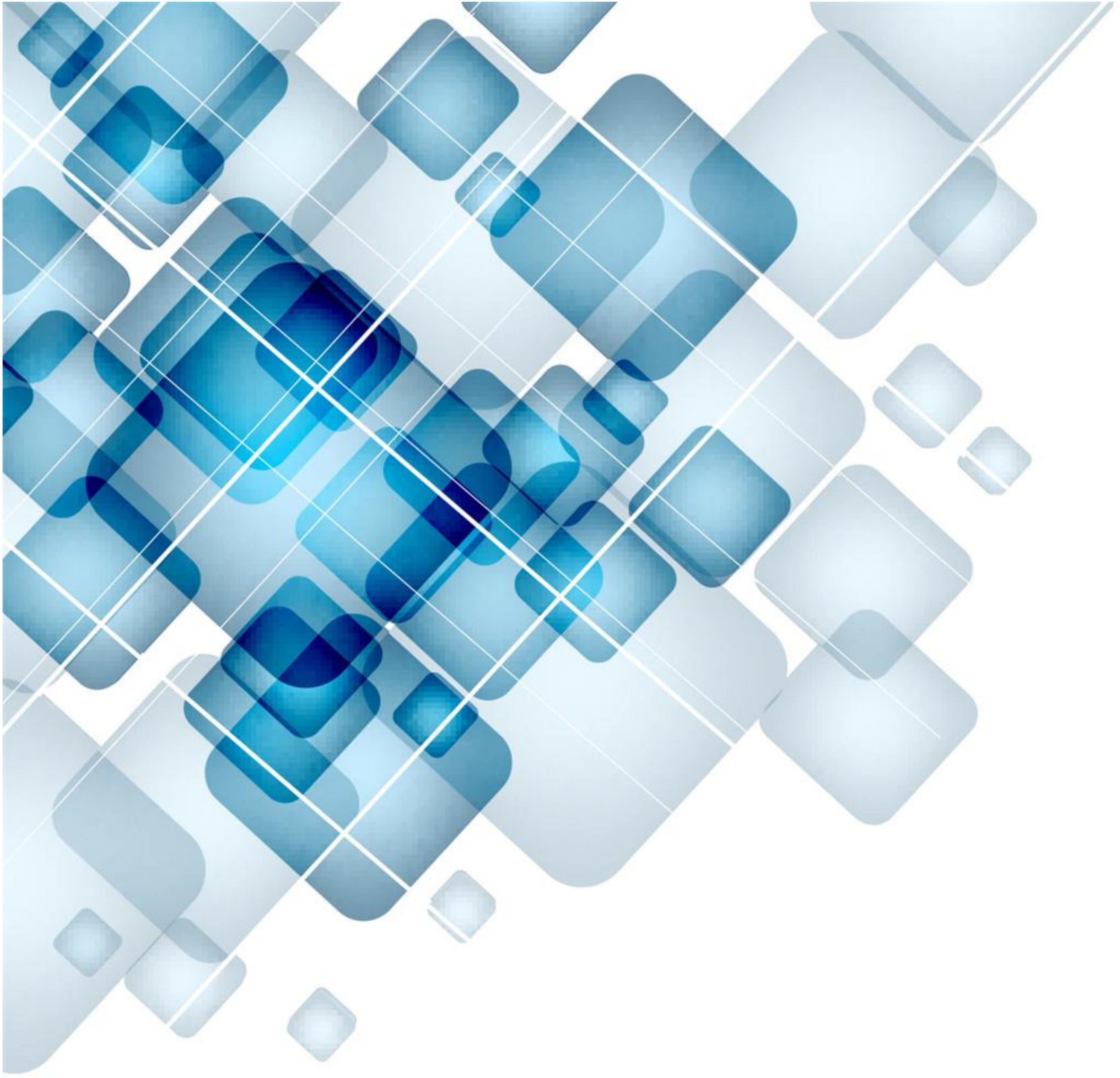


China Mobile Edge Computing Technical White Paper

China Mobile Edge Computing Open Laboratory

Table of Content

1. Background of Edge Computing Development

1.1. Requirements & Use Cases

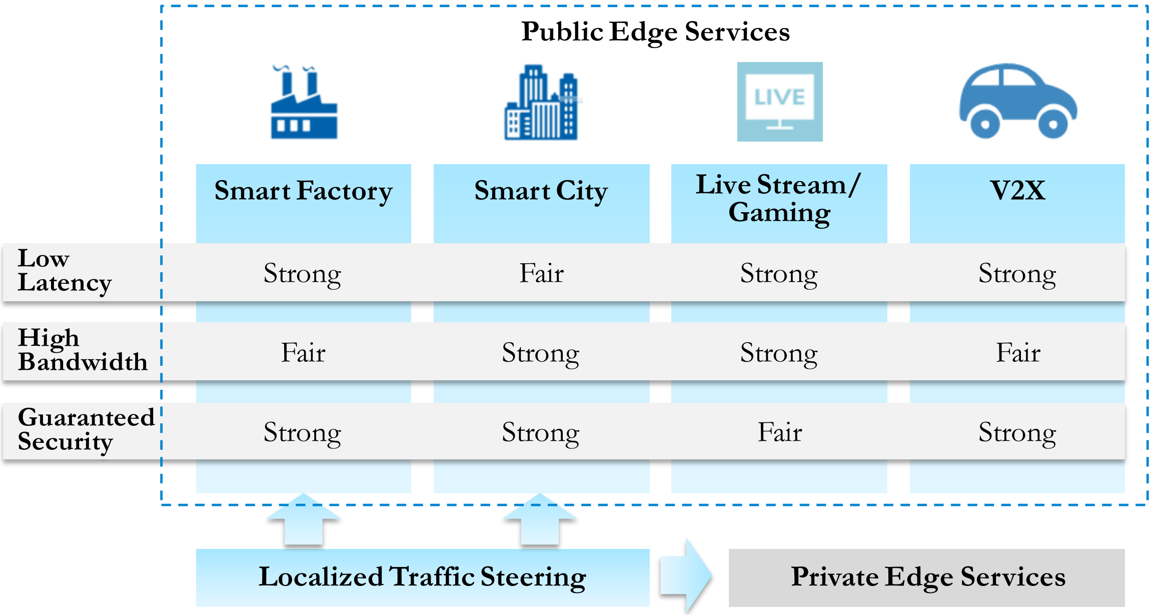
Edge computing provides computing, storage and other infrastructures close to data sources and users. It also provides cloud service with IT environment for edge applications. Compared with centralized cloud computing, edge computing solves the problems of high latency and large aggregated traffic, and provides better support for real-time and bandwidth-demanding applications. With the rapid development of 5G and industrial internet, many emerging services urgently demand for edge computing.

Among those services in various vertical industries, latency, bandwidth and security are the 3 essential technical requirements for edge computing. At present, the four verticals including intelligent manufacturing, smart city, live streaming/gaming and V2X have the most confirmative demand for edge computing.

In the field of intelligent manufacturing, edge computing gateway is used for local data collection, data filtering, cleaning and other real-time processing. Meantime, edge computing can also provide the ability of cross-layer protocol translation and aggregated WAN access for fragmented industrial LAN. At the same time, many factories are investigating the implementation of industrial controllers based on virtualization technology. This realizes the centralized and collaborative control of the mechanical equipment. Similar to the separation of data plane and control plane in SDN, edge computing is enabling the separation of mechanical plane and control plane with the concept of software-defined-machinery.

In the field of smart city, the main applications are concentrated in smart buildings, logistics and video monitoring. Edge computing realizes on-site collection and analysis of various operational parameters of the building and provides the ability of predictive maintenance. It also enables the monitoring and pre-warning of vehicles and goods transported in cold chain. Furthermore, image processing such as face recognition and object recognition can be realized in millisecond with localized GPU servers.

