China Unicom 中国联通
创新·改变世界

China Unicom Edge Computing Technology White Paper

2017-06

Contents

1 Overview......1
1.1 Vision and Objectives......1
1.2 Status......2
2 MEC Drivers and Challenge Analysis......3
2.1 Industry and Market Development Demands......3
2.1.1 Service and Technology Drivers......3
2.1.2 Commercial and Industrial Drivers......5
2.2 Analysis of Telecom Operator's Network Challenge......6
2.2.1 Silo Network Architecture Unable to Meet Service Development Requirements......6
2.2.2 ICT Convergence Drives Operators to Change "Dumb Pipe Operation" Pattern......7
3 China Unicom MEC Platform Capabilities and Application Requirements......8
3.1 MEC Platform Capability Requirements......8
3.1.1 Service Domain......8
3.1.2 Management Domain......9
3.2 MEC Typical Application Requirements......10
4 MEC Deployment Policy of China Unicom LTE Network......14
4.1 MEC Networking Architecture of LTE Network......14
4.2 MEC Deployment Scheme of China Unicom LTE Network......14
4.2.1 Deployment Location......14
4.2.2 Billing Scheme......16
4.3 Problem Analysis of MEC Deployment......16
5 MEC Technology Evolution Route and Planning......17
5.1 Evolution of 5G Network-oriented MEC Key Technology......17
5.1.1 Traffic Grooming Scheme......17
5.1.2 Service Continuity Scheme......18
5.1.3 Intelligent Perception and Optimization Scheme......19
5.2 Evolution of China Unicom MEC Networking Architecture......19
5.3 China Unicom 5G Network MEC Deployment Planning......21
6 Summary and Prospect......23

China Unicom Edge Computing Technology White Paper

1 Overview

1.1 Vision and Objectives

As the information and communication technology is integrated and penetrating into various sectors today, digital information has become a key factor of production. There is a growing trend of transforming to digitization in all economic and social sectors. Through the deep integration with such technologies as cloud computing, big data, virtual reality (VR), augmented reality (AR) and artificial intelligence, 5G network will connect people to everything to become a critical infrastructure for digital transformation in all sectors. 5G mainly consists of three application scenarios: Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC) and Ultra-reliable and Low Latency Communications (uRLLC). Among them, eMBB focuses on services that have high requirements for bandwidth, such as HD video, VR and AR, so as to satisfy our demands for digital life. mMTC focuses on services that have high requirements for connection density, such as smart city, smart agriculture and smart home, so as to satisfy our demands for digital society. uRLLC focuses on services that are extremely sensitive to latency, such as autonomous driving, industrial control and telemedicine, so as to satisfy our demands for digital industry. According to IDC's latest statistical report, there will be more than 50 billion terminals and devices connected to the Internet by 2020. By 2018, 50% of the IoT network will be bandwidth constrained, while 40% of the data will need to be analyzed, processed and stored at the edge of the network.

Multi-Access Edge Computing (MEC) is an open platform which integrates the core capabilities of network, computing, storage and application at the network edge close to people, things or data sources. The platform provides edge intelligence services to meet the key requirements of industry digitization in various aspects such as agile connection, real-time service, data optimization, application intelligence, security and privacy protection. In 3GPP R15, based on the service-oriented architecture, 5G protocol module can be called according to service requirements and provide technical standards for constructing edge network. As a result, the MEC can be flexibly deployed in wireless access cloud, edge cloud or converged cloud in different scenarios based on demands. MEC can provide mobile operators with the following values:

- To reduce the utilization of the core network and backbone transmission network and effectively improve the utilization rate of the operator network through local offload of high-bandwidth services such as 4K/8K and VR/AR;
- The operator network will effectively support future latency-sensitive services (such as the Internet of Vehicles and remote control) and services requiring high computing and processing capabilities (such as video monitoring and analysis) by descending the content and computing capability, so as to help operators transform from connection pipe to information-based service enablement platform;
- As an edge cloud computing environment and network capability open platform, MEC will lay a foundation for operators to construct the network edge ecology.

MEC is one of 5G network enablement technologies, but currently it can be deployed in LTE network due to the openness of its architecture and platform to provide mobile operators with value-added services. In June 2017, China Unicom joined hands with Nokia, Tencent and INTEL to successfully build a network edge cloud system at Shanghai Mercedes-Benz Arena for the first time. They also validated the MEC-based multi-angle live video and broadcaster interaction service by utilizing the existing LTE network. The test data shows a live video latency of only 0.5s in the venue. Compared to the traditional Internet live video with a latency of over 30s, the system significantly improves the user's real-time viewing experience and lays a solid foundation for promoting and constructing China Unicom’s 5G network-oriented intelligent venue solution.

Based on 5G service requirements and the progress of the MEC industry, this White Paper defines the requirements of China Unicom for MEC platform capabilities and application scenarios, and provides suggestions for China Unicom’s 4G network MEC deployment policy and 5G network-oriented evolution planning. We are looking forward to discussing the MEC commercial cooperation mode with all sectors so as to jointly build the network edge ecology and comprehensively accelerate the vigorous development of 5G services.

1.2 Status

This version 1.0 of the white paper may need to be further revised to be more comprehensive, but we hope that the release of this version will be helpful for the industry. With the freezing of MEC technical standards and the deployment of 5G pre-commercial network, new research content may be added to subsequent versions. All comments and suggestions are welcome.

2 MEC Drivers and Challenge Analysis

2.1 Industry and Market Development Demands

2.1.1 Service and Technology Drivers

(1) Service Driver

The LTE network has a simple designed objective and purpose, i.e., to deliver high-speed mobile broadband service at the highest wireless rate possible. With the rapid development of mobile Internet and IoT, 5G service will present the characteristics of diversified demands. As shown in Figure 2.1, 3GPP defines three main application scenarios for 5G, i.e., eMBB, mMTC and uRLLC. On the one hand, 5G will provide users with a more immersive service experience such as UHD video, next-generation social network, VR and AR, so as to promote another upgrade of human interaction mode. On the other hand, 5G will deeply integrate mobile communication with typical application scenarios represented by smart city and smart home. It's estimated that hundreds of billions of devices will access the 5G network. In addition, by virtue of its advantages such as ultra-low latency and ultra-reliability, 5G will also be combined with vertical industry applications such as the Internet of Vehicles, Industrial Internet, mobile healthcare and energy. On the whole, due to the diversity of 5G services, the 5G network will be capable of supporting the demands for ultra-speed mobile ultra-broadband, IoT mass connection, ultra-reliable and ultra-low latency connections. A more flexible, more intelligent, programmable and scalable network can also be provided so as to support new services and new applications.

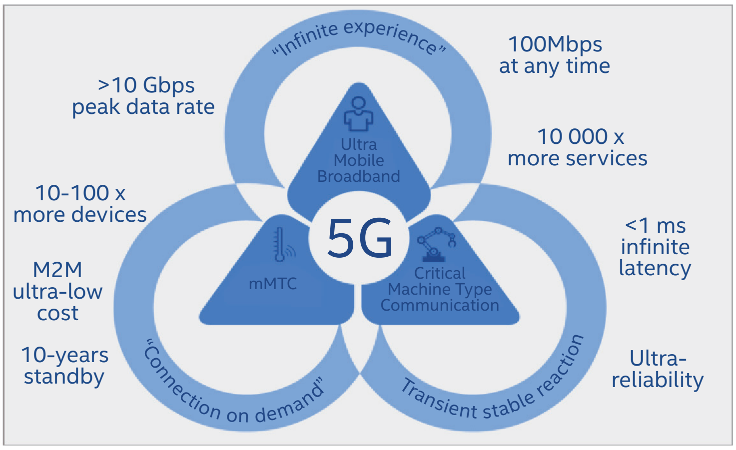

Figure 2.1 Three Main Application Scenarios of 5G

In high capacity hotspot scenarios of 5G network, the user experience rate reaches 1Gbit/s, the peak rate reaches 10Gbit/s, and the traffic density reaches 10Tbit/s per square kilometer. For example, the broadband access requirements of such as services HD video, smart city and B2B is dozens to hundreds of Mbit/s, while the network bandwidth requirements of services such as 4K video, 3D video, AR and VR is dozens of Gbit/s. This will cause great pressure to the radio backhaul network. Therefore, the services shall be shifted to the network edge as much as possible (for example, the edge cloud data center or closer to the base station), so as to implement the local offload of the services. In addition, in 5G uRLLC low latency scenarios, an end-to-end latency at millisecond level is expected (the end-to-end transmission latency and service processing latency of the LTE network are currently above 50ms). This also requires descending the services to the network edge, so as to reduce the network latency caused by network transmission and multi-stage service forwarding. As the key technology of the 5G evolution, MEC can provide cloud computing capacity and IT service environment at the mobile network edge which is closer to customers. With characteristics including ultra-low latency, ultra-large bandwidth, localization, high real-time analysis and processing, MEC can reduce the utilization of the core network and backbone transmission network, as well as the end-to-end latency.

