HUAWEI

Striding Towards
the Intelligent World
White Paper

Cloud Core Network

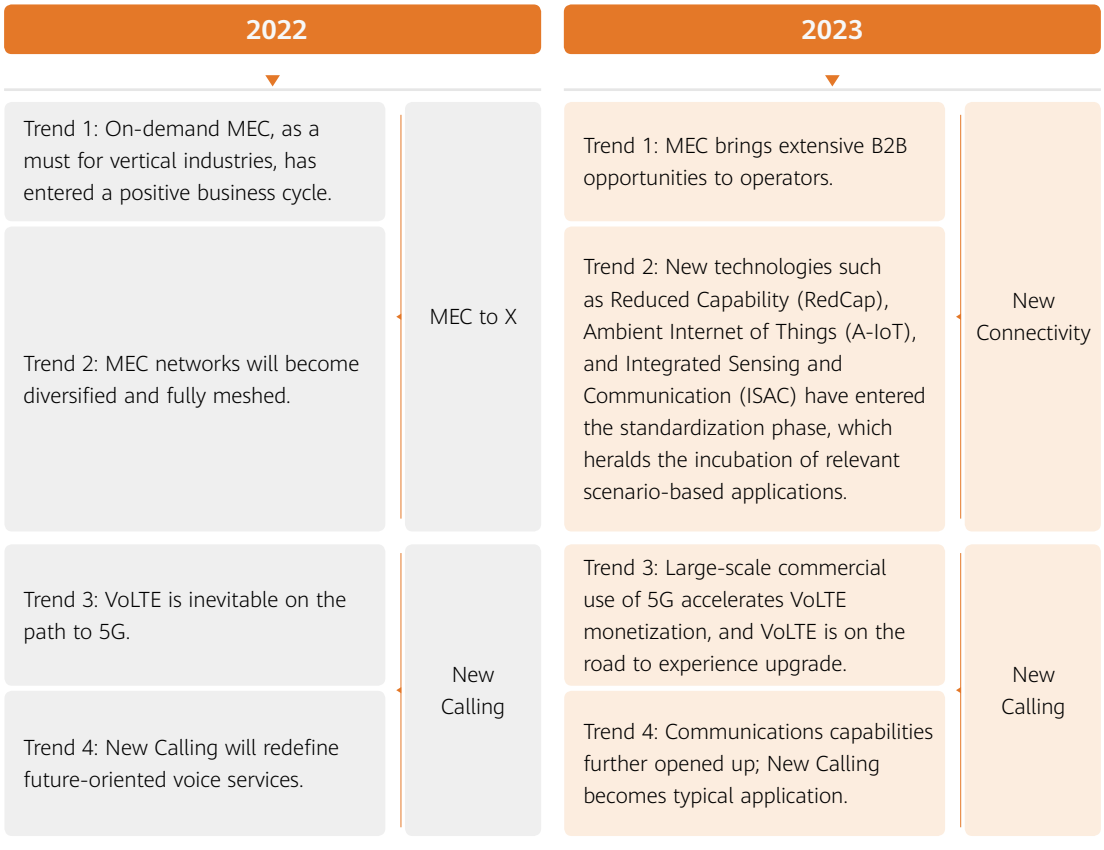
Intelligent Core Network Continuously Empowers New Business

Building a Fully Connected, Intelligent World

Abstract

Since the 3rd Generation Partnership Project (3GPP) officially launched the 5G-Advanced project in 2021 and Huawei proposed the concept of 5.5G Core, the telecom industry has been committed to researching 5G-Advanced and has already made great progress. The publishing of 3GPP Release 18 has highlighted 28 topics that are actively underway, while at the same time, all involved parties in the industry are proactively innovating and practicing, looking to seek out a more promising technical direction.

In 2022, Huawei Cloud Core Network released the white paper *Striding Towards the Intelligent World — Cloud Core Network,* which depicts eight trends of core network evolution and has exerted extensive influence in the industry. This year, Huawei Cloud Core Network continues to work with industry partners to innovate, practice, and gain new insights on 5.5G Core, and hereby releases a new edition of this white paper, delivering the thoughts and suggestions on the industry trends and providing a balanced reference for industry development.



2022		2023	
Trend 5: Operators' video services are converging with OTT applications at lightning speeds, stretching out from TV-based single large screens to OTT-oriented multiple screens, and delivering a better UHD video experience with enhanced network capabilities. Trend 6: Spatial and social-interactive videos will become central to operators' video services.	New Video	Trend 5: Glasses-free 3D is boosting mobile media traffic; ultra-large traffic processing and intelligent experience assurance are becoming trending topics.	New Video
Trend 7: Telecom cloud is the optimal approach to develop telecom services, and will continue to fulfill operators' diversifying requirements.	Telco Cloud	Trend 6: 5G core networks are 100% deployed over Telco Cloud, cloudification becomes the leading network construction model, and future-oriented convergence & simplification, standardization, and automation become the directions.	Telco Cloud
Trend 8: AI is indispensable for developing an IntelligentCore.	Intelligence	Trend 7: Foundation models are turbocharging AI progress, indispensable for 5.5G Core to achieve intelligent O&M, services, and networks.	Intelligence

Table 1 Key trends of core networks in 2022 and 2023

Insights into the core network trends of 2023

Trend 1: MEC Brings Extensive B2B Opportunities to Operators

5G private networks based on Multi-access Edge Computing (MEC) technology have been applied to multiple industries, such as manufacturing, energy, transportation, and government, bringing indisputable value for industry customers and promoting the large-scale deployment of such 5G private networks. Out of the 17,000 5G global private networks, China has a coverage rate of up to 90% in its cities that leads to far-reaching multiple service scenarios such as MEC to mobile office, MEC to electric power, MEC

to mining, and MEC to manufacturing. In these private network scenarios, more than 20,000 projects on 5G industry digitalization are involved, bringing a revenue growth of over 30% to operators' Business-to-Business (B2B) services.

MEC to mobile office: The Mobile VPN solution powered by MEC intelligent traffic steering has served 1300 enterprise customers. To date, it has delivered the mobile office service to 3 million users in schools, government departments, and large- and medium-sized enterprises, offering a superior experience, smooth connection, high reliability, and easy O&M. Leveraging the advantages of those widely-spread network lines, operators build a multi-plane national network to power dedicated services of industry customers. This network reaches far to the edge in the north-south direction through MEC and also enhances east-west interconnection. It allows industry customers to manage data locally, control security policies centrally, and connect service applications to the nearest network nodes.

MEC to electric power: Electric power is a mainstream application field of 5G in wide area scenarios. On a conventional electric power communication network, optical fiber-based direct transmission covers less than 20% of the power distribution stations. In this situation, a wireless private network needs to be constructed. Compared with 4G, 5G features security-oriented isolation, low latency, and high bandwidth. 5G supports integrated generation–grid–load–storage in a new electric power system. Since 2020, the pilot for 5G slice-based power grid virtual private networks covering 274 electric power enterprises has been initiated. More than 300 MEC sets are deployed to support applications such as power distribution telemetry, remote communication, automatic meter reading, distributed PV, and differential protection for power distribution. It is estimated that in 2024, operators will deploy 5G slice-based electric power private networks on a large scale.

MEC to mining: Operators have offered a voice and data converged communication solution for the mining industry such as coal mines. In this solution, MEC is deployed in mining areas and used with the IMS solution to realize real-time communication between personnel underground, between personnel underground and aboveground, and between public and private networks, significantly improving the production efficiency and safety. This solution has been successfully used at over 50 commercial coal mines, promoting smart mining applications and thus making mining intelligent. Furthermore, MEC can create a minimally attended or unattended working environment, improving work efficiency underground by 175%.

MEC to manufacturing: MEC has been deployed in more than 1000 factories and provides various basic applications for industrial enterprises, such as industrial quality inspection, data collection, AR-assisted standard operation, and production security monitoring. MEC adds value by replacing wired networks with wireless networks through 5G, providing basic connections for northbound PLC applications on industrial networks. In addition, MEC can be further applied in areas ranging from auxiliary production to core production, providing industrial connections for southbound PLC applications. As technologies such as 5G LAN and LPHAP mature, OT-UPFs with high efficiency, density, availability, and stability are gradually put into commercial use.

Trend 2: New Technologies Such as RedCap, A-IoT, and ISAC Have Entered the Standardization Phase, Which Heralds the Incubation of Relevant Scenario-based Applications

At this present, 5G networks have not yet possessed corresponding enabling technologies to deal with medium-, low-, and ultra-low-speed IoT application scenarios. While in a not too dissimilar situation, cellular IoT communications are being constrained by technologies and architectures, resulting in considerable power consumption, high costs, and a blunt competitive edge. This is where 5G-Advanced comes in, to support various types of IoT access with the help of innovative technologies mentioned below.

- RedCap: simplifies terminal complexity, for example, by lowering the bandwidth requirement, reducing antennas, and adopting half-duplex communications. Furthermore, Extended Discontinuous Reception (eDRX) is introduced to save terminal power. As such, the costs of RedCap terminal chips and modules are lowered, and terminal power consumption is decreased. RedCap has been standardized through the efforts in 3GPP Releases 17 and 18, and the industry is also proactively promoting RedCap terminal chips. The commercial promotion of RedCap enables 5G-Advanced to tap into the medium-rate IoT industry application market, such as smart grid, video security, and industrial parks.

- A-IoT: Passive IoT has a broad market space. However, conventional passive IoT technologies (such as RFID, short for Radio Frequency Identification) demand manual scanning due to the requirement of short-distance communications. As such, it is difficult to implement automated management or large-scale contiguous coverage.

In contrast, the innovative, cellular-based A-IoT technology supports relatively longer-distance communications as well as greater contiguous coverage and networking advantages, fitting in various IoT scenarios that span wide areas. To address the requirements of ultra-low power consumption and simplified terminals of passive IoT, it is necessary to bring forth lightweight Non-access Stratum (NAS) protocols, tag access management, and security mechanisms.

- ISAC: Proposed by Huawei, it is one of the key technologies to enable 5G-Advanced. Empowered by ISAC, the 5G-Advanced network is able to reuse communication signals, in order to implement sensing functions, such as target detection, ranging, positioning, and identification. The 5G-Advanced core network has introduced the Sensing Function (SF) to control data sensing, perform intelligent computing based on sensed data, anonymize sensed data, expose capabilities, and realize charging based on sensed data.

Trend 3: Large-Scale Commercial Use of 5G Accelerates VoLTE Monetization, and VoLTE Is on the Road to Experience Upgrade

In the 5G era, Voice over LTE (VoLTE) functions as a key link for ensuring 2G, 3G, 4G, and 5G voice service continuity, and has evolved from being an option to a necessity. In addition, the large-scale commercial use of 5G is driving operators to accelerate VoLTE development, helping them refarm the 2G/3G spectrum and optimize their core spectrum assets. Leading operators with well-constructed VoLTE networks are fostering innovative services to unlock the value of VoLTE networks and redefine user experience.

According to GSMA Intelligence, the number of global VoLTE and Voice over NR (VoNR) connections reached 3.9 billion by Q4 of 2022 and is projected to exceed 5 billion by 2025. The rapid development of VoLTE relies on the prevalence of VoLTE-capable terminals and widespread Long Term Evolution (LTE) coverage. On one hand, VoLTE is now automatically enabled on terminals. As smart terminals continue to upgrade, operators can easily migrate calling services to VoLTE networks when onboarding users. On the other hand, global LTE coverage rate is approaching that of Global System for Mobile Communications (GSM), which ensures optimal user experience.

In developing VoLTE markets, operators have refarmed or are in the process of refarming spectrum resources by migrating users to optimize spectrum assets and enhance their value. For example, operator S in Indonesia migrated voice services from Universal

Mobile Telecommunications System (UMTS) to VoLTE and refarmed 5.3 MHz spectrum in the 2.1GHz band to LTE. This increased the Dataflow of Usage (DOU) per user by 34% on average, and annual revenue by US$3.3 million.

In developed VoLTE markets, users demand more diversified communication services, such as video calling and interactive calling. However, legacy calling services cannot meet these requirements for service experience improvement. To enhance user loyalty and transform calling into a platform product, leading operators are introducing New Calling services centered on voice and video calls and built upon a large VoLTE user base.

Given the above, operators, vendors, and industry analysts have reached a consensus that accelerating VoLTE construction and developing New Calling services are crucial.

Trend 4: Communications Capabilities Further Opened Up; New Calling Becomes Typical Application

New Calling is booming in popularity, and commercial deployment has become widespread. In January 2023, China Mobile started commercial trials and beta tests. In June, China Mobile announced that they will complete the construction of a New Calling network serving tens of millions of subscribers by the end of 2023. In this process, Huawei has worked with China Mobile and developed multiple New Calling services, such as fun calling, real-time translation, and visualized voice calling. As the commercial trials approach their end, China Mobile is preparing to release these services for formal commercial use. Operators are currently benefiting from New Calling — they are gradually transforming their business model from voice-only operations to video and content operations.

During MWC 2023, Network as a Service (NaaS) came to the forefront of conversation again. New Calling is considered as a typical NaaS application if you consider the following:

Content-based operations: Take the visualized voice calling service as an example. This service allows users to present customized virtual avatars to the other party during voice calls without having to go on video, enabling them to better express themselves when they chat. For such services, rich content is of great importance. New Calling enables

networks to shine through standard APIs especially when combined with diversified content ecosystems like Professionally-Generated Content (PGC), User-Generated Content (UGC), and AI Generated Content (AIGC), bringing brand-new service experiences. In this way, operators can provide differentiated content for each individual.

Industry-oriented applications: New Calling is a calling-based platform product and provides differentiated service functions for each industry segment. New Calling abstracts and aggregates network capabilities such as dialing, authentication, and interaction, and opens them to industries such as finance and insurance, transportation and logistics, and call centers, helping industries reduce costs and improve efficiency. For example, in the vehicle damage assessment service, traditional offline assessment is time and labor consuming; app-based remote assessment is not widely used due to complex operations. To solve the preceding problems, Jingyou Technology collaborated with Huawei and developed the 5G video-based damage assessment solution. Its assessment system initiates a video call to the vehicle owner and uses the AR annotation function to guide the owner to take photos of the damage. Then, the system captures images and generates quotations. This solution shortens the average damage assessment duration from two hours to five minutes, greatly reducing the operating expense (OPEX) and improving service efficiency as well as customer satisfaction.

Fast service iteration: New Calling decouples services from the network, which helps agile application innovation and service iteration. It sustains the high stability and reliability of the underlying network and ensures the availability of basic calling services. Operators can use this to devise a clear approach to building New Calling networks.

Trend 5: Glasses-free 3D Is Boosting Mobile Media Traffic; Ultra-Large Traffic Processing and Intelligent Experience Assurance Are Becoming Trending Topics

3D video has been through a relatively sluggish development period for a long time due to content production issues. However, as 2D-to-3D conversion technologies mature and become widely adopted, existing 2D content can be quickly converted to 3D, making 3D content production no longer an issue. On top of that, glasses-free 3D screen technology makes 3D-capable screens affordable, and the powerful 5G network is ready for 3D services. With all these, it is now the prime time for glasses-free 3D services to boom.

The explosive development of glasses-free 3D will also promote the development of the entire spatial video industry and numerous other video services. For example, Extended Reality (XR), a cousin of 3D, has great market potential and is growing rapidly in terms of device shipments. According to Huawei Global Industry Vision (GIV) statistics, there will be over 1 billion XR users by 2030, and another set of statistics predict that the XR device shipment this year will be 17% higher than that in 2022. These emerging video services including 3D video, XR, and the increasingly popular cloud gaming and cloud live streaming all bring vast possibilities for mobile media development, but also pose new requirements on the network.

Requirements for ultra-high downlink traffic offloading: The immersive 3D and XR experiences require a bandwidth rate up to 10 Gbit/s. To better processing the large traffic, dedicated traffic offloading nodes need to be deployed on the network and the service flows need to be routed to the shortest paths. The form of the user plane and the network scheduling mechanism also need to change accordingly.

Requirements for high uplink bandwidth assurance: The live streaming market, especially the e-commerce live streaming sector, has experienced remarkable growth in recent years. According to statistics, there are more than 7000 live streaming operation companies and over 60 million live streamers with 10K+ followers in China, producing a Gross Merchandise Volume (GMV) of US$180 billion. Live streaming quality is highly reliant on the network, and in cases of poor network quality and limited bandwidth, this may cause severe impacts on live streaming (especially the walk streaming and outdoor streaming). For example, the streamers may lose viewers, followers, or even deals when network frame freezing occurs. In this context, services such as live streaming have an urgent demand for a high uplink bandwidth.

Requirements for network native intelligent media processing: The smartphone replacement cycle is extended these years, and according to Digitimes, global smartphone shipment has seen a decline for seven straight quarters since 2021. So, how can users enjoy a better experience without changing mobile phones? The network computing and native intelligent processing can help. Numerous emerging services are using network-provided computing capabilities, such as cloud-assisted mobile games and cloud photographing. And by moving media processing capabilities such as intelligent recognition and real-time rendering to the network edge, the service experience on mobile phones with limited computing power will be greatly improved.