In the field of live streaming/gaming, edge computing provides rich storage resources for CDN. It also helps with the audio and video rendering process closer to users, making services such as cloud desktop and cloud gaming easier to be established. Especially in AR/VR scenarios, the introduction of edge computing can greatly reduce the complexity of AR/VR terminal device.



Public Edge Services

	Smart Factory	Smart City	Live Stream/ Gaming	V2X
Low Latency	Strong	Fair	Strong	Strong
High Bandwidth	Fair	Strong	Strong	Fair
Guaranteed Security	Strong	Strong	Fair	Strong

Figure 1 Typical use cases and requirements of edge computing

As a result, the overall AR/VR industry can be promoted more rapidly with significantly reduced terminal costs.

In the field of V2X, the demand for low latency is extremely strong. Edge computing can provide millisecond-level latency guarantee for assisted driving such as anti-collision. It also supports data processing and analysis of high-precision map at the base station, by providing localized computing resources. This can better support pre-warning services for the blind area with line of sight vision.

In addition to the use cases in the listed vertical industries, edge computing is also popular in a special scenario - local private network. Many enterprise users require the operator to provide local breakout to private campuses, and directly divert the traffic of local services to private data center for further process. For example, local area network and file sharing can be achieved in university campus, localized ERP services can be achieve by diverting the traffic to local private cloud, and local data storage can be provided for public services such as library and hospital. In these scenarios, operator provides a dedicated local breakout broadband service for the users.

1.2. Mutual Promotion between Edge Computing and 5G

The three typical application scenarios of 5G networks are closely related to edge computing. The ultra-high reliability and low latency communication of URLLC, the high bandwidth of eMBB and the large connections of MIoT all have the requirement of edge computing. Therefore, edge computing is an inevitable element for 5G. It is one of the most important trend for network evolution and also the key for 5G to successfully provide services in vertical industries.

5G network realizes the local breakout of data traffic through the flexible deployment of the user plane function (UPF) at the edge of the network. UPF is managed by the 5G core network control plane, and corresponding routing policy is configured by the 5G core network. The 5G network also introduces three service and session continuity modes to support edge computing, ensuring user experience with high mobility, such as in V2X scenarios.

5G network capability exposure extends corresponding APIs to edge applications. APIs such as wireless network information services, location services, and QoS services have been defined in the edge computing system. This information can be provided to application by edge computing PaaS platform after encapsulation.

The flexibility of UPF deployment makes it simple for 5G network to be connected with edge computing resources, which further accelerated the development of edge computing. At the same time, edge computing set the technical foundation for the support of new 5G application with low latency, high bandwidth and massive connection requirements。

1.3. Edge Telecom Integrated Cloud

The architecture of China Mobile Telecom Cloud takes full account of the characteristics of telecommunication network functions, including both control and user planes. Control plane functions are suitable for centralized deployment with homogeneous demand for resources. User plane functions are better to be deployed closer to UE for improved user experience. The distributed user plane function is one of the key enablement for edge computing services. With the rise of edge computing, the demand for distributed user plane functions continues to grow. These network elements have new requirements for latency, storage, forwarding performance, computing density, lifecycle management efficiency for different edge computing services.

In order to meet the requirements of telecommunication services, the architecture of China Mobile telecom cloud includes two levels: the core cloud and edge cloud. These two types of cloud resources can cover a variety of data centers/equipment room from the centralized core nodes distributed district-level nodes. The edge telecommunication cloud is an important part of China Mobile telecom cloud architecture, serving both media and forwarding plane network functions. The edge telecom cloud can be deployed at city and district levels. In the near future, it can be extended to even lower level according to service requirements.

The location for deployment of edge computing and edge telecom cloud has many overlaps. Edge telecom cloud with virtualized UPF deployed can provide local breakout service to edge computing nodes. The internet services that are hosted in edge computing have many different requirements compare with the virtualized network functions that are hosted in telecom edge cloud. Edge computing will hugely benefit from the success development of edge telecom cloud in the phase of NFV deployment.

2. China Mobile's Perspective on Edge Computing

2.1. Location of Deployment

Compared with traditional cloud computing, edge computing is deployed closer to users, but the understanding of edge computing deployment locations are different for various industries. Taking the end user of operational technology (OT) field as an example, mainstream companies are exploring the solutions for intelligent upgrading of on-site devices, enabling edge computing services by deploying SDK to on-premise devices. Meanwhile for the Internet companies, besides on-premise devices, edge computing with slightly higher deployment location also draws much attention. Edge computing provides a localized centralization of services, resulting in better resource sharing, real-time performance and cost-saving on bandwidth.

Considering the characteristics of operators' end-to-end infrastructure construction and service development, from the perspective of deployment location, China Mobile categorize edge computing nodes roughly into network-side and on-premise edge computing. Network-side edge computing is deployed in the equipment rooms at city or lower level. Most of these nodes are in the form of cloud, which are mini data-centers. On-premise edge computing is deployed at the access points of the operator network. These nodes are generally located in end user sites, in which no equipment room can be used. These nodes are the first hops for users to connect with the operator network. The typical form of this type of device is CPE such as edge computing intelligent gateway. It should be pointed out that a cellular base station, although recognized as an access point, is considered as a network-side edge computing node since it is deployed in the operator's equipment room.

2.2. All-access Computing Plane

The core of edge computing is to build distributed IT resources that are more versatile, flexible, supporting multiple application ecosystems. The access to edge computing includes 4G/5G, WIFI and FTTx. For a specific edge computing node, services with different accesses can share the same IT resource.

Over the past two decades, China Mobile has created a remarkable network infrastructure plane that covers both wireless and fixed connectivity. The evolution of NFV technology also prompted China Mobile to constructing telecom cloud facilities that host the virtual network elements. Facing the future industrial Internet, artificial intelligence and other emerging technologies, operators need to build an end-to-end computing plane with edge computing on top of the network plane, forming the coverage of

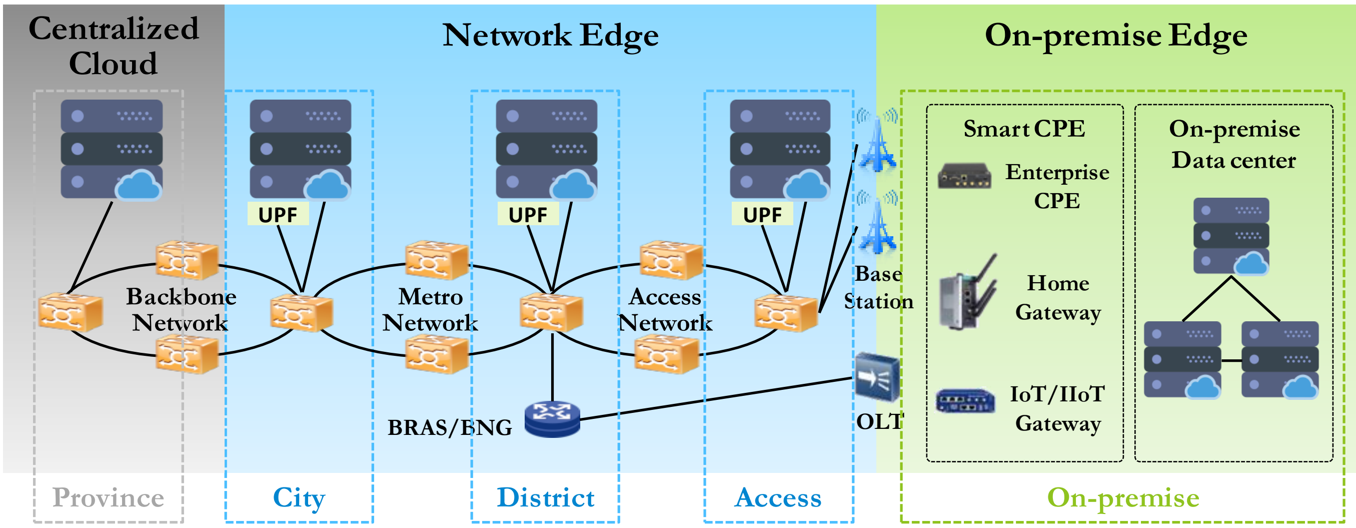


Figure 2 The locations of edge computing in end-to-end network

computing resources. This provides intelligent connectivity infrastructures for vertical industries. On this computing plane, the ubiquitous on-site edge computing nodes provide users with intelligent access and real-time processing of data, realizing flexible access to services. This enables a data ecosystem for edge computing. The network-side edge computing provides abundant computing resources close to end users, realizing the artificial intelligence, image recognition and other new services. This enables the application ecosystem for edge computing. Network resources and edge computing resources converge and integrate, providing extraordinary user experience for vertical applications.

