carried by the communication signal can help devices to have a preliminary understanding of the surrounding environment. Therefore, the device can lock the target more quickly and complete the sensing of its surroundings in a shorter time. Further , through wireless communication, the sensing data can be quickly aggregated, and the instructions formed after sensing can be quickly distributed, so as to better serve the future applications with strict requirements on latency or real-time data. The lower the fusion level (such as the physical layer), the lower the expected latency. Specifically, with the assistance of wireless communication, the sensing performance improves in the sensing range, sensing accuracy and sensing speed.

In terms of improving the sensing range, the sensing range of the network can be improved by sharing the sensing information among nodes. Wireless communication can help network nodes share sensing information.

In terms of improving the sensing accuracy, the collaborative sensing of the same region can improve the sensing accuracy of the network. The communication information transmitted by communication signals is one of the prerequisites for realizing collaborative sensing.

In terms of improving sensing speed, the target location information carried by communication signals can reduce the time to lock the target, thus speeding up sensing.

Collaborative sensing is one of the key techniques to expand the range of sensing and improve the sensing accuracy. Since the sensing environment is rapidly changing and the sensing information has a certain timeliness, it is necessary to complete the cooperative sensing in a short time. Rapid diffusion and efficient fusion of sensing information are the prerequisites for realizing rapid collaborative sensing. Therefore, the main challenge in improving network sensing performance lies in how to realize the rapid diffusion and efficient fusion of sensing information.

However, superabundant sensing information entering the network will occupy excessive network resources, resulting in a large number of redundant sensing information, and thus affecting other services of the network. Therefore, how to reasonably allocate network resources for collaborative sensing is also a key problem to be solved.

4.2.2 Sensing assisted wireless transmission

Sensing assists wireless transmission in multiple approaches. On the one hand, the transmission parameters such as network beam management and resource allocation can be optimized according to the sensing information of propagation environment, terminal orientation, mobility trajectory and attitude, etc., so as to improve the performance of wireless transmission in terms of network capacity and network speed. On the other hand, the wireless signal based on the integrated waveform design can sense the physical space of communication objects and realize wireless transmission at the same time. Wireless communication with known physical spatial information of communication objects has higher security. Overall, with the assistance of sensing information, wireless transmission performance can be improved in terms of network capacity, networking speed and communication security.

In terms of improving network capacity, the physical environment information obtained by sensing can assist the network to optimize beam management parameters, and the electromagnetic environment information obtained by sensing can assist the network to optimize resource management parameters such as spectrum, power, etc. Then, the probability of data packet collision and congestion will be reduced, so as to improve network capacity.

In terms of improving networking speed, the number and location information of surrounding nodes obtained by sensing can accelerate the convergence speed of neighbor discovery and multiple access schemes, shorten the neighbor discovery and multiple access period, and further improve networking speed.

In terms of improving communication security, the integrated waveform design technique can make the nodes to know the physical space information of the communication object while communicating, which can help to identify the malicious nodes. The physical space and digital space information of the communication object can be obtained by sensing and communication, separately. With the assistance of the integrated waveform design, the functions of perception and communication can be realized simultaneously, which is helpful to realize the coupling of physical space and digital space information of communication objects, so as to improve communication security.

Based on sensing information, nodes can obtain the position, speed, trajectory and other information of surrounding nodes, which can help beam alignment and tracking, which result in more efficient neighbor discovery, MAC and routing design algorithms to realize fast networking. However, the reliability of the sensing information cannot be fully guaranteed, the improved algorithm proposed above needs to be compatible with the traditional algorithm to adapt to the situation without the assistance of sensing information.

In addition, through the sensing of location, speed and other physical information of the surrounding nodes, the received message source and characteristics can be estimated, which can be used to determine whether the message comes from a specific node, thus enhancing the information security. In order to improve communication security, how to realize the efficient coupling of physical space information and digital space information is a key problem to be solved. Moreover, unifying the security decision criteria is also essential.

4.3 Deterministic information processing

Deterministic information processing has been discussed preliminarily in 3GPP in terms of deterministic transmission, which focuses on latency and jitter control. To meet the E2E latency requirements, there has been some research on deterministic sensing, deterministic computing, deterministic positioning, and deterministic wireless transmission. Deterministic information processing implies the network transformation from resource-based to service-based.

The features of deterministic sensing, deterministic transmission and deterministic computing are respectively field-level processing, E2E processing and distributed processing.