To meet all the requirements, it is necessary for the network side to continuously optimize network transmission latency and enhance cloud-edge-device collaboration to provide optimal computing power. This is also the basis for ensuring optimal user experience, and poses higher technical requirements for one-hop computing and one-hop connection on 5.5G networks.

Trend 6: 5G Core Networks Are 100% Deployed over Telco Cloud, Cloudification Becomes the Leading Network Construction Model, and Future-oriented Convergence & Simplification, Standardization, and Automation Become the Directions

1. Cloudification becomes the leading network construction model, and 5G core networks are 100% deployed over cloud.

In the past few years, remarkable strides have been made in Telco Cloud development. Over 200,000 servers have been delivered worldwide. So far, 5G core networks are 100% deployed over Telco Cloud. It is estimated that by 2028, with the sunsetting of legacy devices, all core networks will be cloudified.

In addition, operators, standards organizations, and device vendors have made joint efforts in building carrier-grade reliability and high performance capabilities based on IT hardware.

Heterogeneous compute resources have been put into widescale commercial use. Telco Cloud supports both x86 and Arm architectures. Without changing the networking architecture and resource configuration, one cloud is able to manage two types of compute resources and support smooth evolution and upgrades.

2. Deployment of public clouds in the telecom industry faces uncertainties and carries multiple challenges.

In the past year, telecom networks deployed based on public clouds were still rare. Telecom services differ significantly from common IT applications:

The public cloud deployment architecture and network topology cannot meet carrier-grade reliability requirements. Although the on-premises deployment model is

proposed in other solutions to mitigate fault spreading, this model makes a public cloud masquerade as a private cloud to some extent. In this case, public clouds will no longer advance in intensive sharing.

From the perspective of O&M, few public cloud vendors can guarantee carrier-grade availability as expected. The availability of public clouds is far from carrier-grade requirements, especially the high stability and reliability requirements of the core network.

Public clouds are still facing the challenges of data sovereignty, data security, and multi-vendor interconnection, and have not proved their unique advantages over the existing Telco Cloud. As such, Telco Cloud is still preferred by operators.

3. Converged and simplified network construction is recommended.

Containerization is a major concern of operators. What is the best way to introduce containers to the telecom industry? Should operators expand the existing cloud or deploy a new cloud? Striking a balance between the resource management advantages of OpenStack and application-oriented management flexibility (such as using K8s) becomes essential.

It is believed that the current NFV architecture and inventory networks are optimal, and have been developed after years of adaptation to operators' organizational capabilities, network architectures, and personnel skills. It's a natural choice that operators introduce containers seamlessly on an existing network without changing the networking architecture, losing achievements (such as intensification and software and hardware decoupling), or overturning existing interface interconnection. We need one cloud to flexibly schedule resources including VMs, containers, and bare metal servers, share compute, storage, and network resources, and support seamless expansion. This solution has already been piloted at sites of China Mobile and STC in Saudi Arabia.

4. Intent-driven approaches are used for Telco Cloud native automation, leveraging cloudification benefits.

Both CNF-level and network-level automation are required for telecom networks. Deploying a function is not just about the deployment of clusters and Network

Functions (NFs), cross-NF association is also a must to meet the requirements of telecom services such as 5G slicing and private line provisioning.

In addition, the telecom cloud configuration needs to be further simplified. By using automatic network design, configuration, and intent-based configuration conversion, the parameter configuration workload will be reduced by 90%. Northbound interfaces are standardized to shield cross-vendor implementation differences. Through configuration dry run and simulation verification, configuration errors are intercepted in advance.

The objective is to achieve full-scenario automation for the telecom ecosystem, including intent-based health check, automatic test, and online attendance based on intent-related data analytics.

The ultimate objective is to meet the operators' organization and capability requirements, and ease the work of operators. Operator maintenance personnel focus on service and network O&M. For product O&M, AI can be introduced to achieve self-planning and self-optimization for closed-loop operation.

5. ETSI standards are complied with for smooth evolution.

Thanks to the joint efforts of industry partners, the past decade has witnessed the growth of ETSI NFV. Best practices suitable for operators' organizational capabilities, network architecture, and personnel skills have since been explored. NFV's success also relies on the ETSI NFV architecture and a large number of interoperability verifications and certifications. With the evolution of containers, introduction of new hardware, and new automation requirements, joint efforts among all industry players are still needed for advancing the ETSI NFV standards and preventing operators from falling into new vendor lock-in due to new technology silos.

To cope with the growing complexity and to simplify network operations, standardized interfaces are required for southbound and northbound communication. Based on the use of commercial off-the-shelf (COTS) hardware, its drivers should be standardized and open to reduce the complexity of multi-vendor integration and reduce the workload of development and testing as a means to facilitate interoperability. In this way, the cloud platform can quickly integrate and interconnect with various hardware. ETSI NFV standards and solutions mainly focused on MANO's core capabilities such

as lifecycle management. In Q2 2023, MANO evolution was incorporated in NFV's Release 6 Definition with the focus on enhancing digital and intelligent O&M.

Trend 7: Foundation Models Are Turbocharging AI Progress, Indispensable for 5.5G Core to Achieve Intelligent O&M, Services, and Networks

2022 marked the advent of foundation models. The major breakthrough in the foundation model technology will completely transform the AI supply model and make it easier for the supply side to provide diversified services. AI technology is about to be applied on a large scale. AI has progressed from being able to process basic logic and lists of information (AI 1.0) to AI 2.0 represented by foundation models.

In the 5.5G era, Cloud Native and Full Convergence are advancing constantly, greatly increasing the scale and complexity of the cloud-based core network. In addition, service innovation and network optimization need to be enhanced on the core network to continuously improve user experience. Powered by AI foundation models, the 5.5G core needs to ride the wave of AI fever and implement a comprehensive transformation model for intelligent O&M, services, and networks, enabling new service experience and better networks.

CONTENTS

1 Overview

1.1 5.5G Core Network Stepping into the Commercial Deployment Phase After Preliminary Exploration

In 2021, 3GPP initiated the 5G-Advanced project to further refine 5G standards. On the basis of 5G-Advanced specifications and the rollout of multiple new Information and Communication Technologies (ICTs), the concept of 5.5G Core was proposed with an aim to build a mobile network that can continuously empower new business. Thanks to the collaboration of all involved parties, the 5.5G Core industry is advancing in innovative practices, ecosystem development, and standards promotion. Now, 5.5G Core is stepping into the commercial deployment phase.

In terms of the innovative practices of emerging technologies, Huawei is working with operators and industry partners to explore new scenarios and opportunities, aiming to tap into the value of 5.5G Core technologies.

- For the B2C sector, Huawei has worked with China Mobile to launch the User Plane Function (UPF) prototype with high throughputs, to cope with the high concurrency and heavy traffic brought by the increasing High Definition (HD) video, XR, and glasses-free 3D services. This UPF uses technologies such as dynamic multi-core aggregation and intelligent offloading, achieving a 10 Gbit/s throughput for a single user even with the presence of dynamic background traffic. It also has a 3 times higher capacity for concurrent media requests.

- For the B2B sector, Huawei has teamed up with operators and industry partners to offer 5.5G flexible production lines based on Operation Technology-UPFs (OT-UPFs). With cutting-edge technologies such as 5G LAN, multi-channel dual fed and selective receiving, and millisecond-level UPF hot backup, we can provide an industrial network with high reliability, low latency, and deterministic network services for the product lines. By doing this, it will be much easier to deploy Programmable Logic Controllers (PLCs) in a centralized manner and adjust the production lines. Generally, PLC deployment costs are reduced by 50%, production line adjustment time is shortened from 7 days to just 15 minutes, and the comprehensive production efficiency is improved by up to 12%.

- For voice services, Huawei has piloted New Calling services in multiple regions in China and will officially put it into commercial use across the nation by the end of this year. The New Calling solution adopts a 1+3+N architecture, and helps operators roll out new services based on calling, providing users with ultra-HD, intelligent, and interactive service experiences. It allows business transactions to be completed during calls, and enables thousands of vertical industries. For example, New Calling can be used in the processing of online insurance claims. A typical claim can be settled within approximately 5 minutes, which is much faster than a conventional insurance claim that may take up to 2 hours.

In terms of ecosystem development, Huawei has collaborated with industry peers to eliminate the bottlenecks for the 5.5G industry, laying a foundation for the commercial use of 5.5G. For example, to promote the construction of the Data Channel ecosystem, Huawei worked with Global System for Mobile Communications Association (GSMA), the world's leading operators, and industry partners to set up the 5G New Calling Foundry project. Through this project, we have made joint efforts to overcome hurdles and promote the development of a rich ecosystem. It is estimated that data channel–capable terminals will be launched in the fourth quarter of this year.

In terms of 5G-Advanced standards, all of the involved parties have made significant concerted efforts. In December 2021, 3GPP Release 18, the first standard involving 5G-Advanced, identified 28 research topics, introducing new technologies such as enhanced media processing capabilities, New Calling, and industry scenario enablement. However, the standardization process has been delayed due to the pandemic. It is planned that Release 18 will be issued in the first half of 2024. Plus, the development of Release 19 has also begun, which will introduce new technologies such as intelligence, network openness, new Internet of Things (IoT), ISAC, and new media capabilities. All of the above will further enhance the capabilities of mobile networks, enabling a wider range of business scenarios.

Through all these efforts during the past two years, 5.5G Core has demonstrated its technical value and potential, and a development direction has been mapped out for the 5.5G industry. Now, the conditions for commercial use of 5.5G are more or less ready. It is expected that 5.5G will be put into commercial use in 2024, and end-to-end solutions will be offered by then, supporting the construction of the 5.5G network for operators.

1.2 5.5G Core Featuring Enhanced Basic Network Capabilities and Introduced AI Native Technologies

5.5G Core will deliver indisputable business value. It is expected to underpin the application of B2B, B2C, and voice services in more scenarios while further extending to new sectors such as Business to Government (B2G) and Business to Home (B2H). To fulfil these expectations, 5.5G Core needs to incorporate new capabilities in two aspects.

One aspect involves enhancing basic network capabilities, specifically, more connections, a convergent architecture, and agile resource allocation.

- More connections: Multiple networks are required for different industry applications including the IoT, sensing, and communication. To simplify the network construction, Release 19 will define the low-power A-IoT technology and the ISAC that integrates communication and sensing. For AIoT, the core network will introduce the Tag Management Function (TMF) to ensure and manage secure access of massive tags. For ISAC, the core network will introduce the SF to sense multi-dimensional information, making the 5.5G network fit for multiple purposes.

- Convergent architecture: To empower industry private networks in more scenarios, technologies such as wide-area 5G LAN are introduced to converge the networking at Layer 2 and Layer 3. To enable edge applications, native computing is leveraged to integrate connection and computing capabilities. To provide an optimal media service experience, multiple media technologies such as Media over QUIC (MoQ) and XR Media (XRM) are adopted to combine communication and media capabilities.

- Agile resource allocation: The existing VM-based Telco Cloud platform will evolve to a VM-container convergent platform. The convergent platform will be more containerized and service-oriented, and can allocate cloud-based resources more flexibly.

The other aspect for enhancement of the 5.5G Core involves the introduction of AI technologies to enable more business scenarios through AI native.

The popularization of 5G stimulates higher requirements from both consumer and industry customers and also brings new challenges to the evolution of 5G. For example, the innovation of communication services needs to keep pace with users' increasing requirements for new models and content after New Calling is scaled up. Networks are

expected to offer real-time, multi-dimensional service awareness capabilities for service quantity evaluation as well as optimized resource scheduling capabilities for closed-loop user experience guarantee. Cloud-based networks require automated network planning, construction, maintenance, and optimization to respond to more complex networking and more demanding requirements of services on networks in the future. Thanks to the rapid development of machine learning and big models, AI has shown its significant value in fast content generation as well as learning-based awareness and automation. The combination of 5G and AI will bring remarkable improvements and changes to the core network.

- AI will bring about unprecedented changes to network services. It allows diversified communication content such as actions, facial expressions, voices, and texts. In the future, more and more intelligent digital humans will be involved in communication, significantly changing the way people live and work.

- AI will make the network more intelligent. By combining AI with networks, AI computing and applications will be available in various scenarios on demand, bringing ubiquitous intelligence to reality. In addition, AI native will enable human-like natural interaction between humans and digital humans and between humans and robots.

- AI will also simplify O&M. It assists O&M personnel to collect and analyze massive data, plan and manage complex network topologies and configurations. This will significantly improve O&M efficiency. In addition, the intent-driven technology can simplify man-machine interface interaction, helping achieve real automation.

In summary, enhanced basic network capabilities and AI native technologies will help build a 5.5G intelligent core network featuring intelligent services, intelligent network, and intelligent O&M. This network will empower more new business and help various industries move towards an intelligent world.

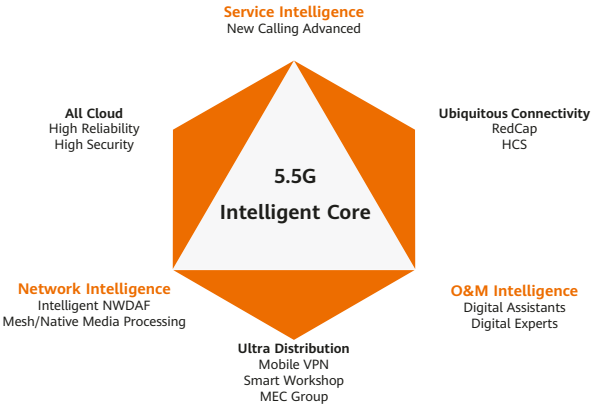


Figure 1 5.5G intelligent core

2

Trend 1

MEC Brings Extensive B2B Opportunities to Operators

Over the past five years of 5G commercial use, the number of subscribed users has increased exponentially and is estimated to reach 3.4 billion by 2030, accounting for 54% of mobile users. Meanwhile, 5G private networks based on MEC technology have been applied to multiple industries, such as manufacturing, energy, transportation, and government, bringing indisputable value for industry customers and promoting the large-scale deployment of such 5G private networks. Out of the 17,000 5G global private networks, China has a coverage rate of up to 90% in its cities that leads to far-reaching multiple service scenarios such as MEC to mobile office, MEC to electric power, MEC to mining, and MEC to manufacturing. In these private network scenarios, more than 20,000 projects on 5G industry digitalization are involved, bringing a revenue growth of over 30% to operators' B2B services.

As a new piece of infrastructure for the digital economy, 5G has accelerated its pace of integration into the real economy. In different industry scenarios, there have been multiple 5G private network solutions tailored for different requirements of industry customers. For example, in the MEC to mobile office scenario, 5G enables users to access the enterprise intranet without changing cards or numbers. The Mobile VPN solution expands the coverage of enterprise private networks and allows users to smoothly switch between Internet services and enterprise services. In the MEC to electric power scenario, the 5G power grid slice private network provides a control plane that is independently deployed. The network can be used for telemetry, remote signaling, and remote control, as well as distributed solar photovoltaic, second-level load control, precise load control, and differential protection, meeting the reliability and security requirements of the electric power industry. In the MEC to mining scenario, the 5G-powered One Number Converged Network solution realizes the interconnection between personnel underground and aboveground, as well as between public and private networks. People can use one number to enjoy both data and voice services on public and private networks. In the MEC to manufacturing scenario, 5G improves its stability and reliability and reduces latency. The Layer 2 private network solution supports the communication of Layer 2 devices, so that wireless networks can replace wired networks in manufacturing factories. The wireless network can then be used in multiple scenarios and for multiple purposes.

2.1 MEC to Mobile Office

Benefiting from the immense popularity of various smart terminals such as smartphones, mobile office or hybrid office users increase rapidly. Against this backdrop, MEC in

conjunction with 5G networks enables flexible, on-demand deployment at the network edge as well as intelligent traffic steering, thereby offering high-quality, efficient, and secure network services for mobile office users.

Smooth Connection, Optimal Experience, High Reliability, and Easy O&M

On the path heading to smart cities, MEC works with 5G networks to provide more efficient network access for public officers. Take the 5G e-government private network of a district government as an example. There were only 5 mobile office applications in 2019, which already increased to over 20 in 2022. The network supports the high-speed download of large 3D map files and fast browsing of multi-channel HD videos, reducing 15.3 minutes of network access time per person per day and cutting the overall network construction and O&M costs by 20%.

In the smart education field, fully-interconnected MEC sites in multiple cities are connected to the research and education private network. For example, there is a 5G education private network that connects 168 universities, bridging the education network and Internet and bringing ubiquitous education services to users. The network access rate is increased by 13 times from 56 Mbit/s (conventional VPN–based access) to 784 Mbit/s, significantly optimizing the experience of smart campus applications such as holographic classroom and cloud VR.