(2) Technology Driver

5G network enables the opening of mobile network capabilities and implements the platform-oriented operation of services by using big data and cloud computing technologies. Figure 2.2 constructs the 5G telecom cloud architecture from four dimensions, i.e., Network Function Virtualization (NFV), Software Defined Network (SDN), Cloud Radio Access Network (Cloud-RAN), and the automatic management and collaboration system. NFV achieves the virtualization of network function and management through decoupling the hardware and software. SDN achieves the on-demand reallocation of transmission and control resources at the data center and transmission node by separating the control and service, so that the network has programmable capability. Based on a generic IT hardware architecture, Cloud-RAN provides mobile service-oriented fronthaul and backhaul network through centralized management and virtualized software functions. Automatic management and collaboration systems manage the life cycle, priority and QoS of cloud network service through the global collaboration of service and network.

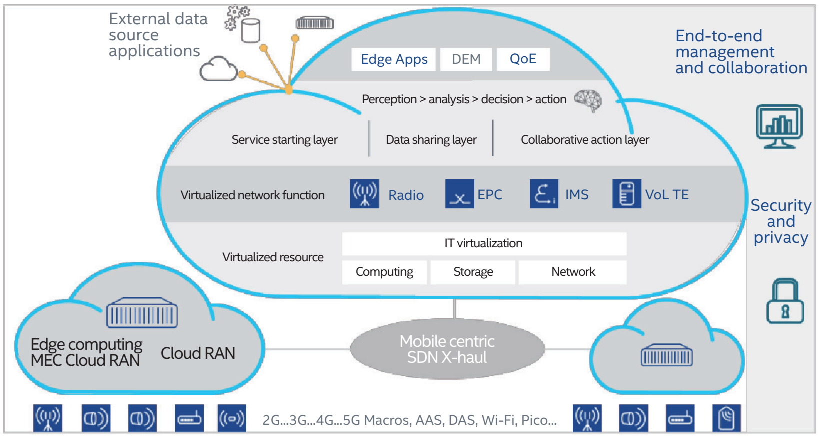

Figure 2.2 Logic Diagram of 5G Telecom Cloud Architecture

Technologies such as cloud computing, NFV, SDN and ICT have led to the development of MEC. The core equipment of MEC system is MEC server built on general IT hardware platform, which provides localized public cloud service through cloud computing facilities (i.e., edge cloud) deployed within the wireless base station or at the wireless access network edge. The MEC server can also connect to the private cloud within other networks (such as the enterprise network) to achieve hybrid cloud service. Cloud-RAN and V-RAN provide an appropriate entry point for MEC deployment. For example, the small Cloud-RAN deployed in a retail center can concurrently deploy MEC functions to manage localized applications, so that application, service and content can be closer to the end users. The 5G network architecture itself will be a service and user oriented network, which can flexibly compose and elastically manage network resources through network slicing according to service requirements. For example, uRLLC service can be implemented through edge network slicing.

2.1.2 Commercial and Industrial Drivers

Due to the rapid growth of mobile services and the increase of operation cost pressure, the entire telecom industry hopes to improve user experience, increase profits, optimize network operation and enhance resource utilization efficiency by introducing new technologies and products. Facing eMBB's high flow capacity and mMTC's mass connections, network operators have to consider the possibility of network congestion and think about how to weaken the influence on network security and service backhaul through local analysis and processing. Enterprise customers hope to provide their customers with more effective and secure network connections with lower latency. Content and service providers are also faced with the challenge of high network latency for user experience, which is crucial in some applications and must be solved.

In addition, as more and more smartphone applications and content are migrated to the cloud, the latency and bandwidth for end customers to access the cloud server must be optimized to ensure a rich experience for consumers. This requires a closer cooperation between network operators and application service providers. Naturally, such cooperation leads to the deployment of application/content at the network edge closer to customers. For this purpose, MEC industry participants are actively validating, testing and deploying new technologies and products to reduce the network latency, improve the network security, and provide a more agile platform for new function development and deployment. Such combined efforts play a great role in improving the end-user experience and network efficiency.

The standards for MEC are steadily progressing in order to attract more participants to promote the development of the entire MEC industry. Since it was established in October 2014, the MEC industrialization standard group already has more than 70 members, including mainstream operators, communication and network equipment providers, IT/middleware vendors, software vendors and OTT application vendors. These members almost cover the complete mobile Internet industry. Member units have jointly proposed and designed rich application scenarios and fully explored the unique value potential of mobile network edge computing. In April 2016, 3GPP SA2 has officially accepted MEC as a key technology for 5G architecture. The development of MEC business and industry will benefit from business transformation, technology evolution and industrial cooperation. These efforts will be eventually reflected in increasingly wider innovation markets, such as smart healthcare, the Internet of Vehicles, industrial automation, VR and AR, gaming and IoT services.

MEC provides location-based cloud computing capability and real-time content information through the IT service environment deployed at the RAN edge. Such an open IT service environment will integrate the applications and services provided by mobile operators and content service providers into the MEC platforms of more providers. The general architecture will promote the rapid deployment of more innovative applications in wider fields and provide the related participants with more value. End users can obtain the extremely-rapid personalized user experience through MEC real-time network connections and content acquisition. By opening the MEC platform computing capability to third-party OTT providers or application developers, and providing services to mobile users, enterprises and vertical industry, mobile operators can redefine their roles in the supply chain, generate new revenue, provide more -value-added services, and create new market opportunities. With the MEC open standard platform, OTT providers and independent application developers can rapidly develop new applications and shorten the development cycle, so as to provide end users with the best experience with near-zero latency. All MEC value chain participants will eventually benefit from the market environment created by the MEC and promote the sustainable economic development through applications covering numerous industries.

2.2 Analysis of Telecom Operator's Network Challenge

2.2.1 Silo Network Architecture Unable to Meet Service Development Requirements

With the explosive growth of mobile Internet, IoT and industry applications, the mobile communication network will face a 1000x growth in data traffic and while requiring the connections of 100B devices. The traditional 4G silo architecture, which can no longer meet the service development requirements, is evolving to a flat "Client-Connection-Cloud" architecture. By adopting NFV and SDN technologies for decomposition, abstraction and reconstruction of element function, 5G networks will form a new flat IT platform consisting of access plane, control plane and forwarding plane. The network will evolve to control function centralization and forwarding function decentralization. The forwarding plane, focusing on the routing forwarding of service data, has such features as simplicity, stability and high performance to meet the forwarding requirements of future mass mobile traffic. The control plane adopts a logic concentration mode to implement unified policy control, ensuring flexible mobile traffic scheduling and connection management. The centrally-deployed control plane achieves the programmable control of the forwarding plane through mobile flow control interface.

The separation of control plane and forwarding plane in 5G the network further flattens the network architecture. The forwarding plane gateway equipment can adopt distributed on-demand deployment to fundamentally solve the transmission and core network overloads caused by silo single service flow of 4G network through diversified multi-stage service flow. Under the centralized scheduling of control plane, the forwarding plane achieves low latency, high bandwidth and uniform load transmission of end-to-end mass service data flow through such technologies as flexible gateway anchor, edge content and computing, so as to improve the carrying efficiency of in-network packet data and users' service experience.

5G organizations including ETSI, 3GPP and NGMN have defined edge computing as a part of 5G architecture, and made efforts in research and standardization. Edge computing can help decentralizing telecom network architecture, achieve operator service localized processing, improve network data processing efficiency, provide end users with an ultimate experience, and meet the demands of vertical industry network for low latency, mass flow and high security. For example, the low-latency automatic driving service requires the network function and service processing function of the core network to descend to the edge close to the access network to reduce the intermediate level and achieve low-latency service processing. In AR/VR video playback scenarios that require large bandwidth, the video needs to be cached at the node close to the edge access side, so as to save a large amount of transmission bandwidth and improve the network processing efficiency. In application scenarios that require improving user experience through video acceleration or implanting push information through Toobar, the MEC open platform can be used for rapid integration of third-party applications.