2.3. Technical System of Edge Computing

The edge computing technology system involves multiple areas. In particular, it can be divided into applications, PaaS capabilities, IaaS facilities, hardware devices, sites planning, and edge network evolution. For various locations of the edge computing deployment, customized technology choices are expected in the above areas.

In terms of applications, edge computing generally inhabits two types of ecosystems. One is the mature services that have been deployed in public clouds. With the growth of user volume and requirement for real-time experience, it is emerged to migrate to edge computing. Such applications tend to have a strong dependence on the original public cloud ecosystem, and at the same time face the technical difficulties of edge-cloud collaboration in the aspect of network, data and functional logic. The other type is the edge-native applications, which require the use of edge computing resources due to the requirements of latency, bandwidth and security. Such applications are less dependent on public clouds, but the ecology is still immature and fragmentation is more serious. The PaaS, IaaS and hardware platforms for edge computing need to be designed to be compatible with both of the application ecosystems.

PaaS, IaaS and hardware platforms are the key enablement in the technical system of edge computing. In terms of PaaS, operators can use the unique advantages of their own network resources to provide various types of featured network capabilities for upper-layer applications running on the PaaS platform. Third-party PaaS platforms are equally important, because in some specific vertical industries, third-party partners have a deeper understanding of industry service logic and can provide PaaS capabilities which solve essential problems in the edge computing ecosystem. Among some of the edge nodes deployed on customer sites, a lightweight PaaS

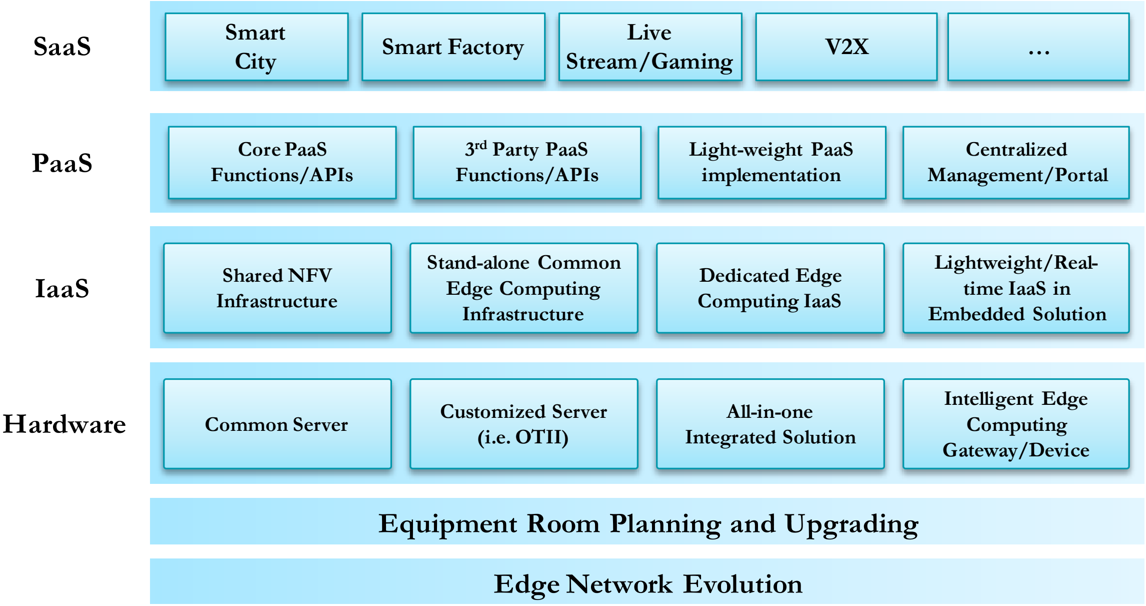

SaaS
Smart City
Smart Factory
Live Stream/Gaming
V2X
…
PaaS
Core PaaS Functions/APIs
3rd Party PaaS Functions/APIs
Light-weight PaaS implementation
Centralized Management/Portal
IaaS
Shared NFV Infrastructure
Stand-alone Common Edge Computing Infrastructure
Dedicated Edge Computing IaaS
Lightweight/Real-time IaaS in Embedded Solution
Hardware
Common Server
Customized Server (i.e. OTII)
All-in-one Integrated Solution
Intelligent Edge Computing Gateway/Device
Equipment Room Planning and Upgrading
Edge Network Evolution

Figure 3 Technical system of edge computing in China Mobile's perspective

needs to be provided. At the same time, the edge computing PaaS platform must also have the ability to be centrally managed. In terms of IaaS, edge computing needs to consider the sharing and integration with NFV infrastructure in some cases, but also the cases where it is independently deployed. In terms of hardware, considering the conditions of the edge computing sites, it is necessary to redesign and customize the layout of the server. For different vertical applications, it is also important to be capable of delivering all-in-one equipment and all types of on-premise intelligent access devices.

In this version of the white paper, we will introduce and illustrate the edge computing PaaS platform, IaaS infrastructure and customized hardware. In future update, we will further elaborate on the transformation of the edge network sites and the edge network related technologies according to the development of new services such as 5G.

2.4. Security in Edge Computing

Security is the key element of edge computing. Firstly, effective security mechanism can avoid the impact of introducing edge computing applications on operators' networks and services; Secondly, only ensuring the security of edge computing in terms of technology and management mechanism can elimite the concerns of third-party applications deployed on edge computing platforms; Moreover, third-party applications deployed on edge computing platforms usually need general security capabilities (e.g. firewall, IDS/IPS, WAF etc.). Perfect network and information security mechanism is the prerequisite for the healthy development of edge computing industry and ecology.

Compared with the traditional operator network, the edge computing system has great changes in network architecture, service delivery mode and operation mode, which pose greater challenges to security. In terms of network architecture, edge computing nodes are closer to users and are more likely to suffer from physical attacks. Edge computing nodes deploy core network elements (e.g. UPF) and communicate with core network data plane gateway, which enlarge the attack surface of core network. In terms of service provision, hosting multiple third-party edge computing applications needs to be well isolated between application and application, application and network element. n terms of operation mode, the experience in collaborative management and operation of edge computing platforms and third-party applications still needs to be accumulated. In addition, exposing network capabilities to third parties involve security issues such as network security, service and user data security, user privacy management and control. Edge computing security should be implemented from the following aspects.

2.4.1. Physical Security

Edge computing nodes, especially deployed in unattended DC or user side, are in open environments that are not controlled by operators and are more likely to suffer from physical attacks in relatively. Infrastructure such as network, electricity and air conditioning should be fully taken into account when selecting deployment location and device types to ensure high availability of devices. In addition, the technology and management mechanisms should also be strengthened to avoid theft and information leakage.