By June 2023, over 7000 smart city and smart education projects have been signed, reaching 31 provinces and cities in China, and more than 1500 MEC sites have been deployed, serving a user base exceeding 3 million.

Mobile VPN Achieves East-West Interconnection of MEC Sites and Offers a Layered Mesh Network for Operators

Huawei's Mobile VPN solution is well-suited for mobile office. This solution supports 4G/5G access, flexible traffic steering to nearby network nodes, and network-wide cross-domain access. With this solution, operators can enhance east-west interconnection of MEC sites and build a layered mesh network by customer or industry. In the future, the Service Routing Function (SRF) can be introduced to select the optimal anchors and steering points that match the cost, security, and data compliance policies, finding the best routes for sessions on the network.

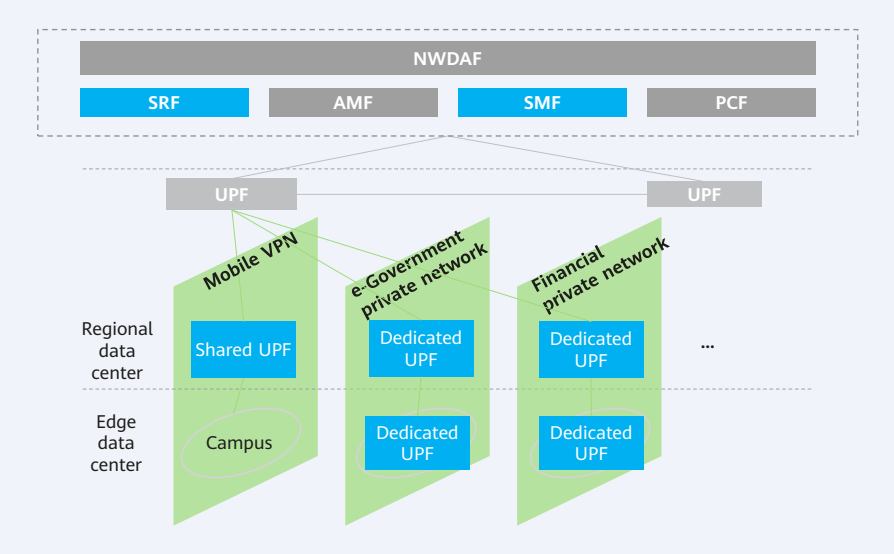


Figure 2 Networking architecture of MEC to mobile office

5G Private Networks Expand Beyond Smart Education and Smart City to Meet Mobile Office Requirements of More Industries

5G private network products are being adopted in a wider range of scenarios from the education and government sectors to healthcare, finance, manufacturing (industrial parks), and tourism (such as museums). In particular, for the healthcare industry, private networks are applied in pre-hospital first aid, inside-hospital guidance, and collaborative office between inside- and outside-hospital personnel; for the finance industry, they play a role in building 5G backup lines and connecting 5G Point of Sale (POS) devices; for the manufacturing industry, they are used in industrial parks for asset monitoring, mobile inspection, and collaborative office; and for the tourism industry, they can work in scenic spots and museums, supporting immersive AR/VR services for example. As more operators' edge nodes are used for east-west interconnection, it is estimated that millions of users from multiple industries will turn to mobile office. This will further boost the deployment of MEC sites in cities and campuses as well as dedicated network planes.

2.2 MEC to Electric Power

Electric power substations and power distribution stations are usually widely dispersed and numerous in quantity. Conventional optical fiber-based networks dedicated for

the electric power industry have prominent deficiencies such as high construction and maintenance costs and insufficient coverage. For instance, in China, optical fiber-based transmission covers less than 20% of power distribution stations providing a voltage below 35 kV. In the 4G era, the electric power industry adopts private networks with dedicated frequencies. Distinct pain points, such as spectrum interference, low bandwidth, and immaturity, emerge during this process. Meanwhile, to achieve the global sustainable development goal of carbon neutrality and carbon peaking, a new electric power system needs to be constructed, with renewable energy as the primary energy source. This causes new constraints such as unstable power generation of renewable energy as well as climate and weather factors. To break the constraints, an advanced wireless communication private network is urgently needed to help construct a new electric power system highlighted by integrated generation–grid–load–storage.

Typical Applications Are Under Rapid Development and Ready for Large-Scale Commercial Use

Compared with the conventional power supply model of generation–transmission–distribution–use, the new electric power system is highlighted by integrated generation–grid–load–storage. For power generation, 5G helps the electric power industry improve working efficiency. For power grid, 5G-powered differential protection for power distribution as well as telemetry, remote communication, remote control, power distribution automation, and secure PV grid connection help improve power distribution efficiency. For power load, 5G second-level precise load control ensures power supply to residential users.

5G has been piloted in the electric power industry since 2020 and gradually put into large-scale commercial use. To date, operators have completed the first batch of RedCap commercial pilots with industry partners, especially Huawei. In Hangzhou, for example, the world's first end-to-end commercial pilot of RedCap in the "5G+electric power" scenario has been established. This promotes the large-scale application of RedCap in the electric power industry and accelerates the construction of a new electric power system.

5G Slice-based Electric Power Private Network Ensures Secure and Reliable Communications

5G provides customized "slice-based private network" services for users in the electric power industry. Compared with 4G, 5G provides end-to-end slicing based on network virtualization functions and achieves isolation of physical network resources. This secures power grid services by isolating services in zone 1 and zone 2 that form the

production control zone from those in zone 3 and zone 4 that form the management information zone. In summary, 5G slicing provides differentiated service assurance in terms of bandwidth and latency and enhances self-control capabilities for electric power enterprises.

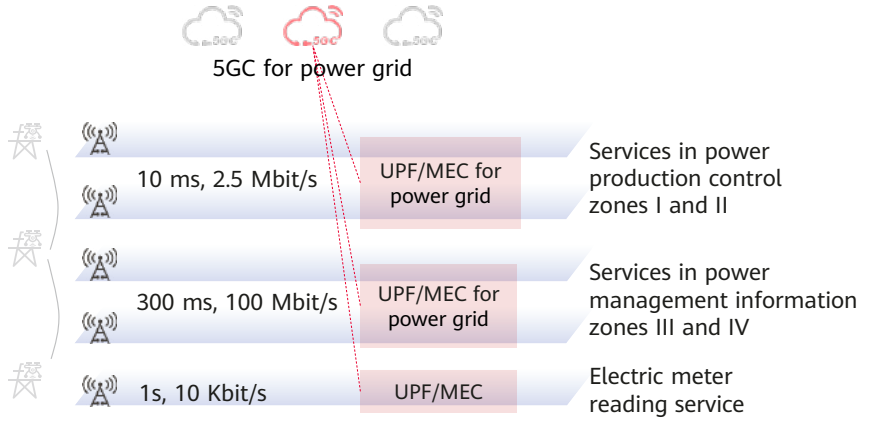


Figure 3 Networking architecture for MEC to electric power

Based on the 5G SA networking architecture, a 5G Core Network (5GC) control plane is independently constructed in the operator's private network resource pool. This ensures the connection to a large number of wide-area base stations and allows numerous power distribution nodes to access the network. Such a private network shares base stations as well as the chip and module ecosystem of the upstream and downstream industry chains with public networks, simplifying network construction.

The 5G SA networking architecture separates the control plane from the forwarding plane. The forwarding plane can be independently deployed nearby to guarantee the quality of services requiring millisecond-level load control and low latency, and avoid traffic detour of services requiring high bandwidth. In addition, their separation facilitates layered deployment of electric power networks at provincial, regional, and substation locations to improve transmission efficiency.

Emerging 5G Scenarios Drive Low-Carbon and High-Quality Development

5G creates new possibilities and opportunities for the electric power industry. The use of MEC in the electric power industry has shifted from pilot demonstration to large-scale promotion. As 5G technologies are optimized to become more mature, and

more spectrum resources are released and utilized, there will be rich and diversified 5G applications in the electric power industry. Currently, intelligent power distribution networks are mainly used to cover tens of millions of power distribution nodes. Looking ahead, thousands of MEC sets will be deployed in a distributed manner for integrated power generation–grid–load–storage. MEC sets are supposed to support latency-sensitive 5G applications, such as differential protection for power distribution, distributed FA, load control at the 100-ms-level, and adjustable load charging piles. In summary, MEC is used to implement a new electric power system, which features high reliability, high efficiency, and low-carbon consumption and aims to promote socioeconomic development.

2.3 MEC to Mining

Mining private networks are in rigid demand for a voice-based dispatch system, which can be widely used at coal mines, metallic mines, and non-metallic mines as well as underground and open-pit mines. Conventional private network solutions have drawbacks such as poor signal coverage, low network bandwidth, poor call quality, and monotonous functions. As such, the 5G private network solution has become a must for intelligent and digital mining.

Cost Reduction and Efficiency Improvement Are Achieved in Valued Scenarios

Operators over the globe are exploring 5G application in the mining industry. For example, at a coal mine in China, a 5G private network is deployed to cover mine plants, living quarters above the ground, as well as mining faces, electromechanical chambers, and the main haulage roadways under the ground. With this converged 5G communication network, applications such as 5G+ intelligent, remote mining control, unattended inspection, unattended mining truck, and converged dispatching communication are deployed. Together, they help increase the production output by 21%, the production efficiency by 20%, and the dispatch efficiency by 50%, as well as reduce the onsite manpower by 12%.

Let's see a typical use case outside China. At a large open-pit mine, operators have offered a 5G private network and deployed digital productivity tools, which span the mining trunk dispatch system, mining trunk fuel consumption monitoring, driver fatigue detection, and voice data communication between employees and visitors. This improves

the dispatch efficiency by 60% and reduces energy consumption of mining trunks by 10%.

By April 2023, 5G smart mining has been implemented for over 500 mining enterprises in 27 provincial-level regions in China. In total, 1043 intelligent mechanized faces and 1277 intelligent excavation faces have been built, and more than 400 MEC sets have been deployed.

One Number Converged Network Offers Both Voice and Data Services

To address the pain points of conventional private networks, Huawei launches the MEC-based One Number Converged Network solution. The 5GC private network uses MEC and the IP Multimedia Subsystem (IMS) as its foundation in mining areas. It connects to the public network to realize real-time communication between personnel underground and aboveground, as well as between public and private networks. In addition, it interconnects with the dispatch system to implement real-time converged dispatching and commanding for safe production in mining areas.

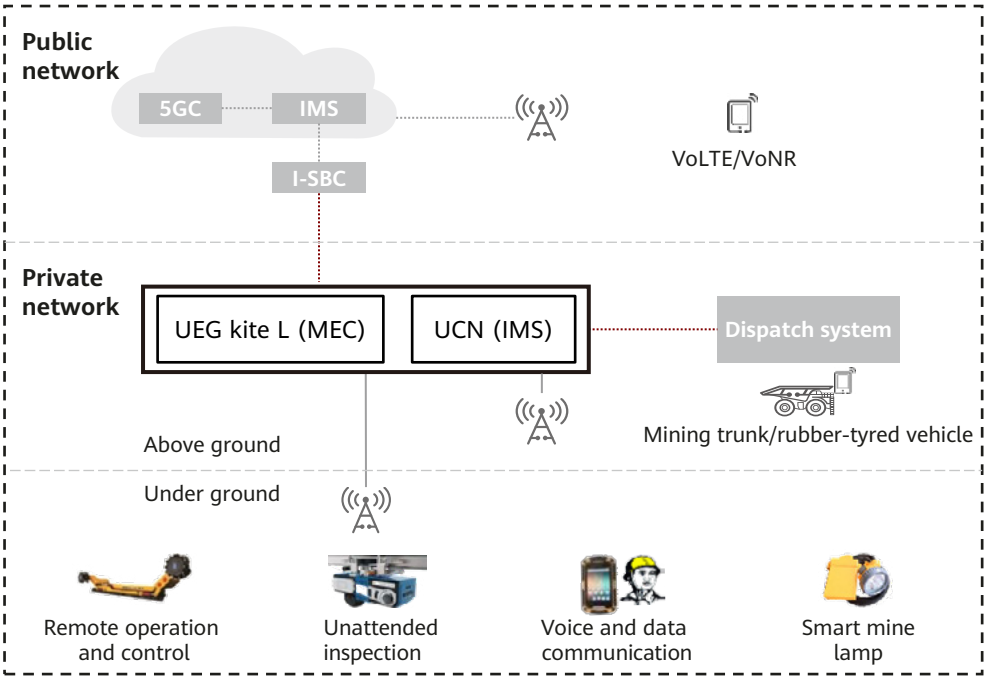


Figure 4 Networking architecture for MEC to mining

With the One Number Converged Network solution, mining employees can use operators' public numbers to make voice and video calls to each other without changing their cards or numbers. This ensures that miners underground can keep in touch with the outside world anytime, anywhere, effectively improving communication efficiency and reducing the dispatching waiting time. This solution provides reliable assurance for mining intelligence, production safety, and mining efficiency.

Large-scale Use and High-quality Development Make a Promising Future

More than 30,000 mining enterprises around the world are seeking for intelligent transformation. Via VoNR, the One Number Converged Network solution will work with growing IMS capabilities and data channel technologies to implement ubiquitous connections and real-time interaction between people and things and between things in the industry. Looking into the future, this solution will support diversified and enhanced voice communication applications, such as remote guidance, AR-assisted maintenance, and push of an emergency rescue route. In addition, this solution deeply integrates advanced technologies with production scenarios to build an organic whole for people-thing interconnection and man-machine collaboration and interaction, incubating integrated and intelligent applications in the mining industry.

2.4 MEC to Manufacturing

More than 1000 MEC sets have been widely deployed throughout industrial manufacturing scenarios, implementing 5G enablement from production assistance to core production. With network slicing and MEC, multiple peripheral auxiliary application scenarios are supported, such as 5G workshop smart surveillance, 5G industrial vision, 5G AR standard operation, 5G AMR-based one-click material transportation, 5G-powered data collection, and 5G-powered action identification. These application scenarios can be found in industries such as home appliance manufacturing. Meanwhile, the application of 5G in the core production phase is also developing, gradually expanding from the northbound PLC application to the southbound PLC application, and entering the industrial control field to meet new network requirements such as high efficiency, high density, high availability, and low tolerance.

One Network for Multiple Purposes Helps the Industry Move from Peripheral Assistance to Core Production

Highly customized factories, represented by the electrical equipment manufacturing industry, pose higher requirements on production line flexibility. By replacing wired connections with 5G connections, the production line can be more fluid in terms of layout, electricity, communication, and process, reducing the cost for equipment investment and maintenance by 80%. PLC can be deployed in a centralized manner based on the 5G private network, with utilization increasing from 20% to over 80%, the footprint of the production line is reduced by 40%, the adjustment time of the production line is shortened by 75%, and the production capacity of the production line is improved by 12%.

In high value-added manufacturing industries such as automobile and medicine manufacturing, automation and robots account for an increasing proportion of the production line. In key production phases, C2IO is widely used in motion scenarios. 60% of I/O devices are continuously moved, twisted, inserted, and removed. As a result, cables are aged or disconnected, causing one or two downtimes per month and 4 hours of production capacity loss on average. 5G reconstruction brings two business benefits. First, one network for multiple purposes improves efficiency, reduces investment, replaces wired networks with wireless networks, and supports real-time (RT) and non-real-time (NRT) services, reducing the cabling time of new factories by 7 to 15 days. Second, it helps improve production capacity. 5G penetrates the field-level OT network of PLC southbound, removing the constraints of traditional wiring and reducing the downtime by 50 hours per year. In this way, devices can remain online for real-time processing, reducing the device adjustment and restart time. The production pace of a single work station is shortened by 500 ms, improving production capacity.

Layer 2 Private Networks Underpin 5G Fully-connected Factories and Build a Converged Network

For industrial manufacturing scenarios, Layer 2 private networks are presented, which simplifies the factory network architecture and provides a converged network solution for fully-connected factories. The solution supports industrial interconnect in all scenarios, such as production assistance, production management (PLC northbound), and production control (PLC southbound).