4.3.1 Synchronization

Clock and time synchronization are the basis on which deterministic information processing can be achieved. 3GPP has discussed the clock/time synchronization for the integration of 5G network and TSN. 5G NR supports air interface capabilities such as ultra short frame structure, large-scale antenna, carrier aggregation, high-precision positioning and massive terminals. It requires that the inter-BS clock synchronization accuracy should be up to 100ns. The dual backup synchronization mechanism is commonly used, which considers both the satellite reference source and dedicated optical fiber network reference source of national high-precision ground-based time service system. The generalized precision time protocol (GPTP) is used in TSN, which should be developed towards 6G network.

In the integration scheme discussed in 3GPP, time synchronization is divided into two parts: the network-wide time domain and the local clock domain. The former is equipped with a master clock, which can theoretically be located in any device of the whole network, and the time synchronization information is delivered to the whole network through deterministic support function. The latter covers all local devices associated with the same task, such as robots, AGVs, sensors and controllers. One or more local devices are set as a UE to access 5G network and receive the time synchronization information. The local clock domains operate independently, and different synchronization accuracy can be set. A UE supports simultaneous connection of multiple local clock domains. Similarly, a base station also supports simultaneous connection of multiple local clock domains. When two or more local clock domains overlap, for example, when a mobile device enters from one clock domain to another, it needs to decide whether to merge multiple clock domains according to the synchronization accuracy requirements of the task process. If the accuracy is tighter, clock domain merging is needed.

In AT interaction scenarios, the coverage of local clock domains will be dynamically overlapped. Two kinds of clock domain merging methods can be considered. One is to merge multiple clock domains with reference to the master clock at the same time; the other is to select a device with UE function in the local clock domain to send synchronization information to the merging domain to realize multi-domain merging. The latter is suitable for the case of temporary interaction or low synchronization accuracy, such as the cooperation between UAVs or autonomous vehicles in the process of unmanned logistics.

4.3.2 Deterministic sensing

Sensing is a typical field-level information acquisition processing, no matter for the internal sensors and external sensors which are equipped within an AT. An unmanned service supported by AT interaction depends on these sensors with sensing on time and in time. Compared with the E2E information delivery and distributed information computing, the main performance of deterministic sensing is the sensing response time with a certain sensing accuracy and range. In cooperative sensing for cooperative ML, deterministic sensing also means the time synchronization among ATs and among the internal sensors.

4.3.3 Deterministic transmission

IEEE began to study TSN network since 2006, and has been expanding from audio and video field to industrial field, IoV field and MNO's transport network. The key issues for TSN are time synchronization, traffic scheduling, and network control and management. Technical standards have been released respectively on timing and synchronization, frame preemption, enhancements for scheduled traffic, cyclic queuing and forwarding, per-stream filtering and policing, stream reservation protocol enhancements and performance improvements, path control and reservation, seamless redundancy, and link aggregation.

3GPP R16 defined integration architecture of TSN and 5G network to provide key industrial IoT functionality. In case of unmanned manufacturing, the deterministic transmission between machine tools should meet the requirements in terms of latency of 100us~50ms, availability of 99.999% and synchronization accuracy of 1us.

DetNet is the generalized TSN applied into the IP and MPLS networks. It is composed of terminal node, edge node and forward node that jointly provide DetNet services. These nodes are interconnected through transport nodes (i.e., routers) as an endpoint, and connected to the subnets. These subnets, such as TSN and OTN, support DetNet flows. Multi-layer DetNet is possible, where one underlying DetNet appears as a subnet and provides services for higher-level DetNet. In DetNet, critical data streams with extraordinarily low packet loss ratios and E2E latency are guaranteed though resource reservation (node buffers, node schedulers, and link bandwidth) via configuration, management and protocol action. Critical data streams and other best-effort streams can co-exist within a single network if the former do not take up too much network resources.

TSN is also introduced into 5G network for the deterministic fronthual. Moreover, the work item “deterministic mobile networking” has been set up in ETSI to identify the requirements of 3GPP mobile network towards deterministic networking capabilities for URLLC applications. Its target is to provide holistic architecture including functions, interface and protocol to realize E2E deterministic transmission across RAN, core network, backhaul or transport and virtual networks within data center. Base on this, 6G is expected to be a deterministic network.

4.3.4 Deterministic computing

Deterministic computing refers to the information computing on time and in time within a task process. An integration of cloud, edge and network, as well as computing resource allocation algorithms are needed to ensure deterministic computing.

4.4 Resource management

6G needs a mechanism for multi-dimensional heterogeneous resource management and scheduling.