2.4.2. Platform Security

The edge computing platform is based on cloud infrastructure. It should consider the virtualization software security, the virtual machine/container security and the data transmission security when management software is deployed remotely.

Operators' network elements such as UPF and edge computing applications should be co-located. Physical security, data security and access control of UPF should be considered to prevent edge computing applications attack the core network through UPF.

After the multi-tenant edge computing application is deployed in the edge computing node, it should provide isolation between tenants and applications and distinguish between tenants' service operation and security management to avoid data theft and user privacy data leakage.

2.4.3. Application Security

In order to evaluate the security of third-party applications, appropriate security assessment control and accreditation should be implemented when the applications are deployed and upgraded.

Edge computing applications should be integrated into security management processes such as security compliance checking and auditing, exposed assets management and virus scanning, so as to avoid the security problems of other applications on the edge computing node caused by the application's own security vulnerabilities.

The content security and information security issues carried by third-party applications should also be controlled.

Reference implementations, capabilities and service models related to high availability should be provided in the edge computing architecture to ensure the availability of applications.

2.4.4. Capability Exposure and Security

The capability exposure of edge computing involves not only user data (such as user location data, behavior preference data, etc.), but also network information of wireless network and core network. It should be controlled in authentication, authorization, monitoring and other aspects. It should implement hierarchical management of capability exposure, match with information security and application requirements, control authorization granularity and measurement means, avoid excessive authorization and abuse of authority, and implement related security audit.

Security capabilities (such as anti-DDoS, IDS, WAF, etc.) are also an important part of capacity exposure. The traditional method based on deploying security devices and setting static security strategy is no longer applicable for edge computing. Therefore, we should consider the virtualization and multi-tenancy of tenant's general security capabilities, and support customized configuration and orchestration under the unified control of security services to meet the customized security requirements of applications.

3. Edge Computing PaaS Technology

3.1. PaaS Overview

Edge computing provides PaaS layer services, which can be a value-added service for telecommunication service provider and reduce the difficulty of application migration. The PaaS platform for edge computing differs from the PaaS platform for public/private clouds. The edge computing data center is too small to deploy all PaaS platform capabilities as a whole. The PaaS capability should deploy in an on-demand manner for various implementation scenarios.

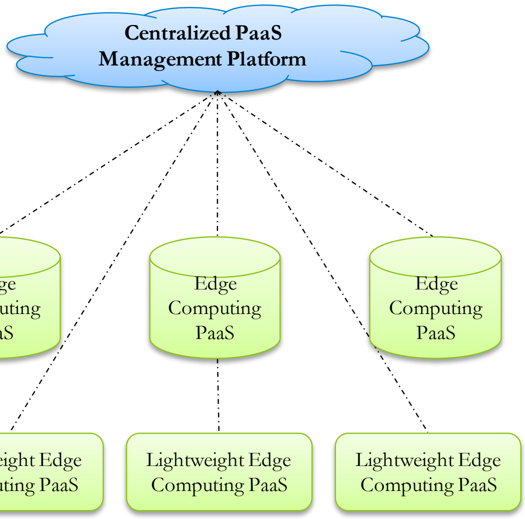

Centralized PaaS Management Platform
Edge Computing PaaS
Edge Computing PaaS
Edge Computing PaaS
Lightweight Edge Computing PaaS
Lightweight Edge Computing PaaS
Lightweight Edge Computing PaaS

Figure 4 Edge Computing PaaS Platform

The PaaS platform can be divided into three layers for different implementation scenarios:

1) **Centralized edge PaaS management platform**

- Manages thousands of edge data centers and tens of thousands of edge gateway platforms
- Provides a unified portal for users and managers
- Displays the number of data centers, resource usage, service running status etc.

2) **Edge data center PaaS**

- Provides application operation and maintenance environment
- Provides maintenance tool kits
- Unified deployment porta
- Provides vertical industry SDK and capability Exposure
- Reports corresponding resource status information and service running status to the management platform.

3) **Lightweight PaaS for edge intelligent gateways**

- Heterogeneous access for terminal devices
- Data acquisition and conversion capabilities

PaaS platform solves the following issues:

1) **Application deployment**

Depending on the location where the applications are deployed, there are two categories of applications. One is for zone-specific: industrial parks, factories, applications only run in this specific area, etc. The other is for common deployment: video, game, etc. These applications for common deployment can be deployed generally in the area with high population density. Generally, the edge computing datacenters in for zone-specific applications are normally centralized, whilst the ones for common deployment distributed. It is very challenging for application developer to manage so many edge data nodes. Thus, it is required that one can deploy and manage the application automatically from a unified portal.

2) **Service activation**

A customer can't access the application right after an operator deploys the application in the edge computing node due to reconfiguration of the network function, such as 5G UPF. Only telecommunication operator has the permission to manage network function. The general process is to provide the requirements of application to centralized edge computing management platform. The platform will check the requirement and send it to the management plane of the network function. If the application needs to change the DNS records at the edge, it should follow a similar process.

3) Introducing wireless capability and core network capability

In general, wireless capabilities and core network capabilities mainly include location services, bandwidth management services, wireless network information services and other information, which are the unique capabilities only provided by operators. These capabilities can be introduced to PaaS platform via Restful interfaces. However, most of the edge applications are yet to make use of these capabilities. On one hand, the developer is not familiar with these capabilities. On the other hand, there is a huge gap between what can be provided by operator and what is actually needed by vertical applications.

4) Edge platform SDKs

As the related edge capabilities are introduced to PaaS platform, it directly provides the native interface to the applications, which is not friendly to for developers to use. SDK encapsulation for native interfaces can help developers to develop edge application more effectively. Providing SDK and message middleware on PaaS platform can further guarantee the stability of open capability. By optimizing SDK, the attachment of application to edge data center platform can be improved. In addition, SDKs such as OpenViNo, CUDA and so on, which are widely used in the industry, can reduce the difficulty of porting related applications to PaaS platform by ensuring optimized compatibility with the platform.

5) PaaS capabilities of third-party platforms

In principle, PaaS platform shall support integrating PaaS capabilities from third-party platforms. Enterprise applications generally rely on their own private or public clouds. When they are transplanted to edge, PaaS platform needs to support relevant capabilities to ensure the normal operation of these applications. For example, enterprise application A was originally deployed in some public cloud and used capability m. When the application is required to be deployed on the edge, application A and capability m are deployed on the PaaS platform together through the unified application deployment process, which requires compatibility for both application A and capability m on the platform.

The introduction of third-party PaaS capabilities are very complex, choosing cloud native software system can reduce the related software development workload.