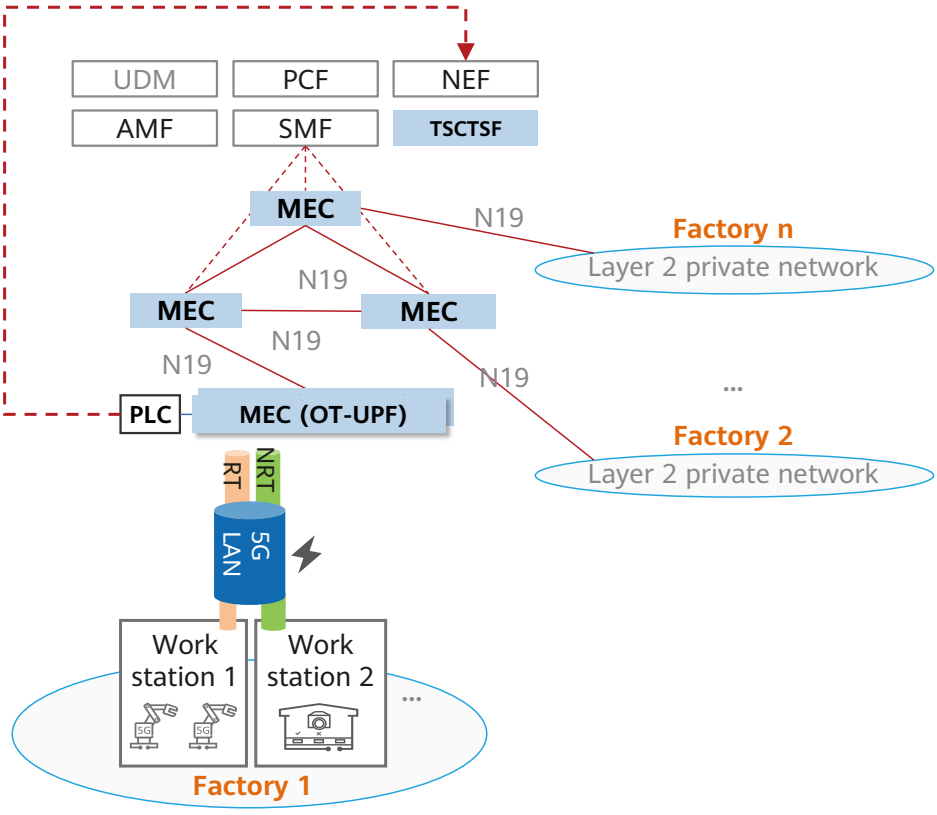


Figure 5 Networking architecture of MEC to manufacturing

5G LAN and 5G LAN QoS enhancement are supported. Different data flow types (NRT and RT) are mapped to different QFIs. Industrial real-time services and non-real-time services can be carried on one network.

XSO cross-layer collaboration is supported to implement bidirectional interaction between the 5G network and industrial devices, dynamic monitoring and awareness of industrial protocols, and interaction of configuration parameters for air interface scheduling. This improves the URLLC scheduling capacity by 3 to 8 times and the determined latency from 99.99% to 99.9999%.

The N19 interface is supported to implement layered wide area interconnection. The N19 interface across SMFs is used to extend the Layer 2 private network in a single campus to multiple campuses, implementing flexible and simplified networking between multiple branches.

Service self-healing networks for high availability are supported. Two UPFs are used for dual fed and selective receiving, achieving E2E high availability of dual fed and selective receiving links, ensuring that no packet loss occurs in case of service faults, and realizing reliable production.

Lightweight UEs Accelerate Project Implementation and 5G Factory Construction Develops on a Large Scale.

It is estimated that the cost of RedCap UEs can be reduced to 20% of that of conventional UEs by 2025, which effectively addresses the pain point of high UE costs in the 5G industry. China's operators and multiple industries, such as industrial manufacturing, smart grid, and electronic manufacturing, have verified RedCap capabilities for commercial purposes. Lightweight RedCap industry UEs will boost the development of 5G industrial interconnect.

For the manufacturing industry, MEC penetrates the core phases in the local area and interconnects with the core network in the wide area, driving intelligent transformation across industries. The 5G fully-connected Layer 2 private network helps continuously implement more than 10,000 projects in advanced manufacturing and high value-added industries, bringing trillion-level economic growth.

Numerous digital transformation opportunities are emerging for 5G, which can be applied across multiple industry sectors. 5G must become an integral part of the production system so that it can serve as a production force and be scaled up in industry applications. MEC serves industries such as mobile office, electric power, mining, and manufacturing. More than 3000 MEC sets have been deployed on a large scale, bringing tangible benefits to industry customers and becoming a powerful tool for operators to develop B2B services.

3

Trend 2

New Technologies Such as RedCap, A-IoT, and ISAC Have Entered the Standardization Phase, Which Heralds the Incubation of Relevant Scenario-based Applications

3.1 Overview

Looking at the current market trends, IoT is more voluminous than Internet of People (IoP) and will be the main driving force for the future growth of connectivity. To make IoT ubiquitous, 5G-Advanced IoT will cover all connectivity scenarios — from high-, medium-, and low-speed to passive connections — comprehensively leveling down the costs of connectivity. 3GPP Releases 17 and 18 have completed the standardization of RedCap, propelling the maturity of the RedCap ecosystem as well as the wide-scale commercial deployment of RedCap. Release 19 is embarking on the standardization of A-IoT.

In terms of developing 5G-Advanced, the industry has proposed ISAC — one of the key enabling technologies behind 5G-Advanced. In addition to inheriting communications functions via the existing 5G air interface technology, 5G-Advanced stands out as a key enabler in wireless signal–based sensing. Specifically, 5G-Advanced is capable of using ISAC to sense the information of target objects or environments and utilizing the information to fulfill sensing functions such as positioning, ranging, speed measurement, imaging, detection, identification, and environment reconstruction, furthering the sensing exploration in the physical world.

3.2 A-IoT

A-IoT covers passive, semi-passive, and active IoT terminals. Compared with existing Narrowband Internet of Things (NB-IoT), A-IoT features lower or even zero power consumption and can combine with cellular networks to enable more IoT scenarios, achieving all-scenario connectivity in industries and opening up a new space for hundreds of billions of connections.

Internet of everything and low power consumption requirements give rise to A-IoT, which predominately covers the following three types of applications:

1. Tag-based connectivity (asset tag identification): Typical scenarios include goods or asset management in the manufacturing and logistics industries.

2. Micro sensor connectivity (sensor data collection): Typical scenarios include wireless

sensor networks in energy, electric power, animal husbandry, and industrial fields.

3. Low-power downlink connectivity (downlink data push): Typical scenarios include applications with Electronic Shelf Labels (ESLs) in the industrial field, supermarkets and retail, and office scenarios.

To meet the demand of passive IoT for ultra-low power consumption and simplified terminals, it is pivotal to simplify the architecture and protocol stack of a 5G-Advanced core network.

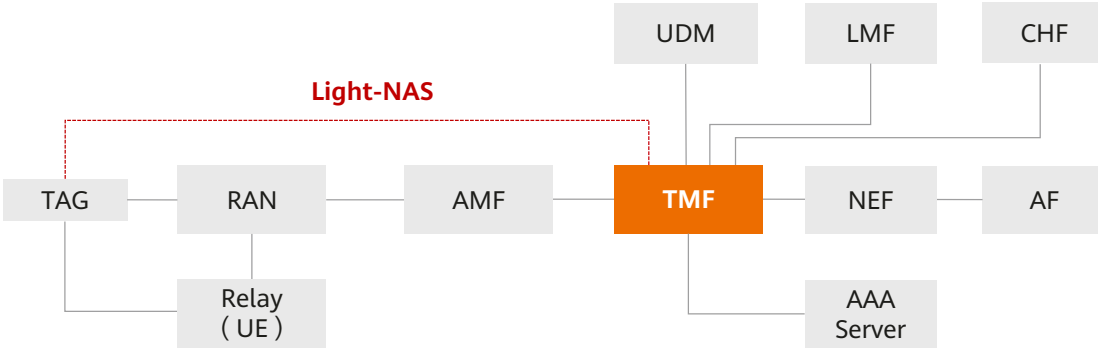


Figure 6 A-IoT architecture

As illustrated in the above figure, the TMF is added onto the 5G-Advanced core network to provide access management and communication based on A-IoT tags. Lightweight NAS protocols are introduced to simplify message structures and IEs. Lightweight tag-based access and mobility management significantly reduce the complexity of terminals. Plus, lightweight, flexible security mechanisms are leveraged to meet differentiated security requirements of various application scenarios and tag capabilities. Users are then able to make a trade-off among terminal complexity, costs, power consumption, and security, addressing the requirements of customers in different industries. What's more, via capability exposure interfaces, network capabilities such as tag management and stocktaking as well as tag-based access and positioning are available. The A-IoT-capable Radio Access Network (RAN) and core network can be deployed together with the existing 5G network, protecting operators' investment in 5G basic networks. In a nutshell, by utilizing the TMF, capability exposure, and tag-based charging, the A-IoT network and services are manageable, controllable, and operable.

3.3 ISAC

ISAC is applicable to wide- and local-area scenarios. Wide-area scenarios include low-altitude inspection with drones, road/railway intrusion detection, environment detection, border defense control, and vehicle-road collaboration. Local-area scenarios include smart home (health detection and home intrusion detection), smart factory (automated guided vehicle detection and tracking in factories), and motion and posture recognition.

To implement ISAC, the SF — managed and controlled by operators — is introduced to the 5G-Advanced core network. It collects data from ISAC-capable base stations, terminals, roadside devices, and among others, and performs intelligent computing based on sensed data.

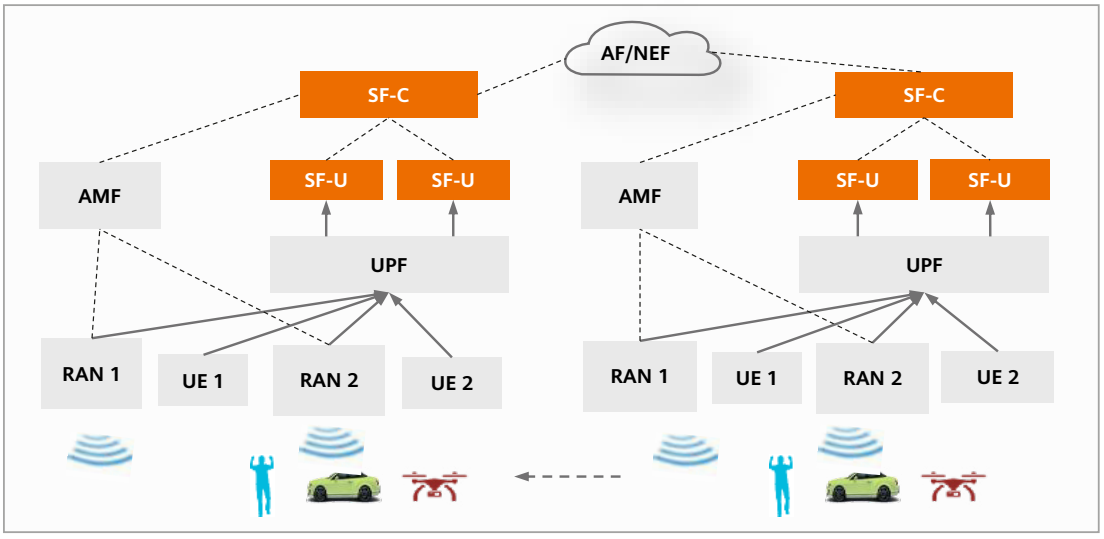


Figure 7 ISAC

ISAC-capable network architecture under Control and User Plane Separation (CUPS)

The control plane sensing function (SF-C) and user plane sensing function (SF-U) are added to the core network. The SF-C interacts with existing 5GC control plane NFs and transmits sensing control information to the RAN and User Equipment (UE) through the Access and Mobility Management Function (AMF), implementing sensing control, such as authentication and authorization on sensing services, and control on sensing QoS, precision, types, and areas, as well as sensing-based charging and sensing capability exposure. In addition, the SF-C is responsible for selecting an SF-U and controlling the

sensed data–based computing on the SF-U. The SF-U collects sensed data from ISAC-capable base stations, terminals, cameras, etc., performs sensed data–based computing, and obtains and reports sensing results. Additionally, the SF-U is able to combine the sensed data from terminals and non-3GPP sensed data (including the data from cameras) with the sensed data from multiple base stations, realizing convergence processing and analysis on multiple types of sensed data.

The sensing operation may involve user privacy security, which requires the core network to anonymize privacy information, for example, performing sensing operations according to users' authorization, controlling sensing areas, and blurring sensed objects.

Based on the sensed data from base stations or terminals and harnessing intelligent computing resources, the SF-U conducts intelligent computing, identifies sensed targets, performs imaging, reconstructs an environment, and more. By leveraging the sensing function, intelligence, and computing, the intelligent computing hardware platform and intelligent software architecture are built for the core network with the ISAC-capable SF-U. This makes value scenarios such as smart lower airspace systems, smart transportation, and digital twin viable..

Reference

[1] Research Report on 5G-Advanced integrated sensing and communication Network Architecture

4

Trend 3

Large-Scale Commercial Use of 5G Accelerates VoLTE Monetization, and VoLTE Is on the Road to Experience Upgrade

4.1 Large-Scale Commercial Use of 5G Accelerates Global VoLTE Development and Monetization

Over the past decade, some operators have deployed all-IP VoLTE networks, upgraded voice service quality, and improved network efficiency and cost-effectiveness. As 5G continues to dominate the telecommunications market and legacy 2G and 3G networks are phased out, VoLTE has become a must-have for operators to provide core services and is imperative to the commercialization of 5G.

In the past ten years, operators deployed VoLTE to improve network efficiency. In the upcoming decade, they will need to leverage VoLTE to commercialize 5G. With the ramp-up of 5G and the gradual phase-out of 2G and 3G, calling services need to be migrated from outdated Circuit Switched (CS) devices to the IMS. The IMS remains the core network for implementing calling services in 5G because VoNR is also based on it. Therefore, operators should deploy VoLTE as early as possible to pave the way for future network transitions while safeguarding voice service quality. According to GSMA Intelligence, by the end of 2022, global VoLTE penetration will reach 46%, and the number of VoLTE-capable terminals will increase significantly. By 2025, the number of global VoLTE and Voice over 5G (Vo5G) connections is expected to reach up to 5 billion, with VoLTE penetration reaching 75%.

More operators will continue to reap benefits from VoLTE, including improved user experience, higher spectral efficiency, and support for 5G evolution.

- **Better experience**

1. VoLTE call connection delay is significantly reduced, taking only 1 to 3 seconds to connect a VoLTE call, which is 2 to 3 times faster than a CS call.

2. VoLTE calls leverage HD audio and video codecs, providing high-fidelity voice services for users.

3. VoLTE users can use data and voice services simultaneously, enhancing their overall service experience. For example, a user can make a call while playing a game or using navigation services. Statistics show that operators providing VoLTE services receive higher scores in P3 network tests.

Voice services pose higher requirements on wireless coverage than data services. Before the widespread commercial use of VoLTE, operators must ensure sufficient 4G signal coverage. To align with operators' VoLTE development and investment pace, Huawei has established network coverage standards for VoLTE in basic, mature, and advanced stages. In 2021, Huawei assisted operator T in country A in building a VoLTE network based on these standards, and operator T successfully put the VoLTE network into commercial use.

Huawei has also established a Key Quality Indicator (KQI) system, assisting operators in establishing a correlation between user experience, network indicators, and NE indicators, encompassing call connection rate, call connection delay, voice quality, and call continuity. This system facilitates operators in optimizing the VoLTE service experience. In 2016, Huawei helped China Mobile significantly improve key KQIs of its VoLTE network, surpassing those of its 2G network, setting the stage for rapid growth of the VoLTE user base and establishing the world's largest VoLTE network.

- **Enhanced spectral efficiency**

The spectral efficiency of LTE networks is 2-3 times that of 2G/3G networks. In other words, LTE networks can support more users with the same spectral bandwidth. For instance, with a 5 MHz spectrum, a 2G cell supports approximately 32 users, a 3G cell supports about 50 users, while an LTE cell supports more than 119 users. As VoLTE can provide higher spectral efficiency, operators aim to migrate their 2G and 3G users to VoLTE networks and refarm spectrum resources to enhance LTE coverage and capacity, thereby promoting the development of both 4G and 5G. According to the Global mobile Suppliers Association (GSA) industry report, 135 operators around the world have either shut down or plan to shut down their 2G and 3G networks, and we anticipate this trend will peak by 2025.

- **Necessity for 5G**

To scale up the commercialization of 5G, a well-constructed VoLTE network must be in place to ensure user experience in voice and data services. The adoption of 5G also boosts VoLTE development. 3GPP mandates that voice calls cannot fall back from 5G to 2G/3G networks. At the same time, some operators are planning to phase out 2G or 3G networks. In this case, VoLTE becomes imperative to ensuring voice continuity for 2G, 3G, 4G, and 5G users.

With the advent of 5G, leading operators still use VoLTE to provide calling services in Non-Standalone (NSA) networking. In SA networking, while options like EPS Fallback (EPS FB) and VoNR are available, VoLTE remains a critical component for EPS FB and an indispensable supplementary to VoNR. According to GSMA, more than 95% of 5G operators have built VoLTE networks.

4.2 VoLTE Is on the Road to Experience Upgrade

In the 4G era, operators' Video over LTE (ViLTE) services were not widely used. However, 5G presents significant opportunities for operators to develop video calling services as essential services. Terminals that support video calls are becoming increasingly commonplace. Moreover, operators can provide video calling services within the same package and at the same price as voice services, allowing users to engage in video calls without worrying about incurring additional fees. These changes lay a solid foundation for the large-scale commercial use of ViLTE or Video over NR (ViNR).