4.4.1 Sensing resource

The most important mean for sensing is to use sensing devices, i.e. sensors. The generalized sensors include Radars, camera, detection devices, positioning modules, and other kinds of sensors. In detail, Radar includes ultrasonic Radar, millimeter wave Radar, Lidar, etc. Detection device include X-ray detection device, infrared detection device, sonar device, terahertz detection device, radio detection device. Another kind of sensors includes pressure sensor, temperature sensor, humidity sensor, gas sensor, acoustic wave sensor, photosensitive sensor, etc. These devices with related spectrum form a sensing resource pool for sharing and scheduling.

4.4.2 Communication resource

Communication resources include routing resources and radio resources. The latter includes spectrum resources, power resource, antenna resources, slot resources, code resources, etc.

4.4.3 Computing resource

Computing resources will be the native resources of 6G and can be scheduled as a service. In order to meet the diverse computing requirements, this kind of resources is distributed among cloud, edge, and terminal, in terms of CPU, GPU, NPU, FPGA, ASIC, etc. Storage resource is also a kind of computing resource. For different applications, a variety of computing power combinations have been extended from single-core CPU to multi-core CPU, and to CPU+GPU+FPGA, etc. Faced with a variety of heterogeneous computing resources distributed in the network, a unified metrics system and an abstract representation method should be designed for computing resource management and scheduling.

In case of computing virtualization, the computing carrier has developed from a virtual machine to a container and unikernel. The mirror size is gradually reduced from the GB level of the VM to the MB level of the container to the KB level of the unikernel.

Computing-aware network is a framework solution on how to provide such services. Through the API call mode, it provides real-time calculation service with ultra-low delay at the nearest edge node, and non-real-time calculation service at cloud node. Ubiquitous field-level edge computing provides users with intelligent access and real-time data processing, and network-side edge computing provides users with rich computing power. Abundant network resources and computing resources will continue to integrate and complement each other, providing the ultimate user experience for new services.

4.4.4 Joint resource scheduling

To ensure the user's SLA requirements, the real-time communication resource status, computing resource status, and sensing resource status should be comprehensively considered for joint resource scheduling.

Joint resource scheduling (namely resource allocation) is a typical optimization problem. The problem may become more complex if the sensing resource, communication resource, and computing resource are jointly optimized, and there are multiple optimization objectives. ML (e.g. reinforcement learning) can directly be used for resource allocation. In this case, the most challenge is to find a suitable ML algorithm and define the resource status space and solution space for it. The problem can be simplified and decomposed into subproblems, such as computing node selection, radio resource allocation for wireless transmission, sensing, and positioning, and radio resource allocation for sidelink, uplink and downlink, respectively.

4.4.5 Block chain based resource sharing

The current mobile communication network is based on the resources controlled by each operator, and the service that each user can enjoy directly depends on the ability of its terminal. Due to the continuous enrichment of user application scenarios, the cost of a single operator to master all operating resources, as well as the complexity and price of each terminal will be higher. Therefore, various resources may be managed by different parties, and different application systems and services can be built through resource sharing.

The multi-party cooperation scenarios include single-type resource sharing and multi-type resource sharing, both of which can occur between different operators and different users.

In terms of communication resources, different operators can share communication resources and serve their own users. If the coverage in a certain region is unsatisfactory, one operator can use another operator’s base station and spectrum resources to serve its users, so as to improve the service quality. Similar problems can be solved by sharing communication resources between terminals. One terminal can seek the help of another terminal based on device-to- device (D2D) to complete its communication through the assisted communication link. Both cognitive radio and interference management can be attributed to spectrum sharing.

With the rapid development of integrated circuits and sensing technology, a large number of multifunctional sensing devices have emerged, which can complete multiple sensing tasks at the same time. The sensing devices deployed for a certain professional purpose may also be used for other purposes, and the related terminal can contribute to various large-scale data acquisition. For example, environmental monitoring can be carried out through a large number of users' temperature sensing data with temperature sensor sharing. A terminal without positioning function can get positioning information from surrounding terminals to complete its navigation task.

Computing (storage) resource sharing also exists between operators and users. When an edge computing operator cannot afford the services due to computing resource shortage, a part of computing task can be transferred to the edge platform owned by other operators around. On the user side, personal computing resources can be rented to other users (or even operators) in idle period to maximize the resource utility. Furthermore, computing service providers can rent edge nodes or user equipment to carry out computing services without hardware resources.