2.2.2 ICT Convergence Drives Operators to Change "Dumb Pipe Operation" Pattern

Mobile Internet has broken the original "walled garden" pattern of telecom operators. The rapid emergence of various OTT service types and the sharp increase in the corresponding business quantity have promoted the gradual provision of the data pipe of the mobile communication network. As a result, the control of operators for users will be weakened gradually, the trend of data "dumb pipe" will accelerate and become a link with low value in the mobile Internet. At present, operators adopt the tariff mode which is mainly based on the usage of data traffic, while this mode is relatively unitary compared to the flexible OTT business model. In the MBB era with exponential growth of traffic, the "scissors difference" between the equipment investment growth (proportional to traffic growth) and income growth of operators is becoming increasingly obvious.

Telecom operators need to grasp the business opportunities of ICT convergence so as to carry out transformation in the aspects of service innovation, platform integration, ecosystem construction and business model exploration. In accordance to that, they will complete the transformation from "traffic dividend" to "digital dividend" and then to "information dividend". On the one hand, smart pipes are to be constructed to implement the visualization, manageability and operability of network resources. Operators need to detect and analyze the traffic data of mobile Internet, collect the behavior characteristics of users, and adopt a control policy, so that the pipes can produce added value. In addition, high-value service applications need to be refined or targeted, services need to be promoted and pushed to improve the user experience, guide the user behaviors, and implement a refined management of traffic. On the other hand, API interfaces need to be unified to build the service enablement platform, and close cooperation with OTT and third-party application developers shall be made to provide users with end-to-end application solutions and improve the users viscosity.

Constructing unified cloud computing infrastructure and providing the service capability with main features of computing, storage, network and security are inevitable choices of telecom operators to change the "dumb pipe operation" pattern in the tide of ICT convergence. As the product of IT and CT convergence, the MEC is a powerful tool for digital transformation for operators, helping operators to quickly build the bridge for cooperation with OTT or application developers. Telecom operators can open the storage and computing capabilities of the MEC platform to application developers and content providers to provide them with a brand-new service development environment and user experience. In addition, the eNB information at the wireless side can be encapsulated into various services (for example, RNIS, location-based service, and bandwidth management service) running on the MEC platform and available to enterprises and vertical industry, so as to provide more value-added services and maximize the network value.

3 China Unicom MEC Platform Capabilities and Application Requirements

3.1 MEC Platform Capability Requirements

The MEC virtualization platform, located between wireless access network and mobile core network, can use the cloud computing facilities (edge cloud) in the wireless base station or at the edge of wireless access network to provide localized public cloud services. It can also connect the private cloud in other networks (such as enterprise network) to form a hybrid cloud. Based on specific cloud computing system (such as OpenStack), the MEC platform provides a virtualization software environment for planning and managing the IT resources in the edge cloud. Third-party applications deployed in the edge cloud in the form of virtual machines (VM) can acquire open wireless network capability through unified API. The MEC platform consists of service domain and management domain. The service domain is used to support the running of third-party applications. The management domain is used to manage the service domain of the MEC platform.

3.1.1 Service Domain

The service domain of the MEC platform includes data plane (DP) and API enablement.

1) Data plane: DP provides a data forwarding path between the wireless access network and core network, achieves the local offload of data traffic, and provides network virtualization support for third-party service hosts as well as data forwarding between internal functional components of the MEC platform. DP must provide the analytical processing capability specific to GTP-U data flow: In the direction of uplink data, DP shall first parse the GTP packets received from the base station and then forward them according to the service requirements. In the direction of downlink data, DP shall encapsulate the data received from third-party services into the proper GTP tunnel, so that the mobile terminal can receive the data through the base station. In addition, DP shall also provide the following functions:

- The wireless access network may introduce IPsec for encrypting the service data on the backhaul line, so DP needs to support IPsec encryption and decryption functions;
- The movement of the mobile terminal may lead to the location change of the network access point, causing the interruption of a service session, so DP must support specific forwarding capability to implement continuous transmission of service data;
- DP must provide a backup link. When partial function fails, the service data of a user can be transmitted through a backup link, thus avoiding network interruption.

2) API enablement: API enablement opens wireless network capabilities to third-party applications. Different network capabilities are open through specific APIs. API calls can be made by an external third-party application or an internal function of MEC platform. A specific network capability is abstracted as a specific API, its access and call must be monitored and managed, ensuring that only authorized third parties can obtain credible capability call service. In terms of call objects, API enablement is mainly classified into five categories: API for wireless access equipment (such as eNB); API for mobile core network equipment (such as S/P-GW); API for operator service and operation support system (such as BSS and OSS); API for internal components of MEC platform (such as DP); and API for user service data flow (such as DNS query request).

API enablement must provide intelligence of third-party applications and improve the development experience of third parties to enable quick and convenient calls of network capabilities and ensure the efficient and intelligent response of the network to external call requests. As shown in Figure 3.1, China Unicom is committed to building a MEC unified API service platform, standardizing interfaces such as MP1 and MM9, providing a unified API for third-party applications, and leading the MEC platform of equipment manufacturers to provide the same open capability so that third-party application providers can provide the application software version supporting compatible deployment.

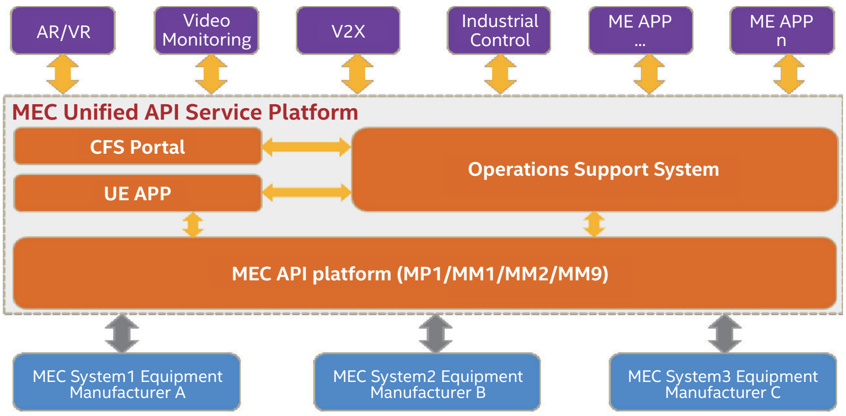

Figure 3.1 China Unicom MEC Unified API Service Platform

In addition, API enablement shall provide sufficient flexibility for supporting network capability openness, so that continuous openness will be provided to meet richer network functions, without making complicated changes to third-party application platforms or network systems. Considering that the current mainstream online applications are mainly based on a Web framework, API enablement can be split into HTTP operation-based logic resources and embedded in a specific HTTP session for delivery together with service data so as to facilitate a third-party call. Third-party application platforms (such as OTT) can be called based on the existing production environment without modifying their own platform architecture or service logic.

3.1.2 Management Domain

The management domain is responsible for managing the MEC platform, including data plane control, API enablement control, resource usage statistics and third-party service deployment management. On the one hand, the management domain shall allow operators and third-party users to manually select and configure the resources through Web-based control plane, or provide API for supporting programming-based selection and configuration services. On the other hand, the management domain can perform dynamic real-time control for service domains based on predefined rules according to the actual operation conditions of the MEC platform.

1) Data plane control: The management domain needs to monitor and manage the data plane, ensuring the controllability and manageability of route forwarding behaviors of the MEC platform.

2) API enablement control: API calls must be authorized by the management domain. On the one hand, third-party applications must complete related API enablement registration with the management domain before calling the API. On the other hand, the management domain needs to license the installation of a specific API and monitor its running status.

3) Resource usage statistics: include IT resource usage statistics, API call statistics and data traffic statistics.

- IT resource usage statistics refers to making statistics on the IT resource usage of the virtual host rented by a third party (for example, CPU and memory).
- API call statistics refers to making statistics on the API call initiated by a third party. The statistics are made based on different rules (such as call times and subscription time) according to different API types and different API call initiation sources (i.e., calls inside the MEC platform and calls outside the MEC platform).
- Data traffic statistics refers to making statistics on the data traffic locally exported by the data plane, and the specific forms include offline statistics, content statistics and real-time statistics. Section 4.2.2 of this White Paper illustrates the real-time traffic statistics scheme.