Thanks to the popularization of the Internet, an increasing number of individual users have become accustomed to using video calls for communication. For example, the number of video call users in China exceeds 1.2 billion, with the average Over the Top (OTT) video call duration over the mobile network reaching up to 300 minutes each month. In contrast to video calling services provided by OTT vendors, operators' video calling services offer unique advantages such as negating the need to add friends and install apps, and high-quality calling ensured by dedicated bearers. These advantages undoubtedly present operators with tremendous opportunities.

For example, China Mobile has committed to developing video calling services as part of its basic services. It has set annual video call development goals at the beginning in 2023, aiming to increase the proportion of video calling services to 10% by the end of 2023. As of now, China Mobile has made measurable progress in improving the video call experience, optimizing tariffs, and adjusting packages.

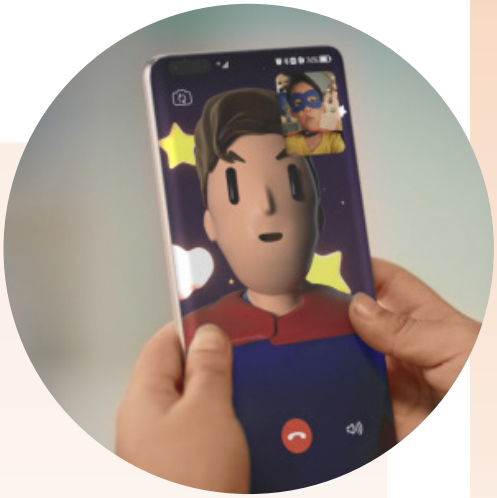
Reference

[2] GSMAi: Building on 10 years of VoLTE

5

Trend 4

Communications Capabilities Further Opened Up; New Calling Becomes Typical Application



Video calls alone can no longer meet people's standards for communication in the digital era. Users have come to expect and seek out experiences that are more interactive and fun. OTT players have made many enhancements to their services, such as multi-party video calling provided by WeChat, background replacement and one-touch beauty supported by MeeTime, and AR emojis provided by TikTok. These new features make calls more enjoyable for the public.

For enterprise users, they expect that the entire transaction process can be seamlessly accomplished with a call, so that their customers do not need to be redirected to third-party applications and jump through extra hoops like registration and authentication. For example, through video calls, car insurance companies can perform damage assessment without an onsite inspection, and device vendors can provide remote installation and maintenance guidance, helping them reduce costs and improve efficiency.

To meet users' requirements and break through conventional calling service limitations, Huawei collaborated with leading operators in China and launched brand-new 5G products. The enterprise aims to empower communication with Ultra-High Definition (UHD), intelligent, and interactive calling. It seeks to help operators transform their business model from voice-only operations to video and content operations in order to take them to greater heights of business success.

5.1 New Calling Is Moving from Concept to Commercial Use

As the first mover in New Calling services, China Mobile has proposed New Calling as one of its main strategic products in April 2022. In January 2023, China Mobile launched beta tests and commercial trials in Zhejiang, Jiangsu, and Guangdong provinces for fun calling, visualized voice calling, speech-to-text conversion, and Chinese-English translation. After commercial trials and several rounds of improvements, at PT Expo China 2023, China Mobile announced that it will complete the construction of a New Calling network that can serve tens of millions of users by the end of 2023. This network is expected to be put into commercial use within the year.

At present, New Calling services like fun calling, visualized voice calling, and real-time translation have already been brought to the market. Fun calling allows users to change their background and send dynamic emojis by issuing voice commands during calls. For

example, if you say "Happy New Year", a Happy New Year animation will pop up on both ends of the call. This service enables people to express themselves in a new and dynamic way.

Visualized voice calling allows users to present a preset short video to the other party during a voice call. Enterprises can also customize their brand promotion materials. In this way, the call screen turns into a new media platform (similar to video ringback tones), providing opportunities for individuals to express their personalities and enterprises to convey their corporate concepts. With services like these, operators can implement content operations and venture into new avenues of monetization.

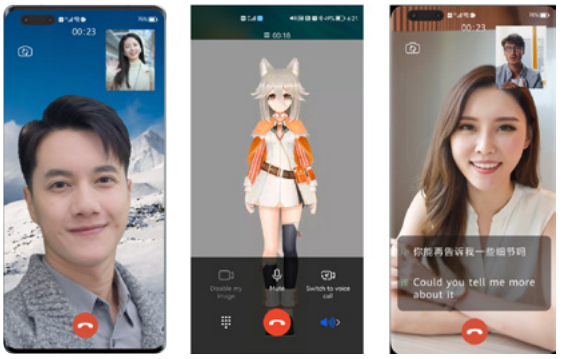


Figure 8 Fun calling/Visualized voice calling/Real-time translation

Real-time translation recognizes speech, translates what the other user is saying, and then converts it into text, overcoming language barriers. In addition, speech-to-text conversion and large fonts are also helpful for the elderly and people who are hard-of-hearing.

In addition to China Mobile, other operators are also rolling out the commercial use of New Calling on huge scales. In the second quarter of 2023, China Broadnet and China Telecom made their first call using New Calling in some pilot provinces.

Internationally, GSMA established the New Calling Foundry project at the New Calling Industry Development Forum during MWC Barcelona 2023. This project attracted the attention of multiple telecom industry partners, including the likes of China Mobile, Kuwait Zain, and Huawei. In October 2023, GSMA will release a New Calling white paper at the Network X New Calling Summit to demonstrate the progress of the New Calling Foundry project. In 2023, some European and Asia-Pacific operators began Proof of Concept (PoC) tests on New Calling. Operator A from Southeast Asia took the lead in completing the PoC verification of New Calling this September. 2023 is seeing the emergence of New Calling in markets outside China. New Calling is no longer a mere concept but has become a reality.

5.2 New Calling Becomes a Benchmark for Network Capability Exposure

NaaS is not a new topic. Operators have been exploring network capability exposure and monetization for many years. Typical examples are short messages and voice calls. For example, when a courier dials the recipient's virtual phone number through an app, it invokes the voice call capability provided by operators, and operators charge the app service provider based on the number of times it is invoked. Operators have achieved great success in calling capability exposure. In China's telecom market, more than 95% of revenue comes from calling capability exposure.

Based on the underlying voice network, New Calling delivers UHD, intelligent, and interactive calling capabilities to revolutionize traditional calling with new services and functions, such as emojis, subtitles, translation, XR, and AI. It enables individuals and enterprises to enjoy fun, real-time, and interactive calling services, and provides more possibilities for operators to widen the scope of network capabilities and discover even more business opportunities. Operators have positioned New Calling as a platform product and will open the platform to numerous developers and industry partners so that they can explore more new service scenarios, enrich digital content, upgrade user experiences, and unlock new potential within the market.

In the past, only basic voice and video calling were opened up to external systems. By introducing the IMS Data Channel (DC), New Calling can open up data interaction, XR, AI, and more on top of traditional calling capabilities. DC-based New Calling enables users to use new features like e-signing and transferring images and location information with a simple touch on a screen. Now, users can enjoy OTT-like service experiences during native phone calls. With New Calling, operators can tailor services for different industries based on their distinctive communication requirements.

Once the New Calling network capability is opened up to enterprises, they can invoke video calling and DC interaction to facilitate their business, including for remote maintenance and remote damage assessment. Operators can charge these enterprises based on the number of times they are invoked. So far, Huawei has worked with China Mobile and Jingyou Technology to apply New Calling to remote vehicle damage assessment. When a vehicle collision occurs, the car owner can make a video call to the insurance company. The insurance company then establishes a unidirectional DC, guides

the car owner to take photos with the help of AR annotation, and finally generates a quotation. In this way, a claim can be processed through a simple call. Onsite inspection can be avoided and the damage assessment time is shortened from two hours to five minutes, greatly reducing costs and improving efficiency. Currently, due to the limited use of DC terminals, unidirectional interaction is generally implemented by means of establishing a unidirectional DC, which restricts the development of interactive New Calling applications. As of 2024, native DC terminals will be gradually put into commercial use, and it is predicted that in the foreseeable future, interactive New Calling applications will mushroom in a wide range of vertical sectors. For example, the on-screen menu function can replace traditional Interactive Voice Response (IVR) systems currently used in sectors like banks and hotels, which will bring operators hundreds of millions in the market space. This is the first phase of New Calling capability exposure.

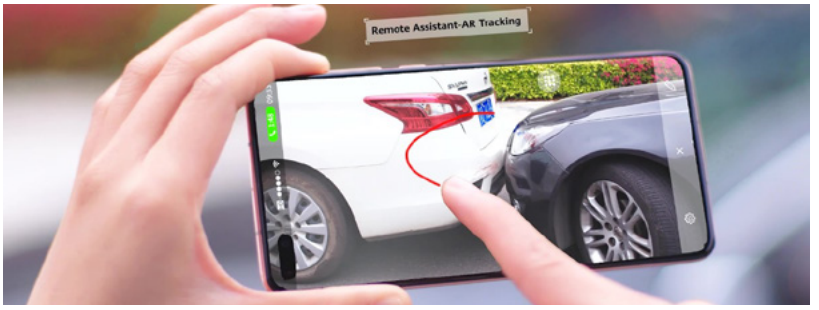


Figure 9 Remote assistance

Looking into the future, the New Calling network will integrate AIGC and multi-modal communication capabilities. This will give rise to more New Calling service forms and unlock tremendous space for network capability exposure. For example, by utilizing AIGC, users will be able to create virtual avatars that can even mimic their facial expressions. During calls, users can show these avatars to the person on the other end of the line to express themselves. In the future, operators will be able to provide an intelligent digital assistant service for enterprises to create digital characters that fit their own industry style and corporate culture. When an enterprise user makes a video call with a customer, the intelligent digital assistant can change the video style based on the company's curated branding. It will be able to match the dress of the digital character to the type of customer it's serving, and can even provide simultaneous interpretation during cross-language communication. The AIGC foundation model alongside multi-modal communication will elevate New Calling capability exposure to a whole new level. It will transition human-to-human communication to human-to-machine, and even machine-to-machine communication, revolutionizing the entire communications industry.

6

Trend 5

Glasses-free 3D Is Boosting Mobile Media Traffic; Ultra-Large Traffic Processing and Intelligent Experience Assurance Are Becoming Trending Topics

Driven by new technologies such as metaverse and AIGC, video services are experiencing tremendous changes in content production, transmission, and display. These technologies are redefining people to people interactions and people to things interactions, while at the same time bringing new challenges to the network.

6.1 Spatial Video Services Like Glasses-free 3D and XR Have Become the Main Trend

Reproducing the real world through images and videos has always been generally pursued by humans, and many people have dedicated great efforts to advance such approaches. Among all these efforts, the idea that binocular disparity can contribute to the stereo visual effect was proposed by Sir Charles Wheatstone as early as 1830, marking the initial start of spatial video technologies such as 3D and VR. However, the commercial use of this idea has been curtailed by the insufficient development of technologies for a long time.

But now, the breakthrough in the glasses-free 3D technology provides an opportunity for spatial video services to grow exponentially. It is worth noting that all the other sectors for 3D video services are also ready. In terms of devices, the screen that natively supports glasses-free 3D display has been put into commercial use, which is affordable and allows a flexible switchover between 2D and 3D display. The screen can also collaborate with eye-tracking analysis powered by AI technologies, so that the display result can be adjusted intelligently and the field of view is expanded for a greater experience. In terms of content production, solutions such as Huawei's Vivision can be used to convert existing 2D content and even 2D live streaming to 3D, solving the problem of insufficient 3D content. In terms of network, the binocular 3D content requires less than twice the bandwidth of 2D videos, which has been fulfilled by the current network.

With all these technological advancements, the glasses-free 3D video service has been streamlined from end to end. The service is now stepping into a rapid growth phase, and will promote the development of content, application, and technologies related to all spatial video services.

Item	Past	Present
3D content	3D video content is insufficient.	With AI-based automatic transcoding technology, massive 2D videos and 2D live streaming can be converted into 3D in real time.
Device costs	Professional 3D display devices are costly.	Multiple 3D-capable tablets have been launched at a price lower than US$500.
Viewing experience	Low resolution and fixed angle of view cause poor experiences.	High-resolution display devices and intelligent eye tracking technologies deliver optimal experiences.
Compatibility	2D and 3D display effects cannot be switched flexibly, affecting 2D video viewing experiences.	2D and 3D display effects are able to be switched at any time on the screen empowered with the dynamic control technology.

Table 2 Rapid development of the glasses-free 3D industry

XR is also entering a fast development phase as 5G expands into more scenarios and new technologies such as AIGC emerge. According to Precedence Research, the XR sector has seen a meteoric rise in adoption rates across industry verticals, and the education, entertainment, gaming, and training sectors have seen unprecedented growth in XR adoption. Precedence Research also notes that the Compound Annual Growth Rate (CAGR) of XR will reach 33.09% from 2022 to 2030 and consumers will collectively use 70 million VR headsets worldwide by 2026. Goldman Sachs states that VR and AR will be a $35 billion annual business by 2025.

As spatial video and HD video services develop and become popular, they will pose higher requirements on media content encoding and transmission.

6.2 Requirements on Downlink Bandwidth

Spatial video services such as glasses-free 3D pose great challenges on the network bandwidth.

The glasses-free 3D display can be achieved with binocular parallax or light field technology. The binocular parallax approach divides a video stream into two streams: one for the left eye and one for the right eye. The two streams will be respectively

projected to the eyes through the parallax barrier and lenticular lens, so that the stereoscopic effect can be perceived by the viewer with binocular disparity. This approach requires one or two times higher bandwidth than 2D videos, which can be fulfilled by the current network. However, this approach can only provide an entry-level 3D experience, and there are still several problems to be solved:

- Vergence-Accommodation Conflict (VAC): It is a visual phenomenon that occurs when the brain receives mismatching cues between vergence and accommodation of the eye. This effect is unpleasant and causes eye strain and dizziness.

- Limited viewpoint: The glasses-free 3D effect achieved with binocular parallax technology is the same regardless of the view angle. The 3D effect feels different from that in the real world.

- Inapplicability for multi-player viewing: As for the 3D effect achieved with binocular parallax, eye-tracking technology needs to be used to adjust the display of the two video streams. However, this mechanism does not apply to multiple viewers.

To solve the preceding problems, light field technology is proposed, which divides a screen into multiple viewpoint sections and allows viewers to see different images from different viewpoints. The glasses-free 3D effect provided by light field technology is more natural and comfortable, and can be experienced by multiple viewers at the same time.

A light-field 3D video superimposes the images of different viewpoints into each frame, which causes a much larger bitrate than binocular 3D videos. In order to achieve smoother 3D video displays, more viewpoints are required with a higher bitrate, which in turn requires much higher bandwidth for transmission.

Media Content Resolution	**Number of Viewpoints**	**Bandwidth Requirements**		
		No Compression	**H.264 (100x Compression)**	**H.265 (350x Compression)**
3840 x 2160 60 fps	30	333.7 Gbit/s	3.3 Gbit/s	976 Mbit/s
	60	667.4 Gbit/s	6.7 Gbit/s	1952 Gbit/s

Table 3 High requirements of 4K videos on network bandwidth

As shown in the preceding table, the bandwidth required for multi-viewpoint light-field 3D videos reaches 6.7 Gbit/s. It is a great challenge for the network to accommodate 3D content and guarantee optimal 3D viewing experiences. Similarly, as XR services grow, they are also bringing large traffic to the network, which requires efficient processing.

6.3 Requirements on Uplink Bandwidth Guarantee

The live streaming market has experienced remarkable growth in recent years. According to statistics, there are more than 7000 live streaming operation companies and over 60 million live streamers with 10K+ followers in China, producing a GMV of US$1800 billion.

Various live streaming activities today require a much higher uplink bandwidth than before, which allows multiple-device live streaming with a larger resolution and greater stability. In addition, individual users today are more and more willing to upload and share content, instead of simply viewing videos. This also demands a faster uplink speed and poses great challenges on the network, especially in scenarios such as concerts and sports events, which tend to attract large crowds.

The resolution and frame rate of live streaming are two important factors relating to bandwidth requirements. Resolution refers to the definition of a video, which is represented by pixel values, such as 720p, 1080p, 2K, and 4K. The frame rate refers to the number of images contained in a video per second, which is represented by Frames Per Second (FPS). A video must have a frame rate of at least 30 fps to prevent frame freezing. The following figure shows the uplink bandwidth required for different resolutions with a frame rate fixed at 30 fps.

Resolution	720p	1080p	2K	4K
Uplink Bandwidth Requirement	3 Mbit/s	5 Mbit/s	10 Mbit/s	20 Mbit/s

Table 4 Uplink bandwidth requirements of videos with different resolutions

If the video frame rate increases to 60 fps, the video experience will be much smoother, but the required uplink bandwidth will double. The application of the VR display on mobile phones further lifts the requirements for uplink bandwidth. For instance, TikTok introduced the VR live streaming function on mobile phones in September 2022, and this

requires a three to five times higher uplink bandwidth than traditional 2D live streaming.