In addition to the above single-type resource sharing, the combination of multi-type resources has more application scenarios. In case of network slicing, a logic private network with different SLA requirements can be built with various link resources owned by different cooperates. Operators with camera resources can cooperate with roadside radar operators and positioning service providers to improve the accuracy of target recognition and tracking.

There are two key issues in multi-party resource sharing. One is how to confirm whether the resource owner has provided the corresponding resource (or transferred the right to use the resource) according to the contract, and the other is how to confirm whether the resource user has completed the payment to the resource provider according to the contract. In general, if there is a trusted third party, then the parties can complete the transaction under its coordination. However, there is no such trusted third party in many practical scenarios. The central third party may abuse its resource allocation right, and a single point of failure usually causes huge security risk.

In the absence of a trusted third party, blockchain technology can coordinate multiple parties to jointly maintain a unified transaction book, so it is very suitable for the resource sharing. Especially, alliance chain is a better solution for different operators to build resource sharing platform.

The alliance chain is constructed based on the interests of all parties in the resource transaction, where all participants determine the resource sharing rules and transaction mode, solidify the transaction execution steps and conditions in the form of smart contract, and independently execute the smart contract, complete transaction books and maintain their consistency based on consensus. The verifiable transaction data is constructed based on resource sharing logic. The identity of resource owner and transaction sender is guaranteed based on digital signature. By doing so, the alliance chain transforms the trust among multiple owners into the trust of chain data and smart contracts. All transaction behaviors are well documented and cannot be denied. All rules are automatically implemented without human intervention. The reputation of each resource sharing participant can be objectively evaluated based on transaction records. With a reasonable incentive mechanism, the alliance chain will attract more operators or individuals to contribute their own resources, thus greatly improving the resource utilization efficiency and expanding more sharing economy modes.

4.5 Service continuity

Service continuity management is required to support the continuity of AT interaction services. The continuity includes coverage continuity, sensing continuity, computation continuity and application continuity. Coverage continuity refers to the access point handover, sensing continuity refers to sensor handover or mobile sensing, computation continuity is for computing task reassignment, and application continuity is for application relocation.

4.5.1 Sensing continuity

Each sensor has its limited sensing range with certain requirements on sensing performance. To support service continuity, the sensors (e.g. mmWave Radar, camera) equipped on BS, or the radio sensing signal should be designed to ensure seamless detection coverage.

4.5.2 Communication continuity

In order to ensure the communication continuity, coverage mechanisms, including inter-frequency and intra-frequency service continuity and mobility management mechanisms are should be considered. That is to say, increase communication coverage or access point switching management. First of all, using Non-Terrestrial Networks (NTN) can expand the network coverage, providing service continuity to reinforce the service reliability. NTN are expected to serve areas that cannot be covered by terrestrial network (isolated/remote areas, on board aircrafts or vessels) and underserved areas (e.g. sub-urban/rural areas). Then, dual-connectivity mechanisms can improve wireless resource utilization, reduce system switching delay, and improve user and system performance. If the UE have TN and NTN connectivity capabilities, the dual-connectivity mechanisms between NTN and TN can be considered. Finally, except for the UE switch within the TN or NTN, the studies are should focus on the mobility mechanisms for the cases where UE's connectivity changes from the NTN to TN ('hand-in') and where UE's connectivity changes from the TN to NTN ('hand-out').

4.5.3 Computing continuity

Cooperative ML needs computing task decomposition and assignment among the cloud, edge and AT. In case of task process change, access point switching, or computing resource scheduling, the computing tasks may be migrated or reallocated. Here, the allocation algorithm needs to guarantee the continuity of reasoning and decision-making.

4.5.4 Application continuity

Due to the access point switching, or the need of traffic and computing power balance, the application instances and contexts need to be migrated from one platform to another.

4.6 Capability exposure

The objective of capability exposure is to provide essential capabilities and information to the third party application service provider. It is based on various functional entities in the mobile network. In 3GPP R15, network exposure function (NEF) is defined to provide standard services for application function (AF), such as QoS capability, event monitoring, parameter configuration, device triggering, traffic guidance, etc. R16 specification further enhances these services defined in R15. At the same time, it also adds some new services such as non IP data delivery, data analysis, network status, IPTV configuration, etc. Besides, more network capabilities are abstracted out for application layer including network slicing service, edge computing service, location service, 5G voice and message, data analysis service, and AI service. These service capabilities will have a wide range of vertical application scenarios in transportation, industry, energy, and many other industries.