4) IT basic resource management: By virtue of virtual machine monitoring, the management domain performs centralized management of physical and virtual IT infrastructures in the edge cloud and implements the resource planning and deployment, dynamic optimization and service composition, including managing the IT resource pool (such as computing capability, storage and network) of the edge cloud and providing support for virtualization technology.

3.2 MEC Typical Application Requirements

(1) QoS Optimization for Mobile Video

Currently, there is lack of interaction between the video content carried by the LTE cellular network and the pipe, so the user experience can hardly achieve the optimal effect. On the one hand, because of the rapid change of channel and air interface resources at the wireless side, it is difficult to dynamically adjust the application layer (HTTP/DASH) parameters to match the change of wireless channel. Similarly, the traditional TCP congestion control policy designed for wired environment is unable to accurately adapt to the changes of wireless channel. On the other hand, eNB doesn't know the content of application layer. Therefore it's unable to dynamically schedule the wireless resources for different types of services, or to provide differentiated QoS for different users of the same type of service. For example, eNB assigns the same QCI, MBR and GBR for online video users.

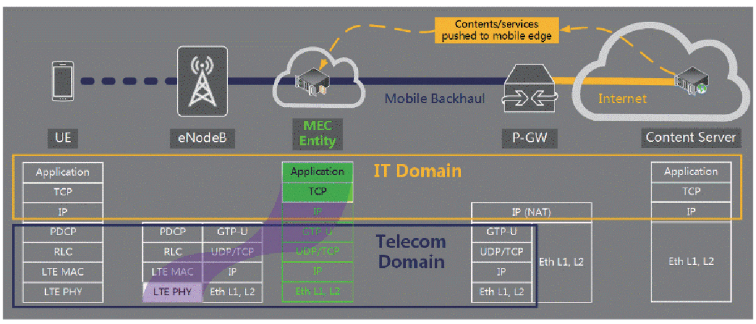

Figure 3.2 Implementing Cross-layer Video Optimization based on MEC Platform

The MEC platform can obtain the application layer and TCP layer information of OTT video services through the northbound interface, or obtain the information such as RAN side wireless channel (RNIS, Location services, etc) through the southbound interface so as to further improve the user perception experience through bi-directional cross-layer optimization and achieve intelligent operator pipes, as shown in Figure 3.2.

(2) Mobile CDN Descending

Generally, the CDN system of current mobile networks is deployed in the provincial-level IDC computer room which is far away from mobile users, instead of running inside the mobile network. Therefore, a vast amount of mobile backhaul bandwidth is utilized, while the services are still not proximate enough to satisfy mobile service scenarios which are more sensitive to latency and bandwidth. As shown in Figure 3.3, operators can deploy an edge CDN system in the MEC platform, OTT rents edge server nodes in IaaS mode to support its service content and orients the service to edge CDN nodes in its global DNS system.

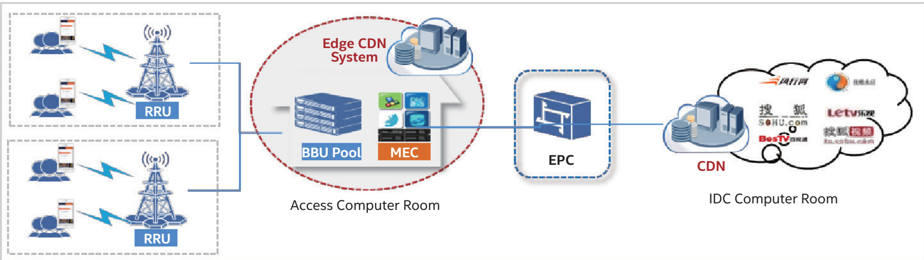

Figure 3.3 Implementing Mobile CDN Descending based on MEC Platform

(3) VR Live Video

In live video scenarios such as large-scale e-sports, ball match, F1 races and concerts, the users have higher requirements for latency for an immersive experience. As shown in Figure 3.4, the MEC platform can implement local mapping and distribution of VR video sources, providing a high-quality VR video experience. It can also provide the audience with a unique angle of view by using multi-angle panoramic cameras. For example, the audience who is far from the field can watch the venue just like sitting on a VIP seat through real-time VR. In addition, with advantages of low latency and high bandwidth, MEC can avoid the spinning sensation caused by restricted bandwidth and latency during a VR experience while reducing the consumption of backhaul resources.

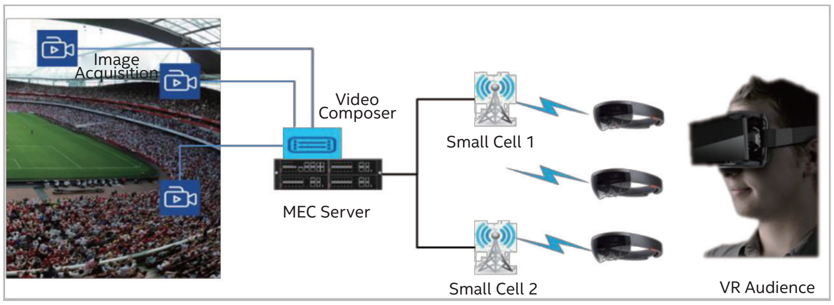

Figure 3.4 Implementing VR Live Video based on MEC Platform

(4) AR

In existing AR solutions, users have to download and install a large APP before experiencing AR, but the memory, battery capacity and storage capacity of smartphones have constrained the development of AR. As shown in Figure 3.5, the MEC platform determines the user location based on network data (such as the Location information fed back from the RAN), uses local AR servers to provide real-time matching, computing and pushing of AR content, and implements the real-time aggregation of local real scene and AR content channels to bring the users a new unique experience. In addition, the rapid and flexible deployment and discovery of location-related AR content can help form new MEC proximity content providing an advertising model.

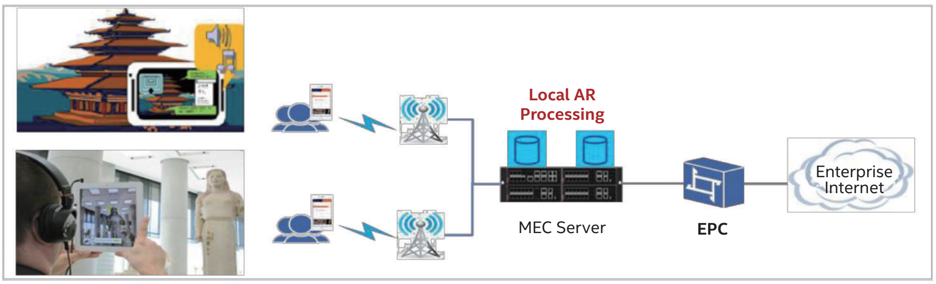

Figure 3.5 Implementing Augmented Reality based on MEC Platform

(5) Video Monitoring and Intelligent Analysis

The backhaul traffic of surveillance video is usually large, but most of the video images are stationary and not very useful. As shown in Figure 3.6, the MEC platform is used for the analysis and processing of video content where only the events and video clips with image changes are hauled back. A great deal of monitoring content is stored locally in MEC server, thus saving transmission resources. This solution can be effectively applied in various scenarios such as license plate detection, anti-theft monitoring and airport security.

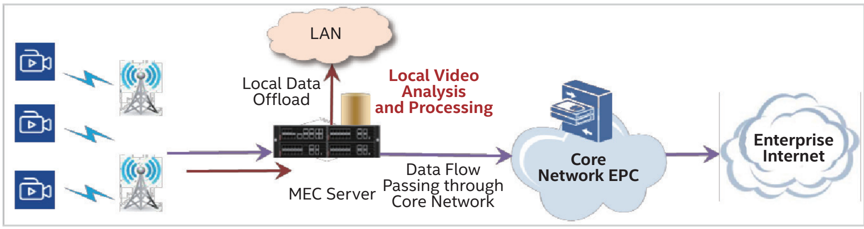

Figure 3.6 Implementing Video Monitoring and Intelligent Analysis based on MEC

(6) V2X Application

The requirements of the 5G network is 20ms for remote vehicle detection and control latency of V2X in uRLLC, while a 5ms or less latency is required for autonomous driving. Edge computing is the enablement technology for reducing latency in the 5G network. As shown in Figure 3.7, the local computing of the LTE cellular network and MEC Internet of Vehicles platform is used to send driver assisted information (such as warning) to an On-board Unit (OBU) in emergency situations. The car-to-car latency can be lowered to less than 20ms, which substantially reduces the reaction time of the driver compared with the latency of existing networks. This has important significance for saving lives and minimizing property loss. In addition, the MEC Internet of Vehicles platform can also achieve path optimization analysis, driving and parking guidance, security assistant information, regional vehicle service guidance, etc.