The network is vital to live streaming, especially to outdoor streaming and walk streaming. In these two scenarios, the streamer may have to move to densely populated areas, where uplink network frame freezing is very likely to occur. Once frame freezing occurs, the streamers may lose viewers, followers, and even deals, causing severe impacts to their business and personal brands. To ensure a good live streaming experience, about 40% of the streamers choose to pay for live streaming service packages and 5% of the total pay over US$150 per month. Optimal uplink bandwidth guarantee is urgently needed in this field.

In addition to personal live streaming, event live broadcasting is also catching on and posing new requirements on the network. For example, for 4K multi-viewpoint interactive live broadcasting of a large-scale sports event, the 5G network can meet the network speed requirement of a single 4K 50 fps camera, that is, 35 Mbit/s to 50 Mbit/s, but it cannot cope if parallel transmission of video streams from multiple cameras is required. In this case, special uplink assurance is required for the 5G network at the shooting site, to guarantee not only an N times higher uplink bandwidth, but also the seamless video synchronization for viewpoint switching.

XR is another key service requiring uplink experience assurance. There are several typical XR services including XR video live streaming (360-degree VR and free-viewpoint), spatial video live streaming, and cloud-based 3D reconstruction. The following figure shows the uplink bandwidth requirements of these services:

Application	**Device Form**	**Resolution**	**Frame Rate**	**Uplink Rate Requirement**
VR video capture	Customer-Premises Equipment (CPE) + 360-degree VR Pan-Tilt-Zoom (PTZ) camera	VR PTZ camera: 4K–8K	30 fps	360-degree VR live streaming: 45 Mbit/s to 75 Mbit/s
Free-viewpoint video capture	Outside-in camera array	Single camera: 1080p	30 fps	Free-Viewpoint Video (FVV): 9 Mbit/s to 45 Mbit/s
Cloud-based 3D reconstruction service	Smartphone	1080p to 2K	-	4 Mbit/s to 24 Mbit/s
Real-time cloud-based 3D reconstruction service	XR headset	720p+depth	-	8 Mbit/s to 24 Mbit/s

Application	Device Form	Resolution	Frame Rate	Uplink Rate Requirement
Live streaming spatial video capture	Outside-in camera array	1080p+depth	30 fps	24 Mbit/s
Real-time cloud-based 3D reconstruction service	XR headset	2K+depth	-	32 Mbit/s to 96 Mbit/s
Live streaming spatial video capture	Outside-in camera array	1080p+depth	60 fps	48 Mbit/s
Real-time cloud-based 3D reconstruction service	XR headset	4K+depth	-	64 Mbit/s to 192 Mbit/s

Table 5 Uplink bandwidth requirements of typical services

In addition to the above services, many application scenarios such as video conferencing, gaming, and large file transfer also require uplink experience assurance. Generally, uplink experience assurance has become a necessary trend for core network development. However, the total bandwidth resources on the network are limited, and different users and services have different requirements on uplink bandwidth. Therefore, to implement differentiated uplink experience assurance, it is fundamental to efficiently utilize network resources.

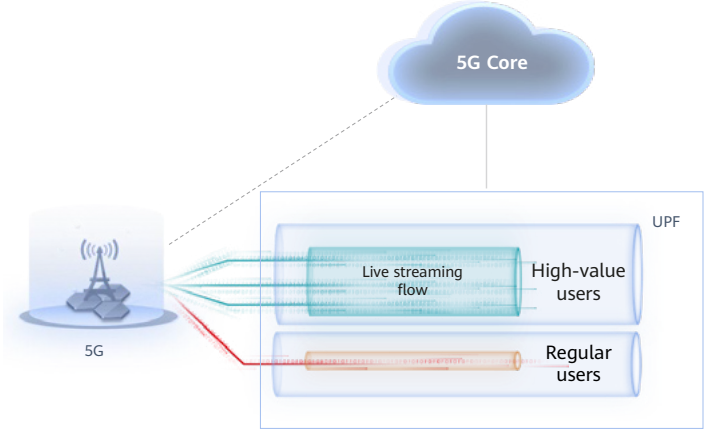


Figure 10 Uplink bandwidth assurance

6.4 New Media Transmission Technology – MoQ

Various media services have flourished on the Internet during the past 20 years, which has brought different media transmission technologies into being, for example, Web

Real-Time Communication (WebRTC) used for video conferences, Dynamic Adaptive Streaming over HTTP (DASH) and HTTP Live Streaming (HLS) used for long video on demand and live streaming services, and the Real-Time Messaging Protocol (RTMP) used for uplink live streaming transmission. These transmission technologies make media services quite complicated, especially for mixed media streams. As such, there are many enterprises, such as Google, working on converging different media transmission protocols. This is also becoming a trending topic.

On a slightly different topic, content encryption is becoming more and more common, out of copyright and privacy protection requirements. Statistics have shown that encrypted traffic accounts for more than 90% of total traffic on the Internet. However, current transmission protocols support only server-side encryption and device-side encryption, and the transmission on the network is transparent. In this way, the transmission security and reliability on the network cannot be guaranteed.

To solve the problems, the Internet Engineering Task Force (IETF) proposed the MoQ media protocol to integrate the media transmission protocols. This new protocol also introduces reliable relay nodes into its architecture, which allow service-network collaboration and dynamic service experience assurance based on network conditions and service requirements.

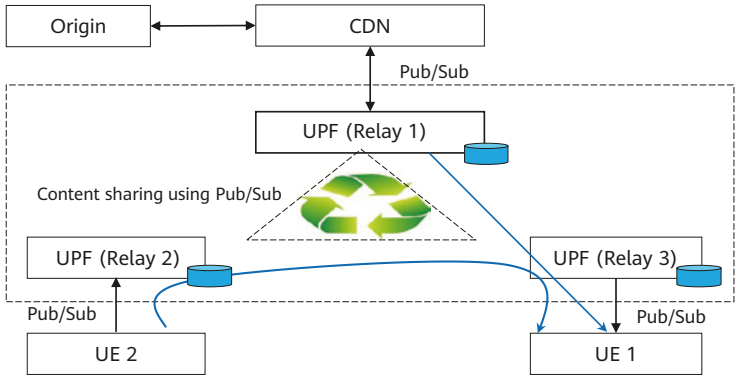


Figure 11 MoQ content distribution architecture

In the MoQ architecture, the MoQ media streams sent by the media source (for example, the XR server) to a relay node contain not only the media content but also the metadata that identifies the media attribute. The relay node detects the network congestion status and performs priority scheduling, caching, dependency, or selective discarding based on the media content attribute to optimize the stream transmission and ensure an optimal video experience. The MoQ is also incorporated into 3GPP specifications on XRM for further development.

6.5 5.5G Core Network Technology Trends for Optimal Video Experiences

To meet the preceding requirements for media services, the 5.5G core network is evolving in the following aspects:

- From packet-level scheduling to frame-level scheduling: For media services such as spatial video and HD video, key streams and non-key streams are encoded separately. These different services are given different QoS priorities, based on which the key streams are preferentially scheduled for processing, so as to ensure optimal experiences for high-bandwidth-demanding media content.

- From static QoS guarantee to dynamic intelligent guarantee: 3GPP defined the Network Data Analytics Function (NWDAF) to introduce intelligent capabilities to networks. With the NWDAF, intelligent service assurance can be implemented through intelligent identification and closed-loop control, and user experience is estimated in real-time and dynamic optimization is performed accordingly.

- From single-user high bandwidth processing to intelligent offloading for heavy traffic brought by highly concurrent connections: With the popularization of high-bandwidth-demanding media services such as glasses-free 3D and XR, the capabilities for processing high concurrency and large traffic is expected from the network. The UPF will be enhanced for that. Leveraging optimized forwarding algorithms and technologies such as intelligent NIC offloading, the enhanced UPF can ensure high bandwidth for services in high concurrency scenarios. In addition, to adapt to the disparity of media services, the network processing capability topology of the 5.5G core network needs to be maintained and ultra-large service traffic can be scheduled to dedicated nodes through intelligent route selection, to improve the processing efficiency.

- From separated computing and storage to integrated computing and storage in distribution: In order to reduce the service latency and repeated forwarding of video streams between devices and the cloud, native storage and computing capabilities need to be supported on the network. In this way, media rendering can be implemented at the network edge, which is much closer to users and can greatly improve service experiences.

7

Trend 6

5G Core Networks Are 100% Deployed over Telco Cloud, Cloudification Becomes the Leading Network Construction Model, and Future-oriented Convergence & Simplification, Standardization, and Automation Become the Directions

7.1 Cloudification Becomes the Leading Network Construction Model, and 5G Core Networks Are 100% Deployed Over Cloud

Since the Network Functions Virtualization (NFV) concept emerged in 2012, Telco Cloud has experienced a decade of development, along with all the challenges in technologies, organizations, cultures, and processes. So far, 5G core networks are 100% deployed over Telco Cloud. According to ETSI's global operator survey in September 2021, the expected NFV values, such as pooled resources, distributed software, and automated O&M, have been basically realized. Over 200,000 servers have been delivered worldwide. In China, there are more than 660 million 5G connections, which are all deployed over Telco Cloud. It is estimated that by 2028, with the sunsetting of legacy devices, all core networks will be cloudified.

To date, numerous achievements have been made with cloudification. For example, operators, standards organizations, and device vendors have made joint efforts in building an eight-level disaster recovery system and proposed the rock-solid reliability solution for cloudified networks, implementing 99.999% reliability on Commercial Off-The-Shelf (COTS) hardware and realizing the world's first cross-DC DR switchover for tens of millions of subscribers. Based on virtualization technologies, carrier-grade enhancements, such as CPU core pinning, Single-Root I/O Virtualization (SR-IOV), Data Plane Development Kit (DPDK), huge page memory, and OpenStack large-scale scheduling optimizations have been made to minimize virtualization overhead and to promote large-scale deployment of Telco Cloud.

Openness is also a development direction of Telco Cloud. Software and hardware decoupling has been put into widescale commercial use. Interconnections based on ETSI specifications have matured and are stable. More than 70% of operators believe that NFV helps avoid vendor lock-in. New hardware forms, such as heterogeneous compute capabilities (a cloud platform supports both x86 and Arm servers), have been put into commercial use on live networks in the China region.

7.2 Deployment of Public Clouds in the Telecom Industry Faces Uncertainties and Carries Multiple Challenges

In the past year, telecom networks deployed based on public clouds were still rare. Telecom services differ significantly from common IT applications:

The public cloud deployment architecture and network topology cannot meet carrier-grade reliability requirements. For example, a region-level fault of a public cloud (more than 40 such faults were reported last year) may spread in a multi-AZ environment. At least two regions are required to realize disaster recovery (DR). However, except for the US, few countries are able to deploy a public cloud in multiple regions. Although the on-premises deployment model is proposed in other solutions to mitigate fault spreading, this model makes a public cloud masquerade as a private cloud to some extent. The cost of the on-premises public cloud solution is 40% to 50% higher than that of the region-based public cloud solution (due to new exclusive hardware and additional equipment room footprint). In this case, public clouds will no longer advance in intensive sharing.

From the perspective of O&M, the Service Level Agreement (SLA) is regarded as the commitment of cloud vendors to their service quality. However, few public cloud vendors can guarantee carrier-grade availability as expected. In most public cloud cases, availability is between 99.5% and 99.99%, which is far from carrier-grade requirements, especially the high stability and reliability requirements of the core network.

After a decade of NFV cloudification practices in the telecom industry, especially in the core network field, IT technologies are integrated into the CT field. Such in-depth integration needs to be adapted again for public clouds and further exploration is required. Public clouds are still facing the challenges of data sovereignty, data security, and multi-vendor interconnection, and have not proved their unique advantages over the existing Telco Cloud. As such, Telco Cloud is still preferred by operators.

7.3 Converged and Simplified Network Deployment in the Future

What concerns operators the most is how to introduce containers to telecom networks. According to the survey conducted by ETSI NFV and CNCF, more than 75% of the feedback received from multiple enterprises indicate that the future application deployment platform will be a hybrid one. That is, the Telco Cloud platform needs to provide VMs, containers, bare metal servers, and other resources. So, how can containers be introduced to telecom networks, by expanding the existing cloud or deploying a new cloud? Striking a balance between the resource management advantages of OpenStack and application-oriented management flexibility (such as using K8s) becomes essential.

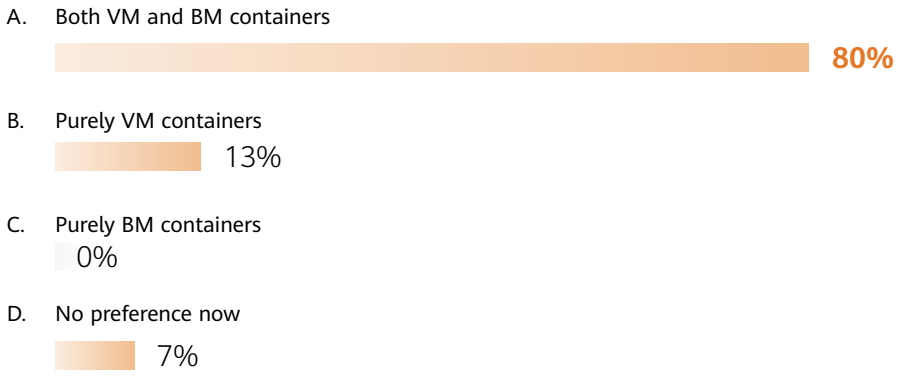


Figure 12 Source: survey result from the ETSI NFV roundtable held at MWC Shanghai (June 2023)

It is believed that the current NFV architecture and inventory networks are optimal, and have been developed after years of adaptation to operators' organizational capabilities, network architectures, and personnel skills. A smooth evolution solution is recommended to better fit operators' evolution path.

Capacity expansion instead of new deployment: On existing networks, the container layer can be smoothly introduced by adding compute resources, so as to provide container resources externally. Deploying an independent container cloud is unnecessary. This will not damage the existing networking architecture. In addition, the cloud-based achievements (such as intensification and software and hardware decoupling) will be reserved, and the existing interface interconnections can be inherited.

Optimal TCO: After new container capabilities become available thanks to capacity expansion, compute servers, storage devices, and network devices can be shared. The network architecture and networking model do not need to be adjusted, additional interconnection and verification are not required, and what's more, investment in new infrastructure is also not required. Only one cloud needs to be maintained. In this way, multiple resources, such as VMs, containers, and bare metal servers, can be flexibly scheduled through one portal. This simplified O&M model greatly reduces the TCO for operators.

Currently, operators such as China Mobile and Saudi Arabia STC are actively deploying this solution on the live network, which has become a new trend.

7.4 Intent-Driven Approaches to Telco Cloud Native Automation, Leveraging Cloudification Benefits

How can complex telecom networks be efficiently operated and maintained? The answer lies in minimizing manual intervention and complexity, and avoiding manual errors. In this way, the value of automation can be brought into full play.

The O&M of a telecom network requires a holistic approach, covering every component and NF. For example, expanding capacity for an NF on a telecom network also requires the automatic association of peripheral components. Additionally, deploying a function is not just about the deployment of clusters and single NFs, cross-NF association is also a must to meet the requirements of telecom services such as 5G slicing and private line provisioning.

The first breakthrough point for simplified O&M comes from removing the complexities of the current telecom cloud configuration and minimizing manual intervention.

1. Simplified configuration parameters: Intent-driven design converts configuration design into network operations to automatically complete network design and configuration, reducing the parameter configuration workload by 90%.

2. Interface standardization: Northbound interfaces are standardized to shield cross-

vendor implementation differences, which facilitates interconnection and interworking between systems from different vendors.

3. Sandbox pre-verification: Through configuration dry-run and simulation verification, configuration errors are intercepted in advance to minimize the risks brought by network operations to the live network.

Another objective is to achieve full-scenario automation for the telecom ecosystem, including intent-based health check, automatic test, and online attendance based on intent-related data analytics.

The ultimate goal is to match the operators' organization and capability requirements. A well-designed autonomous system could help address the complexity of the interworking between components, and thus ease the work of human operators. Operator maintenance personnel focus on service and network O&M. For product O&M, AI can be introduced to achieve self-planning and self-optimization for closed-loop operation.

7.5 Cloudification Specifications Complying with ETSI Standards and Supporting Smooth Evolution

Thanks to the joint efforts of industry partners, the past decade has witnessed the growth of NFV. Best practices suitable for operators' organizational capabilities, network architecture, and personnel skills have since been explored and refined. NFV's success also relies on the ETSI NFV architecture and a large number of interoperability verifications and certifications.

It is becoming apparent that technical divergences and new siloed architectures are emerging due to the increasing number of cloud protocol stacks and there is growing complexity brought by the technology stacks, such as ONAP, Nephio, Sylva, and various public clouds.