In addition to the service capabilities defined in 3GPP R15/R16, the integration of sensing, communication and computing introduces sensing services, synchronization services, positioning services, ML services, as well as the information sharing of user/service/network/terminal attribute and status.

The capability exposure architecture is shown in Fig.7. It contains three layers: application layer, capability layer and resource layer. Application layer is the demander of network capacities. The demand on the capability exposure in terms of sensing, computing, positioning and synchronization is triggered by the unmanned services and assisted services. The northbound API interface is used to bridge the demand and the open capacities.

The capability layer interworks with the application layer through northbound interface and connects with the resource layer through southbound interface. It is the core layer of capability exposure and primarily responsible for capability modeling and encapsulation, capability orchestration and O&M, and capability announcement.

Resource layer completes the abstract definition of the underlying network resources, gathers the information and data on control plane, user plane and AI plane, as well as the information about the user attribute and status, network attribute and status, performs parameter configuration and resource scheduling according to the call logic and strategy of resource and capability from the upper layer.

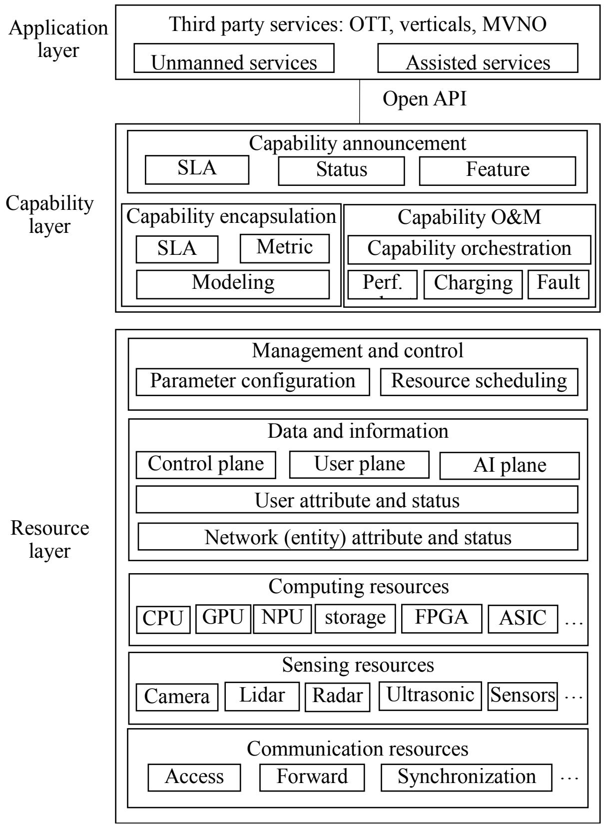

Fig.7: Capability exposure architecture

4.7 Challenges

The demand of E2E information processing promotes integration of sensing, communication and computing, which brings challenges to fundamental theory, devices, chips and algorithms.

- On the basis of information theory, computability theory and cybernetics, the integration of sensing, communication and computing needs to expand and integrate a new theory to analyze the coupling relationship among them.
- The integration of sensing, communication and computing involves ubiquitous heterogeneous communication and sensing. New communication and sensing technologies working on the bands of visible light, millimeter wave and THz need new devices support.

- The integration of sensing, communication and computing involves a large number of heterogeneous computing, which requires the support of various general-purpose and special-purpose chips. In particular, cooperative ML needs more powerful AI chips.
- No matter the detection of wireless signals and sensing signals, or AI reasoning and decision-making, it is necessary to develop new algorithms to ensure the integrated function and performance.
- A mechanism is needed to ensure that all relevant resources are schedulable, sharable and distributable in the integration of sensing, communication, and computing.

5. Application Scenarios

5.1 Interaction of autonomous things

Interaction of ATs is expected to be an underlying capability to support unmanned services, such as auto-driving, unmanned manufacturing, UAV, unmanned logistics, etc.

5.1.1 Interaction Model

AT interaction refers to the behavior of information exchange by means of sensing or communication to achieve an agreement for a common task.

An AT interaction model is shown in Fig.8. It is discussed within a task process and only considers two basic actions: cooperative sensing (perception) and cooperative ML. Cooperative ML involves cooperative training, cooperative reasoning (inference) and cooperative decision-making. Then, the interaction model consists of four levels: sensing, training, reasoning and decision-making. The cooperation level may differ in case of diverse task scenarios and AT capabilities. The different cooperative levels require to exchange different kind of data and information. And then require different radio functions and performance.