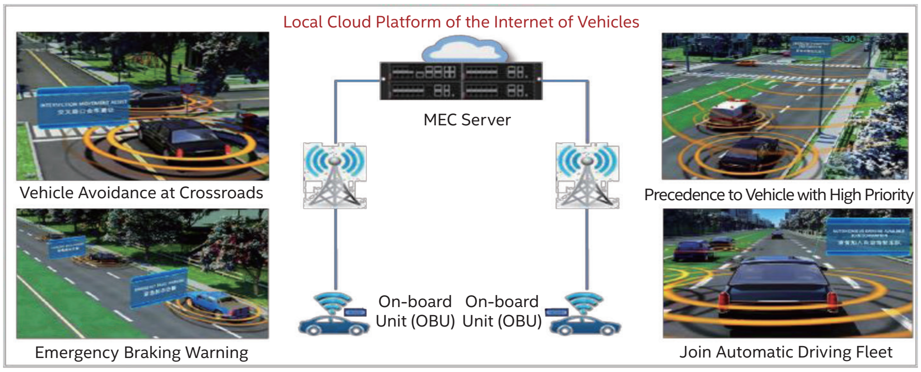

Figure 3.7 Implementing LTE-V2X Application based on MEC Platform

(7) Industrial Control

The rapid development of mobile Internet has led to the increasing demands of industrial parks for wireless communication. Currently, most of the factories/parks implement wireless access through WiFi networks. However, WiFi networks cannot guarantee security authentication, anti-interference performance, best channel utilization rates, QoS or service continuity, making it difficult to meet the industrial requirements. As shown in Figure 3.8, the combination of cellular networks and the MEC local industrial cloud platform can achieve real-time analysis processing and local offload of machine and equipment production data to realize automated production and improve production efficiency in the era of Industry 4.0. As the traditional core network is bypassed, the MEC platform can process and provide feedback on the data collected locally in real time, providing the advantages of high reliability, high security, low latency and high bandwidth.

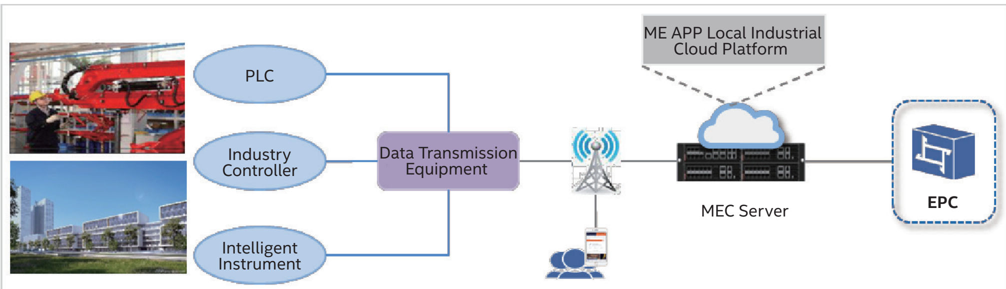

Figure 3.8 Implementing Industrial Control based on MEC Platform

4 MEC Deployment Policy of China Unicom LTE Network

4.1 MEC Networking Architecture of LTE Network

In today's LTE network, a MEC server has two forms: 1) built-in: integrated with the base station as an enhancement function through upgrading software or an add-in board card; or 2) external form, deployed behind the base station or gateway as stand-alone equipment. Figure 4.1 shows the typical MEC end-to-end networking architecture in China Unicom LTE network. The MEC server is located between the base station and core network and implements service offload through parsing S1 messages. There are usually multiple transmission rings between the base station and core network: access ring, convergence ring, and core ring. The MEC server is deployed in a proper position within the network according to the requirements of service type, processing capability, network planning, etc.

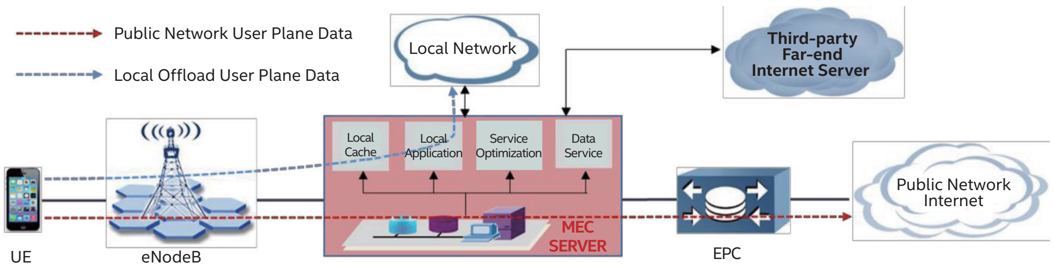

Figure 4.1 Diagram of MEC End-to-end Networking Architecture in LTE Network

The MEC server can run on a physical platform or virtualized platform to provide such functions as local cache, local data service and service optimization. It can also carry local applications. The offload rules of these services are preconfigured for the MEC offload module. When the user plane receives a service data message, MEC parses the feature field of the message (for example, IP quintet) and matches it with the preconfigured offload rules. If the rules are matched, the service flow will be led to the corresponding local application or service, as shown by the blue line in Figure 4.1. In addition, MEC performs a transparent parsing for S1 signaling, which does not affect the signaling process between the base station and core network. For service flows that do not belong to the MEC local service, the MEC transparently transfers the service message received to the core network.

4.2 MEC Deployment Scheme of China Unicom LTE Network

4.2.1 Deployment Location

LTE backhaul networks are relatively closed. The backhaul service traffic is hauled back and converged through a classification of tunnels. Then the traffic interacts with the service node or public network service in the APN network after being uniformly processed by the core network. The use of the LTE network backhaul tunnel is conducive to the QoS guarantee and security policy mechanisms specific to different services in the backhaul process. However, in the scenarios with increasingly prominent localized and regionalized service requirements, the relatively closed backhaul network has a negative impact on the end-to-end quality of services, which requires the backhaul network to perform traffic offload for such services. Meanwhile, the deployment of new services also needs a secure and reliable universal platform. In this regard, a MEC server utilized as stand-alone equipment shall provide the capability of integrated deployment with the existing backhaul network in different scenarios. The typical deployment mode of MEC in China Unicom LTE network is shown in Figure 4.2.

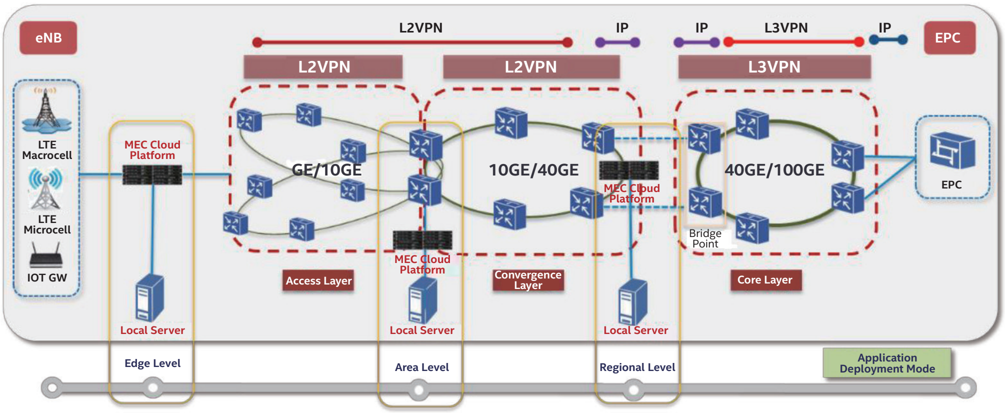

Figure 4.2 MEC Deployment Location

1) Edge level: The MEC server is deployed between the base station and backhaul network. Such deployment is close to the base station (macrocell, indoor cell or Small Cell). The EMC can be deployed in the site computer room, or deployed together with the Cloud-BBU pool in the wireless access computer room. In this deployment model, the MEC covers a small number of base stations and has little influence on the transmission; the backhaul link has the lowest latency and is relatively suitable for local CASH and CDN services. Meanwhile, the coverage performance of this deployment mode has a close correlation with current near-end transmissions, and the coverage requirements and transmission conditions need to be comprehensively evaluated. In this scenario, the MEC server mostly adopts L2 networking mode and needs to provide bypass capability to ensure the continuity and high reliability of services in case of system exceptions.