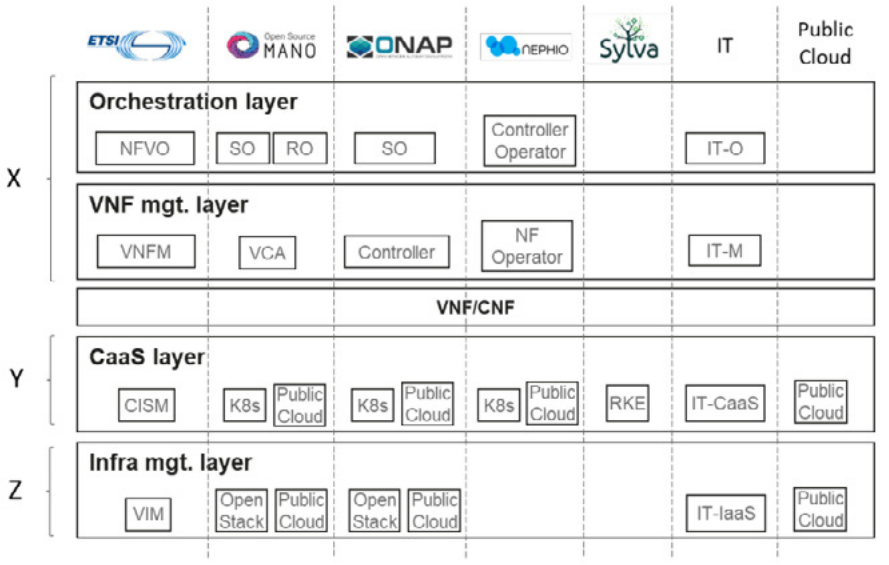


Figure 13 The choices of cloud native implementations for telecom network operators

With the evolution of containers, introduction of new hardware, and new automation requirements, joint efforts among all industry players are still needed for advancing ETSI NFV standards and preventing operators from falling into new vendor lock-in due to new technology silos.

To cope with the growing complexity and to simplify network operations, standardized interfaces are required for southbound and northbound communication. Based on the use of commercial off-the-shelf (COTS) hardware, its drivers should be standardized and open to reduce the complexity of multi-vendor integration and reduce the workload of development and testing as a means to facilitate interoperability. In this way, the cloud platform can quickly integrate and interconnect with various hardware. ETSI NFV standards and solutions mainly focused on MANO's core capabilities such as lifecycle management. In Q2 2023, MANO evolution was incorporated in NFV's Release 6 Definition with the focus on enhancing data-based, intelligent O&M.

8

Trend 7

Foundation Models Are Turbocharging AI Progress, Indispensable for 5.5G Core to Achieve Intelligent O&M, Services, and Networks

8.1 Foundation Models Are Turbocharging AI Progress, Helping 5.5G IntelligentCore Achieve Comprehensive Intelligent Transformation

2022 saw incredible growth in foundation models — AI models trained on a massive scale, and the AI industry is ushering in a period of explosive growth. ChatGPT, which stands for Chat Generative Pre-trained Transformer, burst onto the scene in November 2022, ushering in an explosion of generative AI. With the number of users exceeding one million in just five days, it became the fastest-growing consumer application, reaching an estimated 100 million monthly active users within its first two months. Why did ChatGPT go viral? It went viral because it can mimic a human-like conversation based on user prompts and provide a detailed response. In March 2023, the release of Generative Pre-trained Transformer 4 (GPT-4) further demonstrated the strong potential of foundation models. GPT-4 exhibits human-level performance on various professional and academic benchmarks. Given the breadth and depth of GPT-4's capabilities, Microsoft believes that it could reasonably be deemed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system.

Based on the preceding analysis, we believe that the major breakthrough made by the foundation model technology will completely transform the AI supply model and it will make it easier for the supply side to provide different services. AI technology is about to be applied in industries on a large scale. AI has progressed from being able to process basic logic and lists of information (AI 1.0) to AI 2.0 represented by foundation models.

In the 5.5G era, Cloud Native and Full Convergence are advancing constantly, greatly increasing the scale and complexity of the cloud-based core network. In addition, service innovation and network optimization need to be enhanced on the core network to continuously improve user experience.

As shown in the following figure, powered by AI foundation models, the 5.5G core is positioned to act as a comprehensive transformation model for intelligent O&M, services, and networks, enabling new service experience and better networks. With network capability openness, NaaS business opportunities can be fully activated.

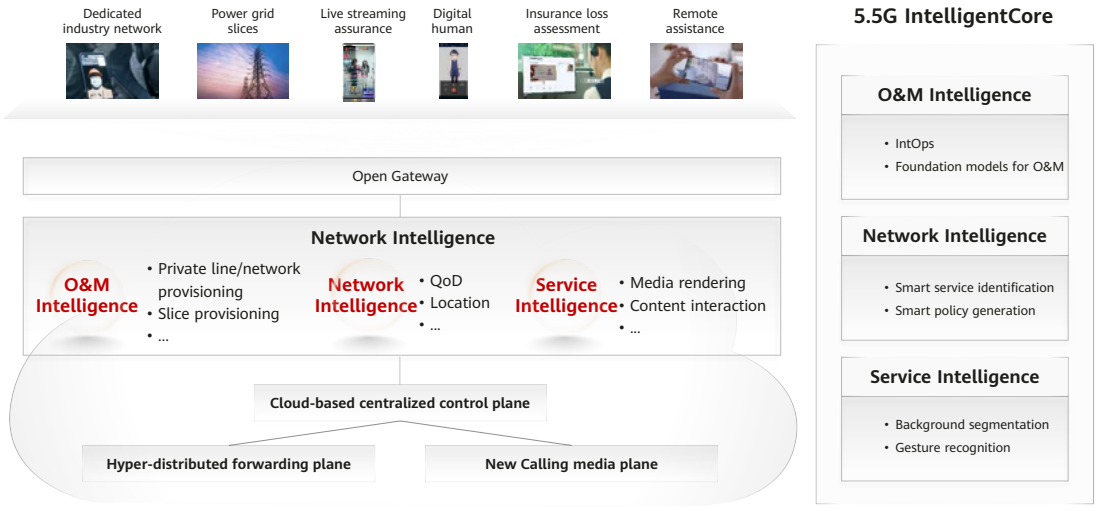


Figure 14 5.5G IntelligentCore

8.2 Intelligent O&M with ICN Master Optimizes Core Network O&M Efficiency for a High-Stability Network with Better User Experience

Industry-leading ICT companies accelerate the introduction of Copilot into their mainstream products, greatly improving product utilization. Microsoft has built-in Copilot in Power Platform, Microsoft 365, Dynamics 365, and Visual Studio 2022. To benchmark with Microsoft Copilot, Google launched Project IDX, an experimental new initiative aimed at bringing the app development workflow to the cloud.

Copilot also needs to be urgently introduced to the core network, to help O&M personnel improve efficiency. Based on the operators' O&M workflow, Huawei introduces ICN Master and suggests that foundation models be built in phases. In the first phase, focus on semantic understanding and API query capabilities for monitoring, troubleshooting, and complaint handling to reduce the number of invoking and switching times between multiple systems. In the second phase, make breakthroughs in advanced service scenarios such as root cause analysis of complex faults.

The following illustrates the application of ICN Master in monitoring and troubleshooting and in routine changes.

· **Monitoring and troubleshooting**

As for troubleshooting, operators care about how long it takes to handle alarms and whether a work order is processed in time. ICN Master can function as the following two roles:

-Alarm analysis assistant

Alarm handling on the live network tends to be time-consuming and labor-intensive. To solve this problem, ICN Master is developed with three key capabilities:

1. Generating the alarm description, alarm impact, alarm handling suggestions, and similar cases based on the alarm information;

2. Aggregating and counting alarms based on the alarm generation time and the NF and device information.

3. Providing detailed alarm handling methods based on the obtained information (through orchestrated commands) and the auxiliary O&M tool.

-Signaling analysis assistant

Typically, operators perform 24/7 dialing tests on live-network services. Once an exception occurs, a work order is triggered and manual analysis is required, resulting in low processing efficiency. In this context, ICN Master steps in to provide two key capabilities:

1. Generating root causes as well as recommended solutions for the dialing test failures. To achieve this, ICN Master performs dialing tests on the failed numbers again, records related signaling, and analyzes the signaling flows.

2. Identifying the root cause of service failures that trigger a user complaint by performing a simulated dialing test and analyzing the signaling flows.

· **Routine Changes**

During routine changes, operators pay close attention to the success rate of site data

configuration. However, there are several obstacles hindering the configuration efficiency and effectiveness, such as deficient automation in script generation, poor adaptability for various scenarios, and intensive manual intervention. After the configuration scripts are generated, manual review is still required. Against this backdrop, ICN Master, tapping Natural Language Interaction (NLI), is introduced to automatically generate data configuration scripts and verify the correctness through simulation for specific NFs and network configurations.

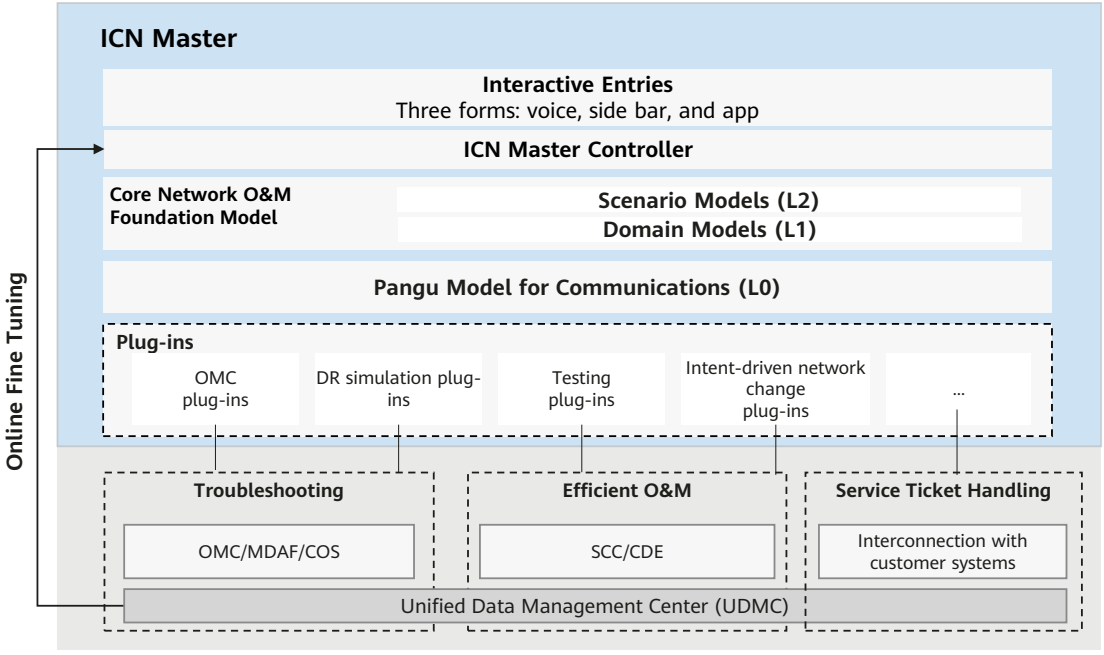


Figure 15 Core network O&M foundation model

As shown in the above figure, Huawei's core network O&M model is built based on the Huawei Pangu communication model. This O&M model is a basic foundation model oriented for communications with strong capabilities. Underpinned by the Mixture of Experts (MoE) technique, it supports trillions of parameters.

The L1 model dedicated for the core network is obtained based on the high-quality pre-training L0 model. Combined with Supervised Fine Tuning (SFT) and high-quality retraining corpuses provided by experienced core network R&D engineers, the L1 model covers numerous device-level positive and negative corpuses and samples. Furthermore, through the efforts of senior core network experts in generating scenario-specific high-quality retraining corpuses and with high-level fine tuning technologies, including Chain of Thought (COT), Reinforcement Learning from Human Feedback (RLHF), and

Toolformer (a tool for training foundation models), the scenario-specific L2 model is achieved, supporting analysis and handling of high-level and complex network problems. After the model is applied on the live network, retraining and fine-tuning are performed again based on live-network data to ensure the adaptability of the model to different sites, further improving the performance of the model.

8.3 Service Intelligence: AIGC Fuels 5.5G New Calling Services, Continuously Improving User Experience

AIGC is also called generative AI. It refers to the process of using AI algorithms to generate content, including text, code, images, videos, 3D content, and audios. Currently, AIGC models are classified into the Large Language Model (LLM) and siloed media single-modal generative models (images, videos, 3D content, and audios are generated separately). Well-known single-modal generative models include:

- Image generation: Midjourney, Bing Creator, Stable Diffusion, and Dall-E2
- Video generation: D-ID and Runway
- 3D: Magic3D and DreamFusion
- Audio: Jukebox and MusicLM

The content generated by these models has been widely used in film and game production.

According to the analysis of research institutes such as Facebook and the Huawei Central Media Technology Institute (2012 Labs), the siloed single-modal generative model will gradually evolve to the multi-modal generative model. Media content such as images, videos, 3D content, and music can be generated through one foundation model, realizing multi-modal input and output.

On top of that, with rising 5G network coverage, the network quality is continuously optimized. For instance, operators in China usually plan 2 Mbit/s uplink bandwidth at the edge of 5G cells, which can adequately serve 720p video calls, paving the way for VoNR video calls to become a basic service. It has become a consensus that 5.5G New Calling will be boosted with the in-depth combination of AIGC and VoNR video calls.

We believe that 5.5G New Calling can be divided into three phases:

Phase 1: Fun Calling

Figure 16 Fun calling

This phase focuses on human-to-human communication. Content is enhanced by the AI-powered capability on the network side. Main application scenarios include background replacement during video calls, avatar replacement, emojis, and gesture recognition. In this phase, AIGC foundation models are not required.

The following figure shows the network architecture in phase 1. It can be seen that New Calling capabilities are added to the VoLTE network, including the New Calling Platform (NCP) on the control plane and the Unified Media Function (UMF) on the media plane.

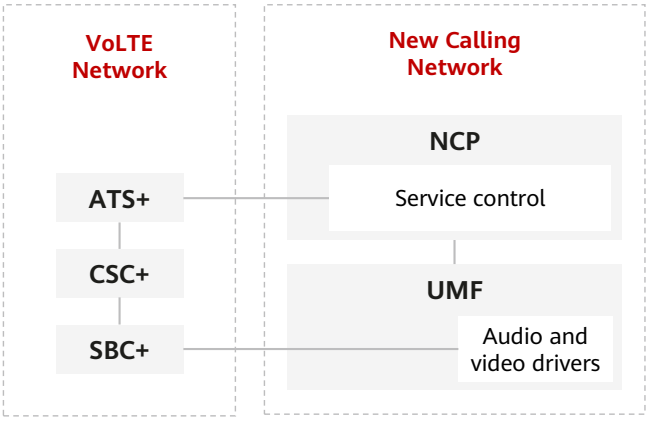


Figure 18 VoLTE network

Phase 2: Digital Characters

Figure 17 Digital characters

In this phase, 2D digital characters are introduced. The AIGC model is brought in to generate personalized digital images and content. The main service scenarios include communication between humans and voice-driven digital characters and communication between digital characters.

The following figure shows the network architecture in phase 2. The Multi-Modal Communication Function (MCF) is added to the New Calling network, further enhancing the AIGC and service openness capabilities.