6) Service operations and maintenance

According to the characteristics of the platform, the edge computing PaaS provides the routine monitoring tools, debugging tools and many to enrich the operation and maintenance capabilities of the application. The platform itself provides multiple capabilities for management and operation, such as an optimized micro-service framework. Introducing new technologies such as Serverless, Service Mesh and

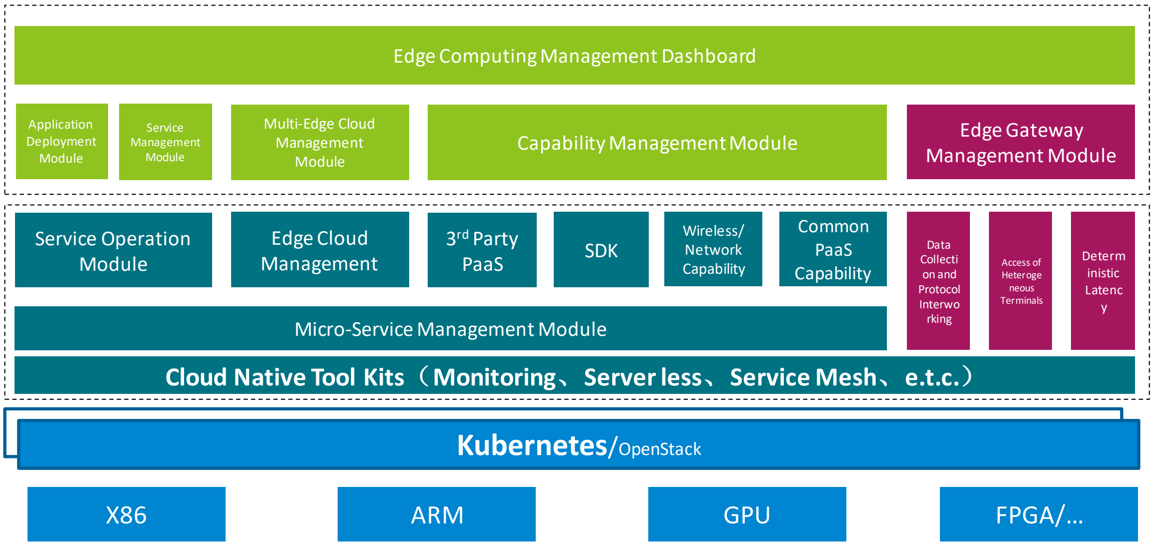

Edge Computing Management Dashboard
Application Deployment Module
Service Management Module
Multi-Edge Cloud Management Module
Capability Management Module
Edge Gateway Management Module
Service Operation Module
Edge Cloud Management
3rd Party PaaS
SDK
Wireless/ Network Capability
Common PaaS Capability
Data Collection and Protocol Interworking
Access of Heterogeneous Terminals
Deterministic Latency
Micro-Service Management Module
Cloud Native Tool Kits（Monitoring、Server less、Service Mesh、e.t.c.）
Kubernetes/OpenStack
X86
ARM
GPU
FPGA/...

Figure 5 Block diagram of China Mobile edge computing PaaS

Microservice and using Cloud Native to develop and maintain applications will greatly enhance the operational efficiency and reduce application failures.

7) Multi-edge cloud management

Due to the large number of edge data centers, multi-edge cloud management is a difficult problem in edge computing. In order to improve resource utilization and enhance user experience, different edge computing nodes may be associated with each other, which further increases the complexity of management. At present, there is no mature open source solution for this problem and it is remained a major challenge for the near future.

3.2. Edge Computing Capability Exposure

3.2.1. Achitecture

As shown in Figure 3-1, the Edge Computing Capability Exposure architecture consists of the edge capability exposure layer, the edge encapsulation & invocation layer and the edge capability-access layer. The edge capability exposure layer provides CT capability, IT capability and specific service capability for applications and developers. The edge capability exposure layer also allows the developer to orchestrate different capabilities online to meet their particular needs. At the same time, it manages the open APIs with the network capabilities, access to application and partners to ensure the stability, efficiency and security of the open API invocation of the edge computing capability. The edge capability encapsulation &invocation layer implements the invocation and the atomic capability encapsulation of the communication network capability, the service capability and the network infrastructure resource capability. To shield the differentiation of different types of networks, the edge capability access layer implements the uniform access to the 5G network, the fixed network and other types of networks.

3.2.2. Capabilities Exposure

The edge computing PaaS platform is an open platform includes three aspects： the openness of edge network communication capabilities, platform management and platform services. Building an open edge computing platform and incubating innovative business with partners is the important means by which operators empower business.

The exposure of edge network communication capabilities is an important means for operators to make money out of communication networks capabilities, including user location, wireless information, and QoS service, etc. The edge

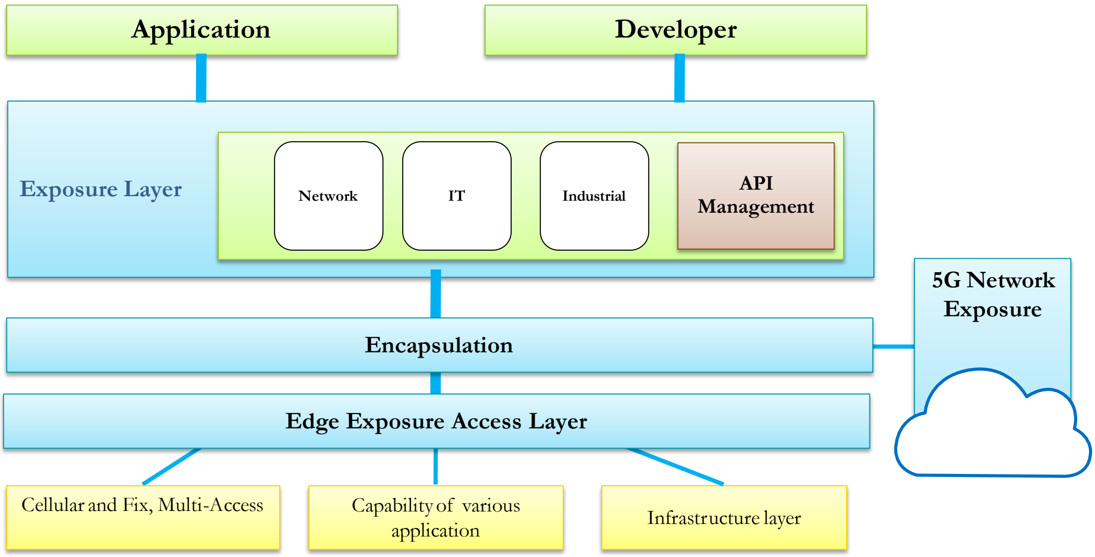


Figure 6 System architecture of edge capability exposure

computing platform can directly provide authorized user location services, and can also provide personalized interactive services in combination with real-time content such as identity information, business information and behavior habit information of the user in the mobile network. The deployment of the edge computing platform at the edge network provides convenient conditions for real-time perceiving of wireless network information, such as real-time wireless network conditions and UE QoS information. To optimize the services of the third-party, enhance user experience and achieve deep integration of networks and services, the network information is provided through an open interface on the edge computing platform to third-party service applications in the form of APIs. Based on the network QoS control capabilities provided by the edge computing platform, the third-party applications can obtain differentiated network services according to their business needs, and then improve user service satisfaction.

The exposure of management capability for the edge computing platform is an important way to enrich the business ecology of the edge computing platform. The third party applies for computing resources of platform on demand, so that the applications of the third-party have a good operating environment. At the same time, it provides APP lifecycle management capabilities, configuration capabilities and monitoring capabilities to the third parties, enabling the third party to have flexible local self-owned APP's operation capability and build an open edge computing business operation ecosystem.

Furthermore, to achieve agility of the third-party application development, accelerate the release of industry applications and realize accurate service provided by the edge network, the edge computing platform can provide customized services to the applications of the third party on demand, such as video codec capability, AI algorithm library capability, which can be introduced from other service providers.

3.2.3. Interface Protocol

This edge computing capability exposure interface is implemented by HTTP protocol that supports multiplexing, traffic priority setting and header compression. HTTP 2.0 is highly recommended. To realize fast connections and efficient concurrent invocation, the exposure interface uses the standard, flexible and convenient RESTful APIs, using Json as the data interaction format.

3.2.4. Encapsulation and Orchestration

The MEC Capability Exposure provides a rich variety of atomic capability APIs for applications. The atomic capabilities APIs can be encapsulated and orchestrated in real time to implement the combined APIs. Further, the parameters of the combined APIs can be masked or mapped to implement the personality encapsulation of the APIs. The MEC Capability Exposure supports to set the invocation priority, time sequence and invocation logic for multiple API invocations. The conflict and recursive detection mechanism guarantee the correct API invocation.