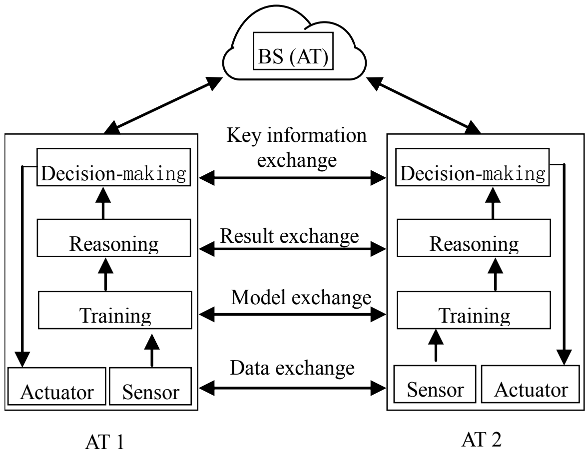

Fig.8: AT interaction model for cooperation ML

5.1.2 Cooperative ML

Cooperative sensing is a process to improve perception dimension and depth for ATs with data exchange through 6G radio. This is also known as data sharing. The exchanged data can be raw data or training sets. The key problem of cooperative sensing is data fusion, which is to delete redundant information and combine complementary information from multiple sensing data in the space-time dimension to obtain the consistent description of sensing objects. According to the fusion results, the system will carry out cooperative training.

Cooperative training is a process to improve training model performance for ATs with preliminary model exchange. An AT with stronger capability can help another AT with model training. ATs can also help each other with model test and optimization. In a more typical case, ATs do local model training and these preliminary local models will be collected together to form a complete global model.

Cooperative reasoning is a process to improve inference precision and reliability for ATs with local inference results exchange. ATs can help each other with result correction, or do local reasoning based on their own resources and models and send the local inference results together to determine the final global conclusion. Several iterations may be needed to achieve a satisfied final conclusion. The ATs joining in the process form a reasoning network, which reasoning ability can be improved by optimizing its architecture.

Cooperative decision-making is a process for ATs to reach an action agreement.

It is an optimization problem with clear objectives and constraints. It can be decomposed into several sub-problems and assigned to each AT for local solution. The final decision solution is formed by these local decisions according to fusion rules. Cooperative decision does not depend on the type of sensors and the sensing data, and has a good anti-interference ability.

Cooperative ML is actually a data fusion process, which is divided into model level, reasoning level and decision level accordingly. Data fusion needs to exchange different information under different ML frameworks. For example, vertical federated learning needs to share the data sets with different features associated with the same target, horizontal federated learning needs to share the data sets of different targets with the same feature, while transfer learning needs the multi-dimensional data sets of different targets.

It must be noted that cloud or MEC may participate in the cooperative sensing and ML through 6G radio. This means the 6G radio must have an ability to support the cloud-network-terminal cooperation.

5.1.3 Federated ML

With continuously improving capability of cameras and sensors on mobile devices, valuable training data, which are essential for AI/ML model training are increasingly generated on the devices. For many AI/ML tasks, the fragmented data collected by mobile devices are essential for training a global model. In the traditional approaches, the training data gathered by mobile devices are centralized to the cloud datacenter for a centralized training. However, cloud-based training means that the enormous amount of training data should be shipped from devices to the cloud, incurring prohibitive communication overhead as well as the concern on data privacy.

In Federated Learning (FL) mode, the cloud server trains a global model by aggregating local models partially-trained by each end devices. The most agreeable FL algorithm so far is based on the iterative model averaging. As depicted in Fig. 9 within each training iteration, a UE performs the training based on the model downloaded from the AI server using the local training data. Then the UE reports the interim training results (e.g., gradients for the DNN) to the cloud server via UL channels. The server aggregates the gradients from the UEs, and updates the global model. Next, the updated global model is distributed to the UEs via DL channels and the UEs can perform the training for the next iteration.