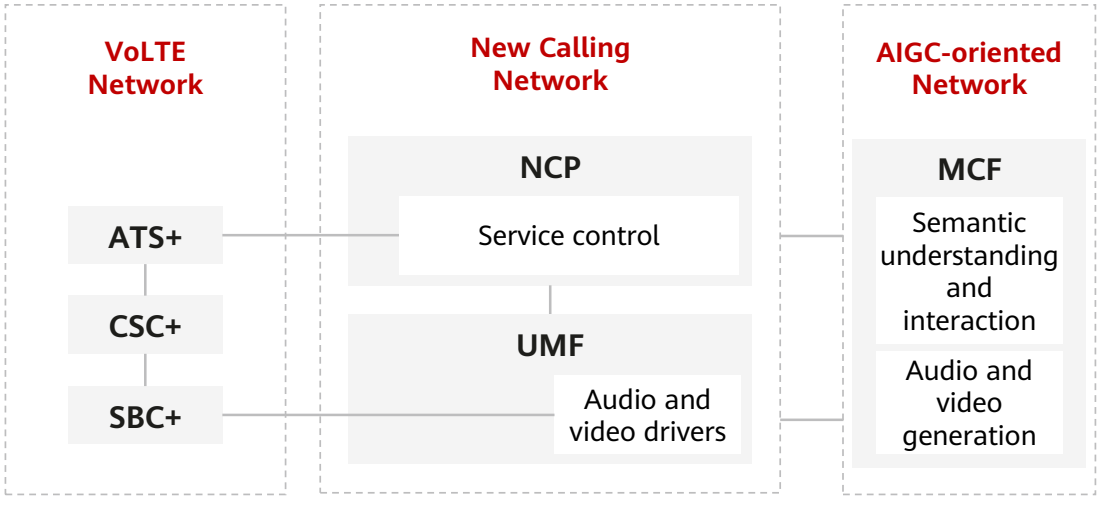


Figure 19 Network architecture for phase 2

Phase 3: Multimodal Interaction

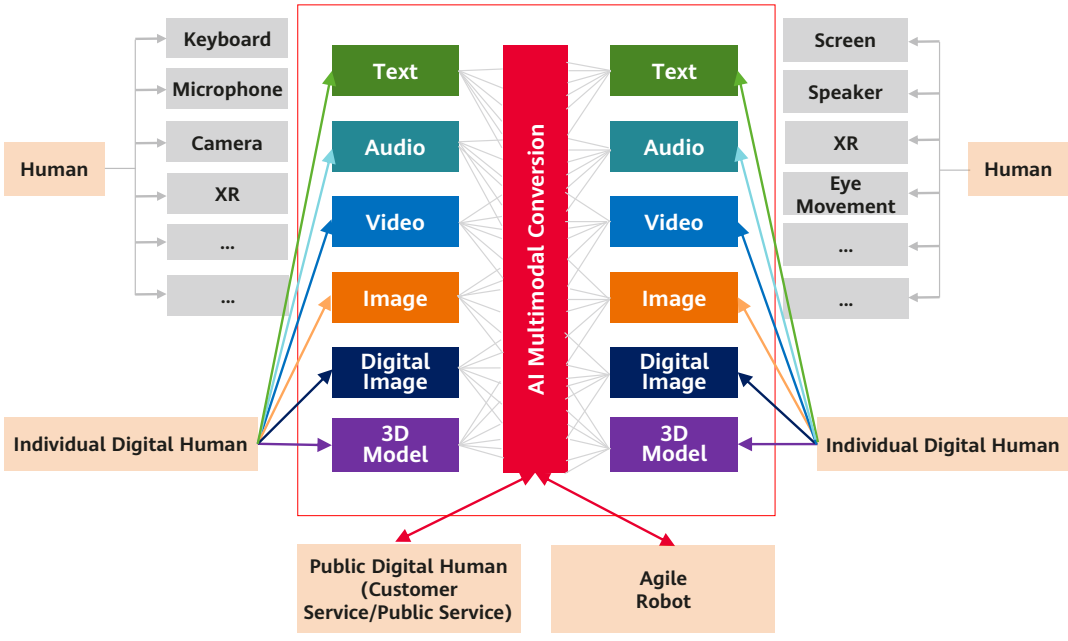


Figure 20 Multimodel interconnection

At this phase, as a digital assistant is introduced to support semantic understanding and information conversion for multiple models, 2D virtual humans are upgraded to 3D with higher authentication. AIGC is used for generating text, audios, and videos for more scenarios, personalizing content and upgrading experience. The network architecture at phase 3 is further enhanced on the basis of phase 2.

5.5G New Calling paves the way for 6G immersive communication while 6G provides basic communications for the metaverse.

Huawei, partnered with operators, has been driving the development of 5.5G New Calling. Huawei first showcased the original model of New Calling at PT EXPO China 2020, generating a huge buzz. In 2023, the four major operators in China revealed their commercial plans for New Calling. At PT EXPO China 2023 – New Calling Industry Development Forum, China Mobile announced their network construction and commercial plans for their 50 million subscribers. It is clear that 5.5G New Calling is now on the fast track.

8.4 Network Intelligence: Digital Twin and Network Foundation Model Create a Closed-Loop Network Autonomous System

First introduced in 3GPP Release 15, the Network Data Analytics Function (NWDAF) is an important function of the 5G core network that collects network data and standardizes AI analytics. It boasts a high performance in real time and makes use of closed-loop management. Pioneering the standard AI-powered system, NWDAF paves the foundation for AI-assisted data analytics on 5G networks.

The NWDAF defined in 3GPP Release 15 can only outline details on network slices while its architecture and analytics are well enhanced in the subsequent Release 16 and Release 17. It is worth mentioning that the NWDAF architecture has seen a major transformation in Release 17, where five logical functions form an initial framework of the intelligence plane. These are defined as follows:

- MTLF: Model Training Logical Function
- AnLF: Analytics Logical Function, which obtains models from the MTLF.
- DCCF: Data Collection Coordination Function, which collects and shares data between the NWDAF and other NFs for data aggregation.
- ADRF: Analytics Data Repository Function, which stores collected raw data and analytics results.
- MFAF: Messaging Framework Adaptor Function, which transfers NWDAF-related messages, including data fetch and collection as well as notifications for fetching data.

3GPP Release 18 has brought about new NWDAF scenarios by making models and analytics more accurate, assisting in application detection and UE Route Selection Policy (URSP) making, and enhancing QoS prediction. In the upcoming Release 19, even more scenarios are expected to be added, including signaling storms, energy saving, recommendation, online learning, reinforcement learning, federated learning, and the coordination with RAN.

The intelligence plane of the 5.5G core network requires the NWDAF architecture to be further evolved so that more complex scenarios beyond those defined in 3GPP can be covered. For example:

1. The high-precision network map enables the real-time display of traffic.

The core network towards 5.5G keeps scaling up by enriching services, making it difficult to ensure a visible, manageable, and controllable network. With the intelligence plane, a high-precision network map can enable the real-time display of traffic.

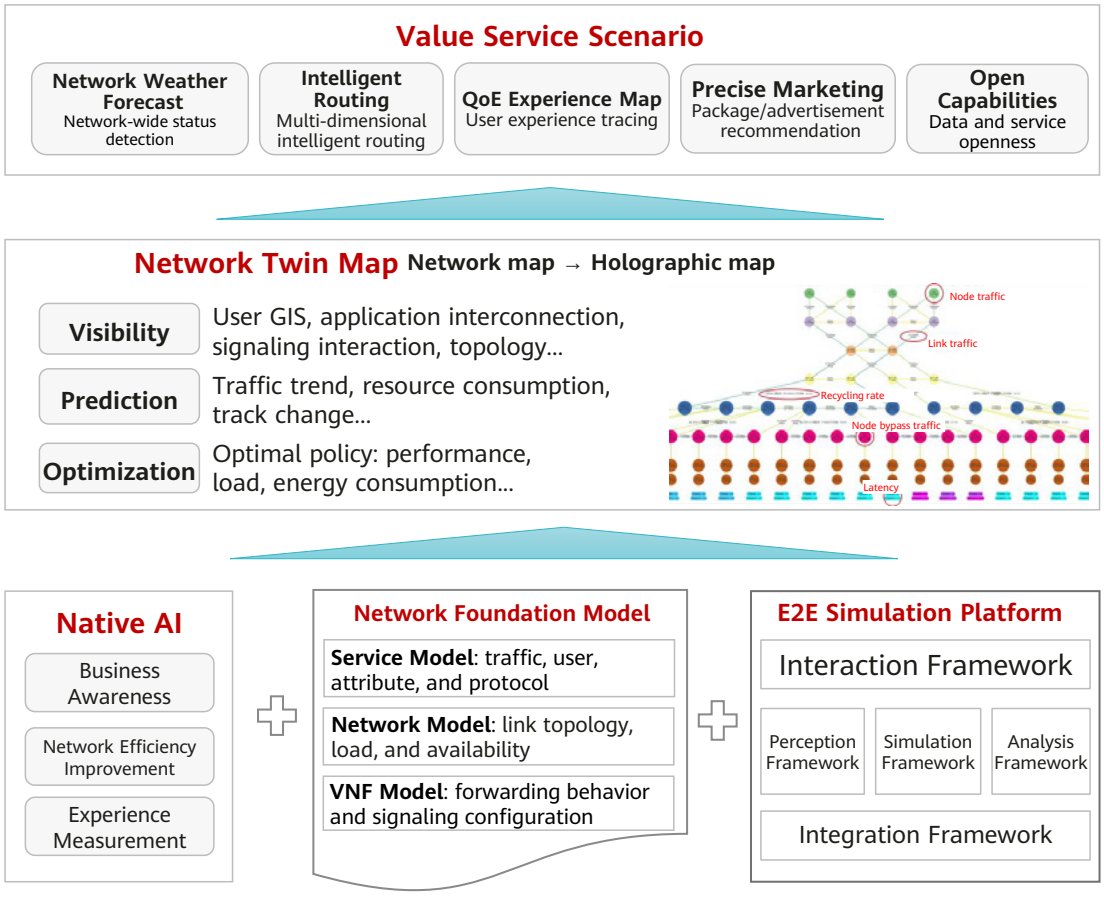


Figure 21 High-precision network twin map

The high-precision network twin map, composed from native AI, the Network Foundation Model, and the E2E simulation platform, underpins several value service scenarios. The key technologies and application scenarios are as follows:

- Network twin map: This map collects network and service data in real time, converts the data into structured data, and displays it holographically. With the help of algorithms, like operations research and network calculus, resources can be optimally allocated in complex networking, and network compliance and service experiences can be accurately predicated.

- Cross-domain all-service network model: By using the adaptive modeling on various complex B2B and B2C traffic, modeling on network functions, network topology and load, as well as cross-domain modeling on the peer-network-cloud synergy, a white-box service scenario repository can be built, paving the way for E2E network simulation.

- E2E simulation platform: This platform enables E2E quantitative simulation for key services covering the core network, wireless network, and IP transport network. It underpins service scenarios like problem analysis and experience optimization, making it flexible to deploy services online or offline.

2. Intelligent closed-loop experience is assured through RAN collaboration.

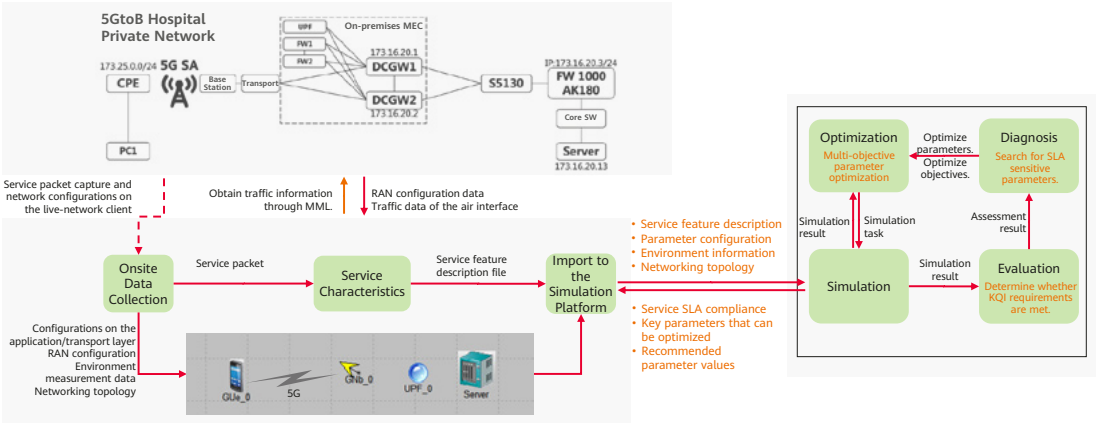


Figure 22 Intelligent closed-loop experience assured through RAN collaboration

As outlined, the key capabilities on the intelligence plane can be provided after services are migrated from industry private networks to the 5G core network:

- After services are migrated to the 5G core network, SLA deterioration may arise. The root cause for this can be located by reviewing, simulating, and analyzing onsite data based on the intelligence plane.

- The optimal parameter portfolio can be obtained based on the different SLA requirements of customers, like optimal performance or optimal resource consumption. In addition to the core network parameters, the configuration parameters of the application layer and the radio air interface can also be a part of a portfolio.

3. Network efficiency is improved.

On the 5.5G core network, the user plane is moving down towards the network edges. Considering the need to build a hyper-distributed and heterogeneous user plane, it is pivotal that routes can be efficiently and intelligently selected.

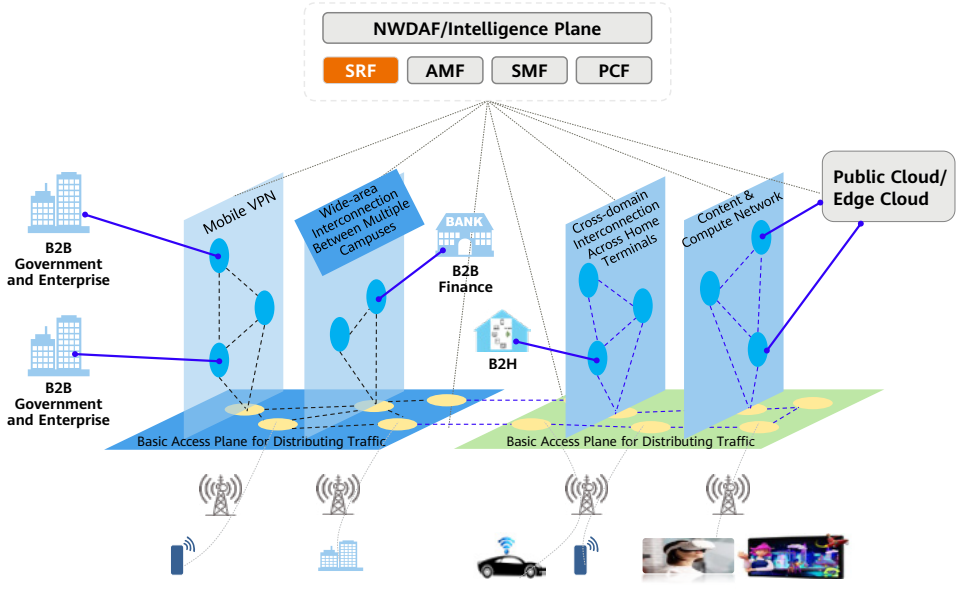


Figure 23 NWDAF/Intelligence plane

The figure demonstrates that the 5.5G core network carries varied and flexible B2B, B2C, and B2H services, including Mobile VPN, wide-area interconnection between multiple campuses, cross-domain interconnection across home terminals, and content & compute network. The collaboration between the intelligence plane and SRF schedules overall service-oriented network and compute resources, bringing with it the ultimate user experience. To be more specific, the intelligence plane can detect, analyze and make decisions on network changes in real time, while the SRF, assisted by the intelligence plane, can select routes based on multiple dimensions, including terminal and service type, access points of users and services, application and path status as well as service and compute offloading capabilities. The collaboration between the intelligence plane and SRF meets users' service requirements, maximizing network efficiency.

9

Summary and Prospect

The intelligent era has arrived. Not only will AI technologies reshape the world by changing how individuals learn, work, and live, but will gradually revolutionize how industrial production, management, and transactions are done. As network capabilities are increasingly improved and combined with AI, the ubiquity of connectivity and intelligence is a not too distant reality.

This white paper focuses on the core network architecture, service requirements, and technology trends for the transition to an intelligent world, hoping to provide guidance for the next phase of 5G network development. Continuous enhancements of network capabilities enable core networks to be fully cloudified, fully connected, and super-distributed, which in turn fuels the sustainable optimization of basic network services. Furthermore, progress is increasingly made towards the intelligence of services, the network architecture, and O&M. For the aforementioned reasons, core networks will lead to the creation of new businesses. While we look forward to 5G-Advanced, research on 6G technologies are underway. Intelligence will become a crucial engine for network evolution and service enablement, driving the combination of technologies and businesses, creating more opportunities, and achieving a win-win outcome in the industry chain.

We hope that this white paper can provide reference scenario requirements and technical directions for the development of intelligence-oriented core networks, promote consensus in the industry, and with concerted efforts, we will open a new chapter for the mobile communications network.

Abbreviations

Aabbreviation	Full Spelling
XSO	Cross-layer Service Orchestration
5GC	5G Core Network
AAA	authentication, authorization, and accounting
AF	Application Function
AMF	Access and Mobility Management Function
CHF	Charging Function
LMF	Location Management Function
NEF	Network Exposure Function
NF	Network Function
NWDAF	Network Data Analytics Function
RAN	Radio Access Network
SF	Sensing Function
TMF	Tag Management Function
SMF	Session Management Function
UDM	Unified Data Management
UPF	User Plane Function
VoLTE	Voice over LTE
VoNR	Voice over New Radio
CS	Circuit Switched
Vo5G	Voice over 5G
KQI	Key Quality Indicator
EPS FB	EPS Fallback
ViLTE	Video over LTE
NaaS	Network as a Service
API	Application Programming Interface
XR	Extended Reality
AI	Artificial Intelligence
DC	Data Channel
AIGC	AI Generated Content
IVR	Interactive Voice Response
PGC	Professionally Generated Content
UGC	User Generated Content
MoQ	Media over QUIC

HUAWEI TECHNOLOGIES CO.,LTD.
Huawei Industrial Base
Bantian Longgang
Shenzhen 518129,P.R.China
Tel: +86 755 28780808

www.huawei.com