3.2.5. Capability Management

The edge computing capability exposure includes CT capabilities, IT capabilities, specific service capability and the third-party capabilities. It provides the standard registration, logout, activation, deactivation, release, subscription update, and notification update mechanism to unified management and operations for the open API.

The edge computing capability exposure monitors the status, performance, and concurrent data of the capability API in real time. The performance data such as the invocation success rate and latency of the capability API are also collected and statistically analyzed for a specified period. The behavior of each application, including but not limited to the number and frequency of the called capability API, will be monitored and continuously recorded.

3.3. Introduction of Application and Capabilities on Edge Network

In order to expand its application ecology, the edge computing platform actively introduces the public cloud PaaS capability in the market. If the application is bound to a certain capability of a public cloud, the application and the capability can be deployed on the platform simultaneously by refer to the list of edge computing capabilities in the centralized management platform. Based on the main vertical areas China Mobile recently focuses on edge computing, video procession and V2X are the verticals with most clear requirements.

1）VIDEO

With the rapid development of Internet services, mature CDNs are already standard on edge computing. The emergence of 8K, 12K and other video requires not only the high bandwidth capability of 5G and the content cache of CDN, but also the edge calculation to encode and decode the data stream. Edge computing introduces video processing capabilities to provide a better user experience for new services such as video ring tones and 12K.

2）V2X

V2X applications have requirements for network latency, bandwidth, business continuity, and location information. For the low-latency and high-bandwidth requirements, the local breakout device and the edge computing platform can be deployed in the edge network.. For the service and session continuity requirements in high-speed mobile scenarios, the edge computing platform that carries the vehicle networking application needs to cooperate with the basic network, and the network-side session management mechanism is used to ensure the continuous service experience in the fast moving state of the vehicle.

4. Edge Computing IaaS Technology

4.1. Design Concept

Edge computing services need to be deployed at the edge of the network close to users and end devices. Based on the difference of business model, resource condition, business needs, operation and maintenance requirements, deployment form of edge computing can be integration of hardware and software, or cloud. Cloud can be more flexible in deployment, operation and maintenance, and billing. Edge service providers can use resources on demand with cloud-based edge computing infrastructure, and avoid heavy assets and maintenance. Therefore, edge service providers can get lower marginal cost of developing business in the same marginal region. Edge Computing IaaS is the cloud-based edge computing infrastructure, in which edge computing services and related network functions of cloud form can be deployed. Edge Computing IaaS is the combination of cloud computing technology and edge computing scenarios.

The types of services or applications that need to be deployed for edge computing mainly include MEC applications and MEC PaaS platform, also including network functions like UPF and CU, etc. Edge computing IaaS should be able to provide a cloud infrastructure for these services and applications to meet the needs of different services and applications.

The following factors should be considered when designing the edge computing IaaS:

(1) Edge computing applications focus on cloud native design, quick start and stop, quick update. Telecom network functions focus on performance, reliability, manageability. There are differences in requirements for Edge computing IaaS between telecom network functions and edge computing applications.

(2)There may be huge number of distributed edge cloud nodes. Unattended remote operation and maintenance should be considered in edge computing IaaS.

(3) The pattern of edge computing services may be introduced point by point by location. Therefore, edge computing IaaS in each edge node should be able to serve the deployed services without relying on the IaaS of other edge nodes.

(4) The use of cloud resources by edge computing applications should enable on-demand usage and billing.

(5) From the perspective of operation and maintenance, it should be able to have a unified view and allocation authorization management for edge computing IaaS resources.

(6) The resource conditions of edge nodes such as space and power distribution are limited. Edge computing IaaS needs to consider optimizing resource allocation so that the business gets more available resources.

(7) Edge computing should have the ability to combine with 5G slicing, etc. Starting from these factors, the following figure shows the architecture design of the edge computing IaaS.

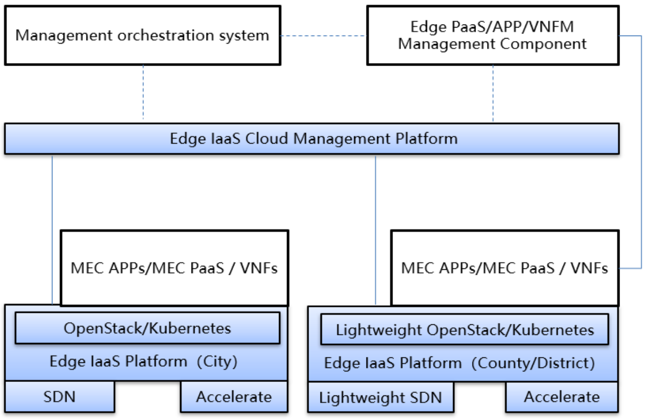

Management orchestration system
Edge PaaS/APP/VNFM Management Component
Edge IaaS Cloud Management Platform
MEC APPs/MEC PaaS / VNFs
OpenStack/Kubernetes
Edge IaaS Platform (City)
SDN
Accelerate
MEC APPs/MEC PaaS / VNFs
Lightweight OpenStack/Kubernetes
Edge IaaS Platform (County/District)
Lightweight SDN
Accelerate

Figure 7 The concept for Edge Computing Design

The edge computing IaaS architecture design fully considers the above considerations and embodies the following core concepts (corresponding to figures in the figure):

(1) Unified operation and maintenance

Introduce edge computing IaaS cloud management platform as a unified entrance for operation and maintenance of IaaS for all edge nodes in the jurisdiction. Realize remote operation and maintenance of unattended edge nodes. And converge the northbound interface to the management and orchestration system such as Management orchestration module to save network overhead.

(2) Autonomy

The edge nodes exist in multiple locations such as cities, districts, and accesses, but edge computing IaaS of each location is an autonomous cloud, and edge computing IaaS is not dependent on other edges.

(3) Heterogeneous cloud platform

Edge computing applications and related telecom network functions may adopt different designs. From the perspective of the cloud, it is needed to support both virtual machine and container resources. Therefore, edge computing IaaS needs to support OpenStack cloud and Kubernetes cloud. The relationship between the two is discussed later.

(4) Lightweight IaaS

The resources of the edge nodes in network edge locations are limited. Management overhead of edge computing IaaS can be lightened by using the fusion nodes and the compression of IaaS management component resources to maximize the available resources of the service.

(5) On-demand usage and billing

The edge computing PaaS platform or service can apply for edge computing IaaS resources on demand, and the edge computing IaaS is charged according to usage.

(6) Unified cloud resource view

The management and orchestration system (such as Management orchestration module) should have a unified view of the edge computing IaaS resources to manage the resources consistently and authorize the edge computing PaaS platform or service to request IaaS resources.

(7) Lightweight network

Flat networking is needed in the edge nodes such as counties and access locations.

(8) Support SDN

Edge computing IaaS needs to support SDN, as well as SDN-based slicing capabilities.