2) Area level: The MEC server is deployed between the convergence ring and access ring. In this case, MEC shall be deployed at the UNI interface of the transmission equipment which is connected to the two rings, and the base station traffic to be offloaded should be groomed to pass through the MEC. In this scenario, the coverage area of MEC may be the base stations on one or multiple access rings and can be selectively offloaded for different base stations on the rings. This coverage mode has a large coverage area and a low latency, however, the VRF relationship needs to be configured or updated on the transmission equipment specific to the base station to be offloaded. This scheme is suitable for scenarios such as venue, factory and mine field which have a relatively large coverage area.

3) Region level: The MEC server is deployed in the convergence or core layer. This coverage model is mainly designed for large-area offload service, or for the situation that the area to be covered contains an access ring island. This deployment type has a higher latency compared with the other two modes, but it can solve the problem of trans-regional transmission coverage. This mode is mainly adopted for the deployment of industrial services or public services and is conducive to the openness of network capability at the core side. In this scenario, MEC servers mostly adopt L3 networking mode and needs to modify the transmission configuration of the docking element to ensure that the message can be sent to the MEC server and other transmission paths can be selected when the MEC server is inaccessible.

In summary, the deployment principle of different locations depends on multiple factors including the requirements of service deployment, the requirements of coverage area and the status of the network. On the whole, it should be deployed as close to the wireless near-end as possible on the premise that the coverage requirements are met.

4.2.2 Billing Scheme

Currently, 3GPP has no clear and specific standardized recommendation for MEC billing related interfaces. Taking local offload as an example, this White Paper provides only the scheme of S/P-GW and MEC joint billing, which is available for reference by China Unicom for billing when deploying MEC in the LTE network.

1) IP-CAN session is established between terminal and mobile network, S/P-GW assigns Charging ID for IP-CAN bearer and indicates the MEC billing module to perform local traffic billing. The indication message contains the Charging ID, terminal identification (including IMSI and IP address), and terminal location (such as cell ID);

2) MEC billing module receives the indication message and establishes the binding relationship between Charging ID and terminal identification;

3) According to the requirements, the MEC billing module makes local service traffic statistics on the terminal bound to the Charging ID at a regular interval or when triggered by events, and generates Local-CDR. The information provided by Local-CDR includes local area identification, application identification, and traffic information (such as service duration and usage amount);

4) Local-CDR is transmitted between MEC billing module and CG. CG merges the CDRs of the same terminal according to Charging ID and S/P-GW address.

The CDR merged by CG includes the outgoing remote service traffic through S/P-GW and MEC local traffic. Based on two statistical dimensions, i.e., terminal and edge application, BOSS system can implement the billing function for MEC local traffic.

4.3 Problem Analysis of MEC Deployment

ETSI defines the function of MEC standards, while the definition of concrete implementation is incomplete. There was no standard interface established with the 3GPP elements in the network and the commercial deployments are faced with the following challenges:

- **Billing:** At present, there is no complete traffic billing scheme in the existing network application, further research and evaluation need to be done for statistics and report of local traffic through MEC, the newly-added node at the core network side (or P-GW upgrade) is responsible for the scheme of generating the CDR and reporting to BOSS system;
- **Security:** The security of the MEC platform is a prerequisite for the deployment of third-party applications, and further research needs to cover physical port isolation, logic port isolation, firewall security control and access control;
- **Lawful intercept:** Providing the listening and monitoring function at the user level should be considered when deploying MEC;
- **Mobility management:** There is no well-validated mobility scheme, and the continuity of service (between MEC servers) needs to be ensured in the handover scenarios.

5 MEC Technology Evolution Route and Planning

5.1 Evolution of 5G Network-oriented MEC Key Technology

5.1.1 Traffic Grooming Scheme

5G core networks can implement local traffic grooming through a flexible session management mechanism of System Management Facilities. 5G networks can adopt "uplink classification" functions and IPv6 Multi-Homing to implement local traffic offload.

1) Uplink classification (UL CL) scheme: As shown in Figure 5.1, the addition and deletion of UL CL is determined by SMF according to the terminal location in the process of handover. When the terminal moves into a MEC coverage area, SMF adds UL CL function and a PDU Session Anchor through a N4 interface, completing the creation for a local traffic path. SMF can introduce multiple UPFs that support UL CL functions to the data path of a PDU session. PDU Sessions can be IPv4 or IPv6, and UL CL implements offload by recognizing the transmission characteristics information of service flow.

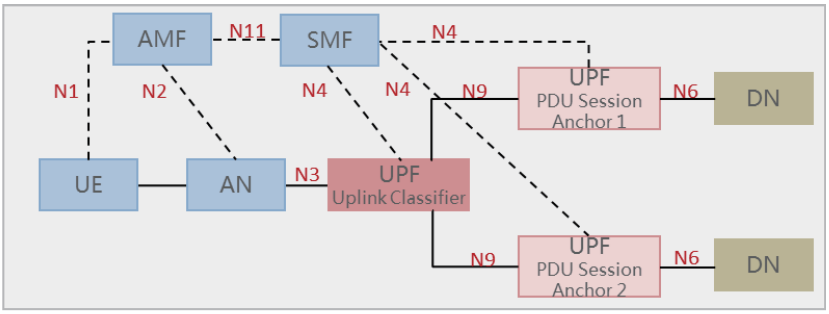

Figure 5.1 "Uplink Classification" Scheme

2) IPv6 Multi-Homing scheme: As shown in Figure 5.2, the creation of local service Anchors is completed through the addition and deletion of Branching Point in the Multi-Homing scenario while the offload function is implemented. SMF controls UPF function through a N4 interface. For IPv6 type sessions, the local traffic to be offloaded is groomed to the local Anchor through a Branching Point. PDU session can be associated with multiple IPv6 prefixes to provide multiple IPv6 PDU anchors for accessing data network (DN).

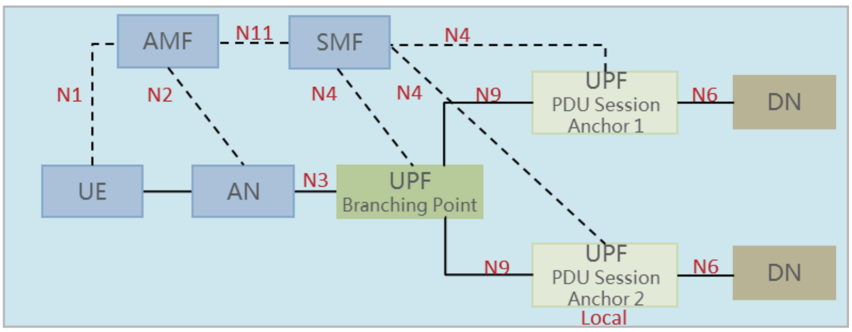

Figure 5.2 "IPv6 Multi-Homing" Scheme

5.1.2 Service Continuity Scheme

To support the session and service continuity (SSC) in the mobility scenario, the 5G network provides three different SSC modes.

1) SSC Mode1: In the movement process of the UE, the Anchor UPF at the establishment of PDU session remains unchanged regardless of the access technology adopted by the UE. This mode is similar to the mode in which the PDN anchor does not change in the LTE network. The UE IP remains unchanged at this time.

2) SSC Mode2: When the terminal leaves the service area of the current UPF, the network triggers the release of the original PDU Session and indicates the UE to immediately establish a new PDU session with the same data network. When establishing the new session, a new UPF may be selected as an Anchor UPF of PDU session, while the UE IP of newly-established Session information needs to be identical to that of the original Session information.