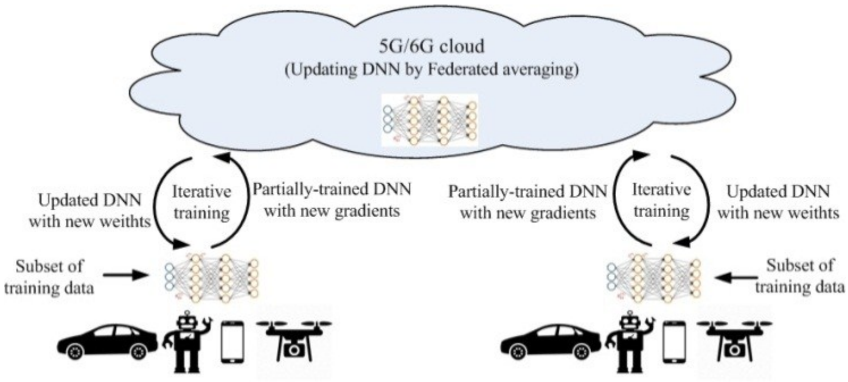

Fig.9: Federated Learning over wireless network

Different from the decentralized training operated in cloud datacenters, FL over wireless communications systems need to be modified to adapt to the variable wireless channel conditions, unstable training resource on mobile devices and the device heterogeneity. The FL protocol for wireless communications can be depicted in Fig. 9.

For each iteration, the training devices can firstly be selected. The candidate training devices report their computation resource available for the training task to the FL server. The FL server makes the training device selection based on the reports from the devices and other conditions, e.g. the devices' wireless channel conditions. After the training devices are selected, the FL server will send the training configurations to the selected training devices, together with global model for training. A training device starts training based on the received global model and training configuration. When finishing the local training, a UE reports its interim training results (e.g., gradients for the DNN) to the FL server. In Fig. 10, the training device selection is performed and the training configurations are sent to the training devices at the beginning of each iteration. If the conditions (e.g. device's computation resource, wireless channel condition) are not changed, the training device re-selection and training re-configuration are not needed, i.e. the same group of training devices can participate the training with the same configuration for multiple iterations.

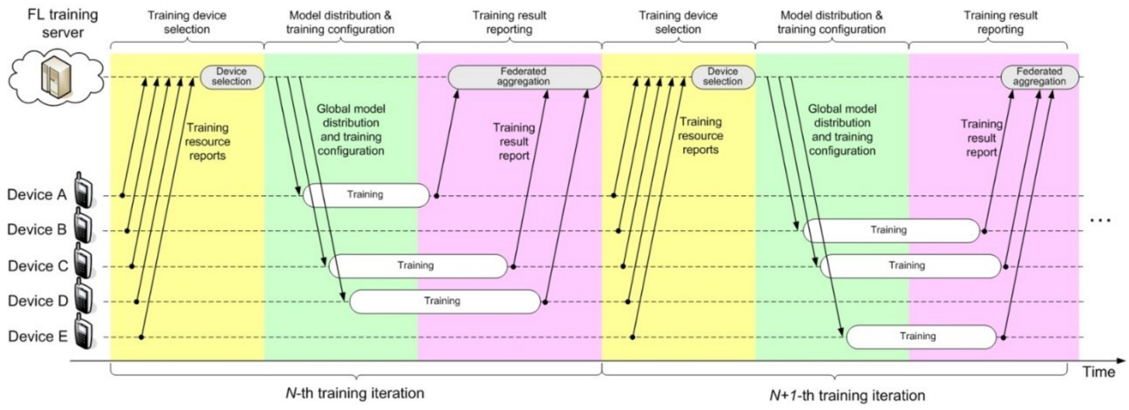

Fig.10: Typical Federated Learning protocol over wireless network

5.2 Immersive XR

Immersive XR refers to an interactive XR space which provides a natural experience through intelligent HMI. HMI usually refers to the perceptible (visible, audible, touchable, etc.) part of XR system, through which the user can communicate with the XR system and operate. To realize the interaction naturalness, HMI may use user model, domain model, task model, conversation model and media model to perform reasoning and expression. The main research contents include: multi-modality interaction, emotional computing, natural language understanding, virtual reality, intelligent user interface, etc. Modality covers various communication methods for users to express their intention, perform actions or perceive feedback information, such as speech, eye, facial expression, lip movement, manual, gesture, head movement, body posture/gesture, touch, smell or taste. For example, to track user’s posture, XR system should do 3D imaging and positioning using 3D light field technology.

Immersive XR requires deterministic information processing with multi-dimensional 3D sensing, cloud modeling, edge computing offloading and broadband transmission.

5.3 Digital twin

Digital twins are the artificial intelligent virtual replicas of physical spaces/ world/ systems. It develops with the spread of broadband and ubiquitous connectivity, AI, sensors, and improvements in big data processing and cloud computing. Digital twin technology is used to solve the problems in digital world which are hard to be resolved in the physical world with data analysis，ML, and simulation technologies. Digital twin is nowadays being applied into manufacturing, city, aviation processes, human body and even 6G network itself. Digital twin systems must be equipped with networking devices and sensors to guarantee a seamless connection and a continuous data collection and exchange either through direct physical communications or through indirect cloud-based connections (computing).