(9) Support acceleration

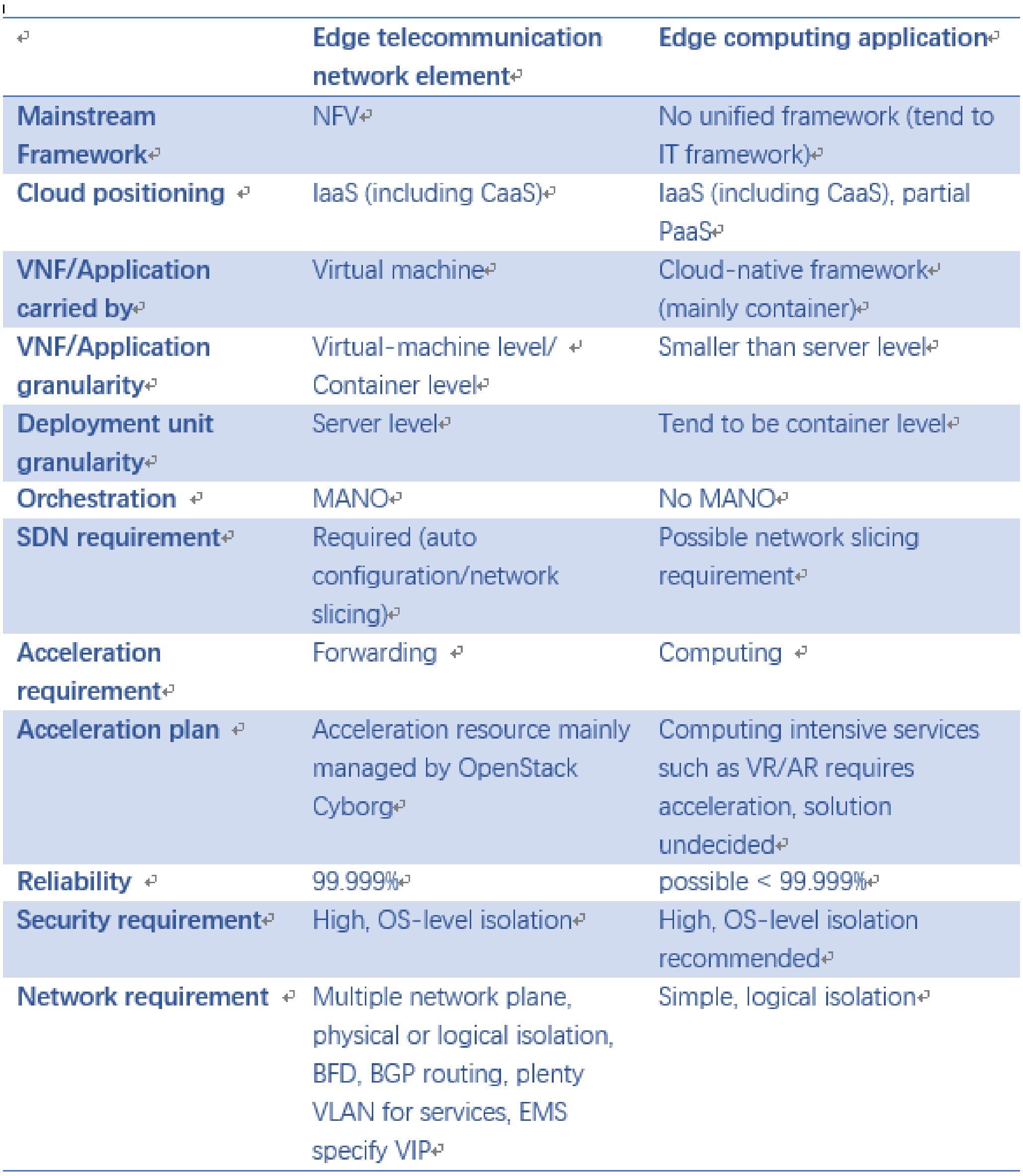
User plane network functions such as UPF and edge computing applications with higher computational density are more stressful on the CPU. Edge computing IaaS is needed to support the offloading of the acceleration function to the hardware implementation.

4.2. Various Forms of Edge Computing IaaS

There are two existing forms of edge computing IaaS: virtual machine and container, the corresponding cloud management system includes OpenStack and Kubernetes. As the functions of current telecommunication network element are complex, NFV is the main cloud technology in telecommunication industry, that is, virtual machine and OpenStack. The possibility of using container to carry telecommunication network elements after using cloud native design is being explored. The applications of edge computing, which is similar to internet and IT applications, prefer to use cloud native design concepts such as micro-services, and are more suitable to be carried by container. However, considering the potential security hazards of shared kernels in containers where different edge computing applications reside, there is also the need to adopt pure virtual machines or container over virtual machines. The following table shows the differences in demand for edge computing IaaS between telecommunication network elements and edge computing applications.

There are many types for edge computing IaaS result from the different requirement for IaaS of VNF and vertical applications.

Table 1 Comparison of different requirement of NFV and Edge Computing



	Edge telecommunication network element	Edge computing application
Mainstream Framework	NFV	No unified framework (tend to IT framework)
Cloud positioning	IaaS (including CaaS)	IaaS (including CaaS), partial PaaS
VNF/Application carried by	Virtual machine	Cloud-native framework (mainly container)
VNF/Application granularity	Virtual-machine level/ Container level	Smaller than server level
Deployment unit granularity	Server level	Tend to be container level
Orchestration	MANO	No MANO
SDN requirement	Required (auto configuration/network slicing)	Possible network slicing requirement
Acceleration requirement	Forwarding	Computing
Acceleration plan	Acceleration resource mainly managed by OpenStack Cyborg	Computing intensive services such as VR/AR requires acceleration, solution undecided
Reliability	99.999%	possible < 99.999%
Security requirement	High, OS-level isolation	High, OS-level isolation recommended
Network requirement	Multiple network plane, physical or logical isolation, BFD, BGP routing, plenty VLAN for services, EMS specify VIP	Simple, logical isolation

1） Container cloud based on bare-metal (for Edge Application)

The edge network element is not carried by IaaS for edge computing, but by VNF integrated machine (internal support OpenStack, support northbound interface), while edge computing services are carried by bare machine containers (or virtual machines managed by Kubernetes such as Katacontainer, Kubevirt, etc.). Edge Computing PaaS platform or service needs to request the authorization of edge computing IaaS resources from NFVO (or other management and orchestration module). NFVO needs to have global view and comprehensive information of resources used by edge computing applications. Edge computing IaaS needs to be able to charge for edge applications based on their usage conditions. In this situation, bare metal management should be considered.

2） Unified virtual-machine cloud(managed by OpenStack)

Both edge computing services and edge telecommunication network elements are carried by virtual machines (or containers inside virtual machine, where containers are invisible). OpenStack divides resources according to different tenants for edge computing services and edge telecommunication network elements, and isolates them from each other. Edge Computing

PaaS platform or service needs to request the authorization of edge computing IaaS resources from NFVO. NFVO needs to have global view and comprehensive information of resources used by edge computing applications. Edge computing IaaS needs to be able to charge for edge applications based on their usage conditions. There are no modifications for NFV MANO in this form.

3）Unified container cloud based on bare-metal (managed by Kubernetes)

Both edge computing services and edge telecommunication network elements are carried by bare-metal containers (or virtual machines managed by Kubernetes such as Katacontainer, Kubevirt, etc.). Kubernetes divides resources according to different tenants for edge computing services and edge telecommunication network elements, and isolates them from each other. NFVO needs to have global view and comprehensive information of resources used by edge computing applications. Edge computing IaaS needs to be able to charge for edge applications based on their usage conditions. In this situation, NFV MANO needs to be reconstructed to support containers, and bare metal management needs to be considered.

4）Mixed cloud (one cloud with two domains, managed by both OpenStack and Kubernetes)

The edge computing service is carried by bare-metal container and the edge network element is carried by virtual machine. Telecommunications applications use forwarding hardware, while edge computing applications use computing hardware. NFVO has global information, and interface with both the container domain and virtual machine domain. Edge Computing PaaS platform or service needs to request the authorization of edge computing IaaS resources from NFVO. NFVO needs to have global view and comprehensive information of resources used by edge computing applications. Edge computing IaaS needs to be able to charge for edge applications based on their usage conditions. In this situation, NFV MANO needs to be reconstructed to support containers, and bare metal management needs to be considered.