3) SSC Mode3: When the terminal leaves the service area of Anchor UPF, the original PDU Session and Anchor UPF persist. Meanwhile, a new Anchor UPF needs to be selected for establishing a new PDU Session. At this time, the UE simultaneously has the PDU Session of two Anchor UPFs, and the original PDU Session is finally released. The UE IP remains unchanged in this process.

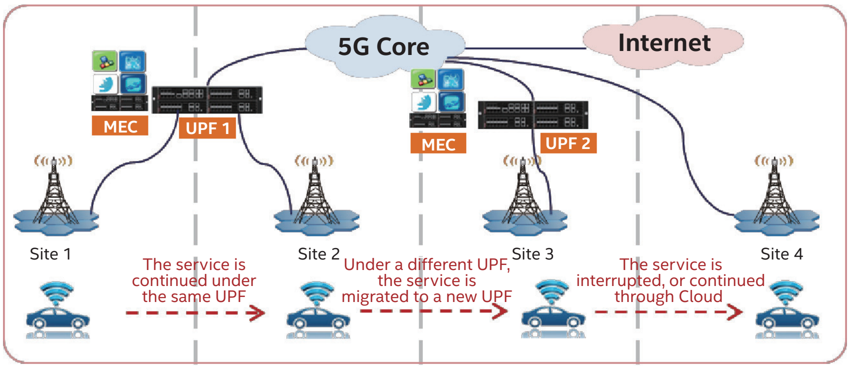

Figure 5.3 Schematic of 5G Network Session and Service Continuity

SSC mode selection policy is configured according to the operator network, and the UE can select a proper SSC mode for an application or a set of applications. In this policy, a default SSC mode can be configured for all applications. If the UE does not select a SSC mode for an application, the network can select a proper SSC mode for the application according to the subscription information, local configuration and application request, so as to support the service continuity of edge computing. As shown in Figure 5.3, the UE moves to the UPF1 coverage area, the 5G core network adopts SSC Mode1 and maintains the continuity of local offload service through uplink classification or IPV6 Multi-Homing mode. When the UE moves to UPF2 coverage area, the 5G core network adopts SSC Mode3 and migrates the service to a new UPF2 without interruption of service. When the UE leaves the MEC coverage area, the 5G core network adopts SSC Mode2; the service is interrupted, or continued through the Cloud.

5.1.3 Intelligent Perception and Optimization Scheme

MEC will be deeply integrated with 5G network architecture. Its functions such as service offload, policy control and QoS guarantee will be implemented through standard 5G network functions. As shown in Figure 5.4, the edge computing application (ME APP) performs real-time interaction with the 5G network through Network Exposure Function (NEF). On the one hand, NEF transmits the perceived measurement information related to the UE and service flow, such as UE real-time location, wireless link quality and roaming state, to MEC server; MEC server optimizes the service performance of the application through intelligent analysis and abstraction based on the above-mentioned measurement information (for example, to adjust video playback bitrate), thus improving the QoS.
On the other hand, NEF transmits the perceived information related to the application service, such as service duration, service life cycle and mobility pattern, to the network. The network perceives the information provided by the analytical application and further optimizes its UE resource configuration (for example, to allocate proper bandwidth resources for VIP users) and session management.

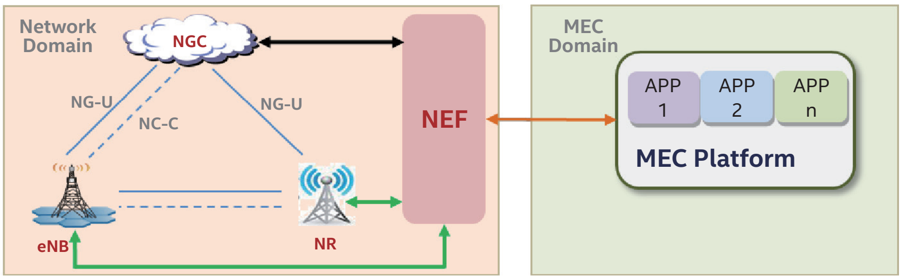

Figure 5.4 5G Intelligent Perception and Optimization Scheme

For various types of MEC applications, the users can query the status of the network through NEF, such as billing policy and network capability, or submit various task requests, such as location and bandwidth.

5.2 Evolution of China Unicom MEC Networking Architecture

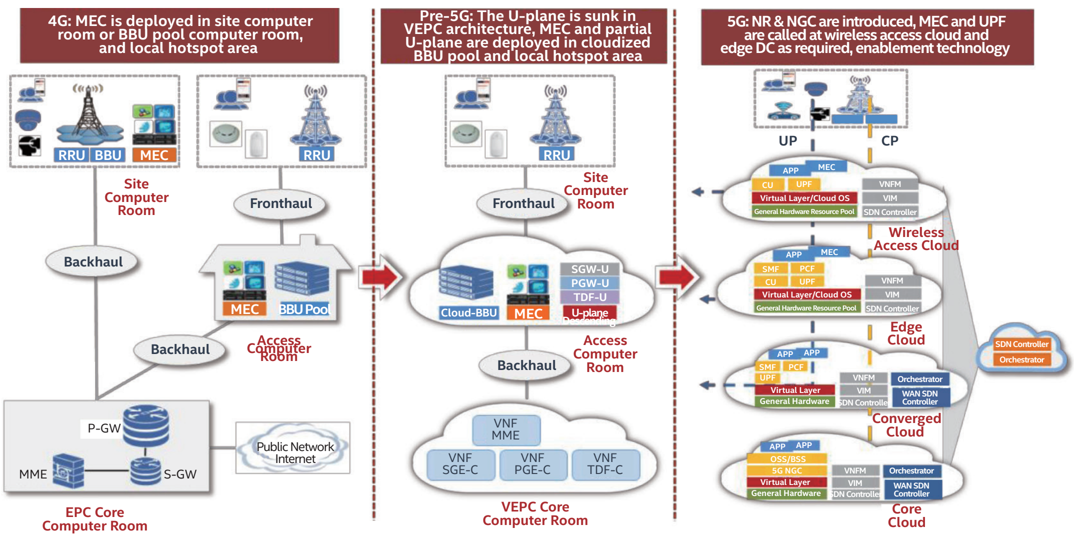

Figure 5.5 Evolution Vision of China Unicom MEC Networking Architecture

With the rise of services such as AR/VR, HD video, and the Internet of Vehicles, the users put forward higher requirements for latency and bandwidth, which leads to the deployments of business service to the network edge. MEC provides an edge computing environment for such service deployments. MEC is a kind of 5G native technology, but it can be applied and deployed in the LTE network due to the openness of its architecture. The vision of China Unicom MEC networking architecture is shown in Figure 5.5.

1) 4G phase: MEC implements the service data offload through transparently parsing the message of the S1 interface located between the base station and core network. In hotspot areas, MEC and the base station can be simultaneously deployed at the same location at the base station access side. After passing through the base station, the local service is offloaded by the MEC platform to the local server to avoid the impact on the backhaul network when the hotspot handles large traffic. In addition, in the scenario of base station BBU pooling, MEC and the base station BBU pool can be simultaneously deployed in a wireless access computer room to improve the service reusability of MEC and reduce the resource requirements for MEC equipment.

2) Pre-5G phase: As the NFV and SDN technologies become mature, the access computer room implements virtualization transformation, and provides the virtual resources required for the operations of base station, core network and services. After the completion of base station BBU virtualization, the Cloud-BBU is formed and deployed in the data center of the access computer room. The core network elements completed the virtualization and the separation of control forwarding, the gateway node was separated into control plane elements SGW-C, PGW-C and TDF-C, and forwarding plane elements SGW-U, PGW-U and TDF-U. According to the need, forwarding plane elements SGW-U, PGW-U and TDF-U are deployed in the data center of the access computer room to form a pre-5G network architecture. In this architecture, the network interface adopts a 4G standard interface, the network architecture embodies the characteristics of 5G network, and the control plane and forwarding plane elements can be evolved to a 5G network architecture through a software upgrade protocol. In the pre-5G phase, MEC, which is based on 4G form and supports virtualization deployment, is deployed in the data center of access the computer room together with Cloud-BBU and core network forwarding plane elements.

3) 5G phase: Taking the data center as infrastructure and cloud computing as infrastructure platform, the 5G network architecture forms four cloud centers from the access layer to the core layer: wireless access cloud, edge cloud, converged cloud and core cloud. The full virtualization and service orientation have been completed for core network element functions which are deployed in the cloud center as required. In addition, the CU/DU architecture is adopted at the wireless side. Based on virtualization technology, the CU can be deployed in the wireless access cloud and edge cloud.