Taking digital twin network for example, networks failures can be early detected and prevented, new features can be fully verified before being activated. The introduction of new functions can be greatly accelerated and network operation efficiency can be extremely improved. The network will become an autonomous network that can evolve by itself.

6. Summary

This paper discusses the three fundamental functions of 6G: sensing, communication and computing, aims to define the basic concept of integration of sensing, communication and computing, and provide solutions for the E2E information processing for 6G. 6G is still in the initial conceptual stage, and needs a global cooperative and ecosystem to promote 6G innovation better.

References

[1] White Paper, “6G services and requirements,” FuTURE Forum 5G/6G SIG, Nov. 2020.

[2] Gartner. “Gartner Hype cycle for Emerging Technologies,2014~2019,” gartner.com/ smarterwithgartner.

[3] C. Sturm and W. Wiesbeck, "Waveform Design and Signal Processing Aspects for Fusion of Wireless Communications and Radar Sensing," in Proceedings of the IEEE, vol. 99, no. 7, pp. 1236-1259, July 2011.

[4] G. N. Saddik, R. S. Singh and E. R. Brown, "Ultra-Wideband Multifunctional Communications/Radar System," in IEEE Transactions on Microwave Theory and Techniques, vol. 55, no. 7, pp. 1431-1437, July 2007.

[5] P. Kumari, J. Choi, N. González-Prelcic and R. W. Heath, "IEEE 802.11ad-Based Radar: An Approach to Joint Vehicular Communication-Radar System," in IEEE Transactions on Vehicular Technology, vol. 67, no. 4, pp. 3012-3027, April 2018.

[6] J. A. Zhang, A. Cantoni, X. Huang, Y. J. Guo and R. W. Heath, "Framework for an Innovative Perceptive Mobile Network Using Joint Communication and Sensing," 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, 2017, pp. 1-5.

[7] Chiriyath AR, Ragi S, Mittelmann HD, Bliss DW. Radar Waveform Optimization for Joint Radar Communications Performance. Electronics. 2019; 8(12):1498.

Abbreviations

- **6G** The Sixth-Generation Mobile Communications
- **AI** Artificial Intelligence
- **AF** Application Function
- **AT** Automations Things
- **CBS** Control base stations
- **CSS** Chirped Spread Spectrum
- **CT** Communication Technology
- **D2D** Device-to- Device
- **DNN** Deep Neural Networks
- **DSSS** Direct Sequence Spread Spectrum
- **E2E** End-to-End
- **eMBB** Enhanced Mobile Broadband
- **FL** Federated Learning
- **GPTP** Generalized Precision Time Protocol
- **HMI** Human-Machine Interface
- **ICT** Information and Communication Technology
- **ICDT** IT, CT and big data technologies
- **IDC** Internet Datacenters
- **IoV** Internet of Vehicles

- **MF** Matched Filtering
- **ML** Machine learning
- **MMI** Machine-Machine Interface
- **NEF** Network Exposure Function
- **NFV** Network Function Virtualization
- **NTN** Non-Terrestrial Networks
- **PBS** Perception base stations
- **SDN** Software-Defined Networking
- **SIC** Serial Interference Cancellation
- **TBS** Traffic base stations
- **UAV** Unmanned Aerial Vehicle
- **URLLC** Ultra-reliable and Low Latency Communication
- **XR** Virtual/Augmented/Mixed reality

Acknowledgement

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

Editors：Chengkang Pan

Contributors：

China Mobile: Chengkang Pan, Ailing Wang, Yajuan Wang, Liang Ma, Guangyi Liu, Chunfeng Cui, Yifei Yuan, Qixing Wang, Junshuai Sun, Jing Jin, Sen Wang, Zhiming Zheng,

Xidian University: Liqiang Zhao, Kai Liang, Xiaoli Chu

BUPT: Zhiyong Feng, Zhiqing Wei, Qixun Zhang, Xu Chen, Wangjun Jiang, Qi Wu, Zeyang Meng, Xinyi Yang

Tianjin University: Zhen Gao

vivo: Jian Yao

OPPO: Wenqiang Tian

China Unicom: Rong Huang, Shan Liu

China Telcom: Xiaoyu Qiao, Song Zhao

Datang Mobile: Xin Su

未来移动通信论坛
FuTURE MOBILE COMMUNICATION FORUM