In the 5G network architecture, the UPF provides offload function and PCF implements offload policy control. 5G MEC is deeply integrated with 5G network compared with 4G MEC. The offload rules required by MEC services are informed to PCF through interfaces. PCF configures SMF with the offload policy which would be sent by SMF to the base station and UPF, and then the offload function is finally implemented by the UPF, as shown in Figure 5.6. In the 5G network, the UPF can be deployed in different positions within the network as required, so as to offload the service to the MEC platform. For example, the UPF and MEC can be deployed in the access cloud for AR/VR services with higher requirements for latency; HD video services can be deployed in the edge cloud or higher-level converged cloud to improve the hit rate of the services.

In addition, based on a 5G network architecture, the offload function and policy function of 5G MEC are standard UPF and PCF network functions. Therefore, 5G MEC can fundamentally solve the existing billing and policy problems of 4G MEC and truly implement the commercialization of MEC. The computing capability and platform architecture provided by 5G MEC can be coordinated with flexible virtualization deployment to construct an open network edge ecology environment while providing the users with a high-quality mobile communication service experience.

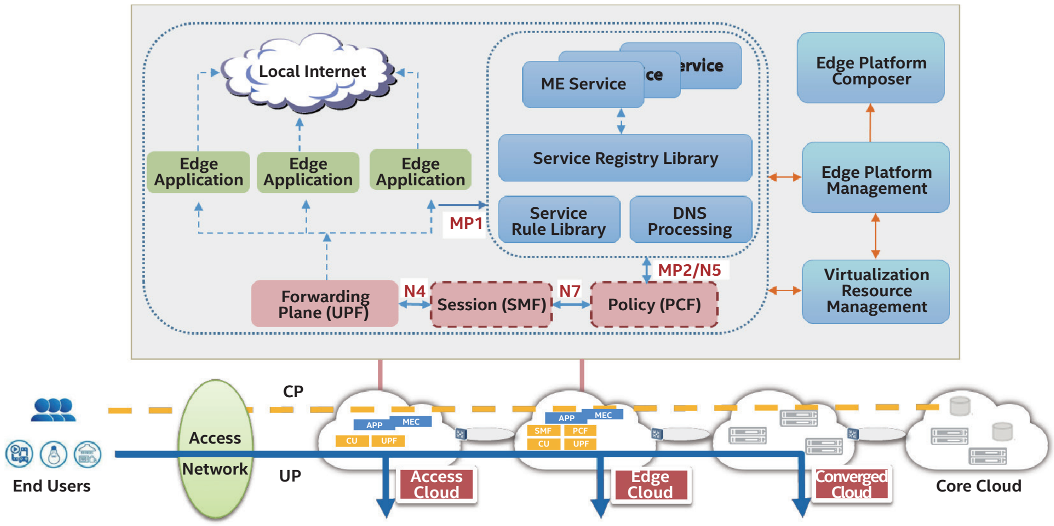

Figure 5.6 Converged Architecture of 5G Network and MEC

5.3 China Unicom 5G Network MEC Deployment Planning

In the 5G network, the deployment location of MEC is closely related to the service scenario, and MEC can be deployed in the wireless access cloud, edge cloud or converged cloud as required. In general, in uRLLC low latency scenario, MEC can be deployed in the wireless access cloud in close proximity to the base station. In the hotspot area with large traffic under eMBB scenario, MEC can be deployed in the edge cloud. In the mMTC scenario, MEC can be deployed in the converged cloud at a higher location to meet the service requirements of a larger coverage area.

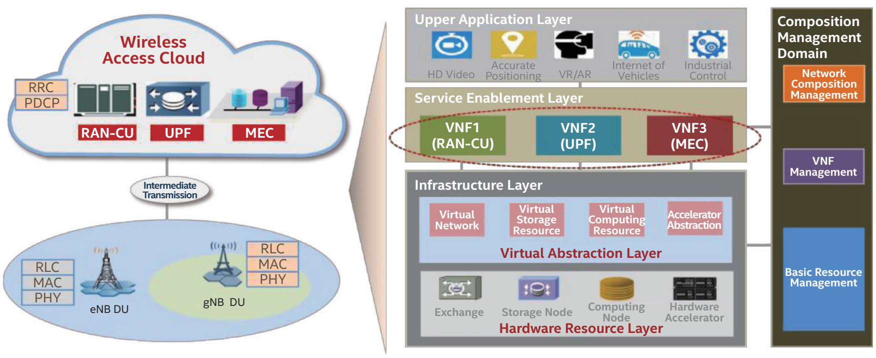

Figure 5.7 Deployment of 5G Network MEC in Wireless Access Cloud

- **Scheme 1: Wireless Access Cloud**

The wireless access cloud is located at the wireless side; MEC, base station CU, and core network forwarding plane UPF are deployed in the wireless access cloud, as shown in Figure 5.7. Local services are deployed at the base station side to provide the users with lower latency (< 4ms) services. The CU consists of RRC and PDCP layers, and the DU consists of RLC, MAC and PHY layers. The CUs of 4G eNB and 5G gNB CU can be jointly deployed. In this mode, MEC service has a relatively small coverage area and is suitable for latency-sensitive services which move at a low or zero speed, such as AR/VR services and local services (for example, stadium, venue, and scenic spot).

- **Scheme 2: Edge Cloud**

MEC and core network forwarding plane UPF are deployed in the edge cloud and are connected with 5G gNB CU through the transmission access ring (the CU is deployed in the wireless access cloud). Alternatively, the MEC, UPF and CU are deployed in the edge cloud and are connected with the DU through the Midhaul network. Compared with the deployment in wireless access cloud, MEC provides a slight increase in latency when it's being deployed in the edge cloud, while the coverage area of MEC service is expanded. Therefore, the services with strict latency requirements, such as the Internet of Vehicles, can be deployed in this scenario.

- **Scheme 3: Converged Cloud**

The converged cloud is usually located in prefecture-level cities and has a large coverage area and numerous types of services. The deployment of MEC and core network forwarding plane UPF in the converged cloud can give full access to the advantages of the MEC platform and provide various services. This scheme is suitable for the scenarios involving wide coverage, large traffic and massive connection.

In summary, the MEC can be flexibly deployed based on the three typical scenarios defined by ITU, under the unified control of the management composition system and by taking into consideration service latency and service bandwidth requirements. For services such as CDN, the same service can be deployed at different levels. If the end user hits the MEC of the wireless access cloud, the service is provided by the MEC of wireless access cloud; if the end user hits the edge cloud or converged cloud, the service is provided by the edge cloud or converged cloud.

6 Summary and Prospect

For more than a decade, the innovation of mobile technology has continuously boosted the development of the mobile industry. Compared to 4G, 5G has become the primary productivity of the society and is about to implement industry-wide digitization. On the basis of improving the user experience of mobile Internet services, 5G will further satisfy the massive requirements of IoT applications and deeply integrate with such industries as the Internet of Vehicles, industrial control, telemedicine and energy, so as to implement the "Internet of Everything" in a real sense.

International standard organizations such as ETSI, 3GPP and NGMN have defined MEC as a part of the 5G architecture and have made strides in research and specifications. In the 5G era, the applications of MEC will extend to such fields as transportation system, intelligent driving, real-time tactile control and Augmented Reality to become the key enabler for the digital transformation of operators. This will bring about a transformation of the network from access pipeline to information-based service enablement platform. The development of the MEC industry standards and the deployment of general virtualization platforms will provide a new network ecosystem and value chain. MEC provides operators with a platform for third-party application registration and management. The platform can quickly integrate third-party applications and achieve the openness of network capabilities to facilitate the service customization in the vertical industry and flexible release of third-party applications, and help operators develop new services.

Regarding the construction of the MEC industry ecology, China Unicom has been making active efforts together with industry partners to discuss, test and commercialize new services. We sincerely invite all mobile operators, telecom equipment manufacturers, OTT providers, application developers, as well as research institutes and enterprises that pay close attention to the development of MEC to participate in the research of MEC technology standards, platform development and existing network validation. Together we can explore new service scenarios, discuss business cooperation models, and promote the development of the MEC industry.