5）Container cloud embedded in virtual-machine cloud (virtual-machine cloud and cloud of container inside virtual machine)

The edge element is carried by virtual machine, and the edge computing service is carried by virtual-machine container (virtual machine is provided by virtual-machine cloud managed by OpenStack). OpenStack divides resources according to different tenants for Kubernetes cloud and edge telecommunications network elements and isolates them from each other. NFVO has global information, and interface with both the container domain and virtual machine domain. Virtual machines as container resources will require additional management processes. Edge Computing PaaS platform or service needs to request the authorization of edge computing IaaS resources from NFVO. NFVO needs to have global view and comprehensive information of resources used by edge computing applications. Edge computing IaaS needs to be able to charge for edge applications based on their usage conditions. In this situation, NFV MANO needs to be reconstructed to support containers.

4.3. Key Techologies for Edge Computing IaaS

4.3.1. Edge Computing IaaS Platform

The edge computing platform is a resource collection infrastructure for deploying and running edge VNFs and IT APPs. It provides a platform for providing common technical component capability such as VNF holds physical and virtual resources. Virtual machine is an important way for edge computing IaaS platform. With the development of container technology and the analysis of real-world scenarios of edge computing services, virtual machines, containers, and all-in-ones may be included in the future, and many forms will coexist for a long time. In addition, edge computing IaaS can also implement lifecycle management functions for edge data center IT equipment (computing, storage, network), life cycle including automatic discovery, automatic management, automatic configuration, automatic monitoring, automatic management, Automatic repair, automatic testing, automatic online and includes the configuration, operation, and management of hardware resources and software resources. In the interface and adaptation of the edge computing IaaS layer, it needs to interface with x86 server, various storage, network equipment, virtualization software, SDN, properly adapt to distributed storage.

The virtual machine is an important existence mode of the edge computing IaaS layer. It is very mature in the NFV field. It manages various types of hypervisors and virtual machines hosted by the hypervisor through OpenStack technology, and needs to ensure that the service is stable running on the bearer platform. Hypervisor is an intermediate layer of software that runs between a physical server and an operating system, allowing multiple operating systems and applications to share a single set of physical hardware. At present, the mainstream hypervisors in the x86 architecture field include KVM, Xen, ESXi, etc. Edge computing IaaS also considers a variety of hypervisors to carry VNF and other services, and promote and enrich the industry development. As an important part of management, OpenStack is the de facto industry standard for cloud computing IaaS. Relevant functions and interfaces have been widely recognized by the community and vendors, and can basically meet various functions, performance, reliability and operation and maintenance management, northbound interface and many other requirements. At the same time, if the NFVO, VNFM, OSS and other functional modules of the NFV architecture are used, the advantages of resource management, scheduling, expansion and contraction can be brought into play. These features can continue to play a role in the edge computing IaaS.

With the development of network elements and edge applications, more forms of bearer have evolved, especially the rapid development and gradual maturity of micro-services and new IT technologies. Container technology is also increasingly valued by edge computing developers. Edge nodes will likely be introduced step by step into containerized network elements and container management platforms. Business systems are built flexibly based on the Platform as a Service architecture. Plan, analyze, and resolve problems between business processes and the IT process's own systems through a global, systemic perspective. Containers are used to package various applications and operating environments. Provide a unified development, testing, and production environment for upper-layer applications. The sub-tenant performs centralized security management, image management and distribution, automatic service online, application unified configuration, data backup, and unified management of related basic services (database, message, log) to meet the flexible use and rapid

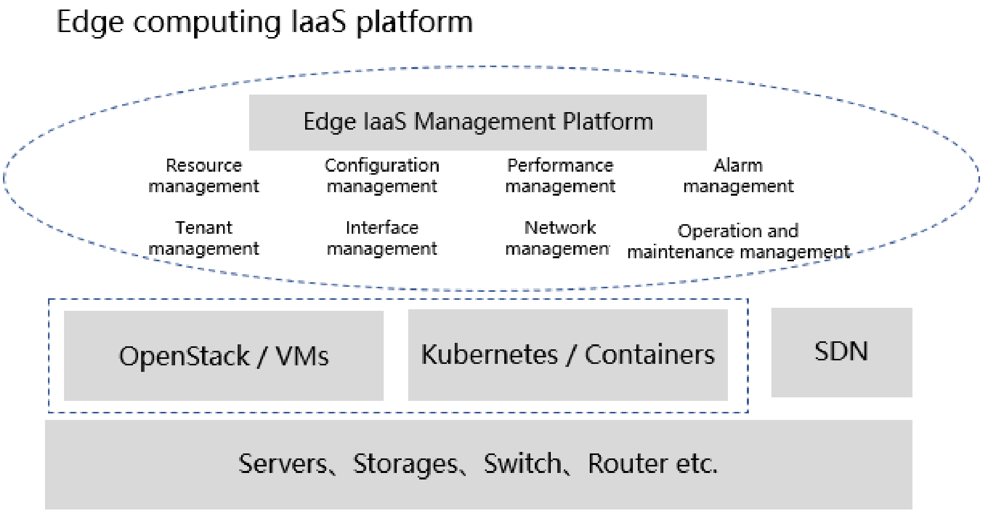


Figure 8 Schematic function blocks of for China Mobile Edge Computing IaaS Platform

iteration capability of enterprise applications. demand. Choosing container virtualization software and Kubernetes management platform to build the core of microservice management and containerized operation, IaaS can be calculated by a wider range of services, and can form standardized, flexible and open core capabilities and platform support.

In order to solve the realistic needs of the location (district and access) scene closest to the user side, the edge computing IaaS platform requires lighter weight, This part of the node usually does not meet the hardware conditions of the centralized computer room, such as space, power, and refrigeration, at the same time, the size of this single resource pool is relatively small and scattered. According to the estimation of the edge calculation IaaS layer resource pool, the resource pool of the edge node of the district and county is about dozens of servers(The total occupied space is about 40U-150U).For edge nodes that can be extended to the access side, the scale can be as low as a few or a dozen servers(The total occupied space is about 2U-40U).Considering the small and scattered characteristics of the edge computing IaaS resource pool, it is especially important to improve the comprehensive utilization rate and management lightweight. In the virtual machine scenario, the management units of the edge nodes are mainly OpenStack and SDN, in the future, we will focus on lightweight deployment of key components of OpenStack and lightweight deployment of SDN controllers. Continuously promote management component virtualization and containerized deployment, reducing management component resources from physical server level to CPU core level. In the container scenario, based on the lightweight features of the container itself, the management unit of the edge node is mainly Kubernetes, and the subsequent focus is on the deployment of Kubernetes and the container in the extreme edge scenario. Through the unified management at the upper level, the service container resource pool is faster and simpler, and the utilization efficiency of the container resource pool is improved.

4.3.2. Edge Computing IaaS Network

Within the edge nodes, the way of networking varies greatly with the resource of the data center and the scale of the servers. For edge nodes with good conditions, such as city level nodes, because of the better conditions in the data center and the larger scale of servers (100 level), the Spine-Leaf switching architecture is adopted in the internal networking, which realizes the physical isolation of service, storage and management traffic at the server port and switch. For smaller edge nodes, such as those in districts and counties or below, the expansion of room space and power supply is limited, and the scale of servers is about 10 or less. Single-layer switching architecture can be used for logical isolation. In this way, the traffic of service, storage and management can be isolated by setting. Currently, there are some customization of switches, which need to be promoted by the industry.

The requirement of network configuration automation in edge nodes, as well as the requirement of rapid service on-line and 5G slicing, requires that deploy SDN in edge nodes on demand gradually. For edge nodes with better conditions in data center (such as city level), SDN deployment solution can refer to the network scheme of large-scale core data center. For edge nodes with poor conditions (such as districts, counties and below), limited by space and resources, SDN software should reduce its occupation of server resources as much as possible. It is necessary for the industry to jointly promote the SDN controller resource occupation to be more lightweight, remote SDN controller deployment, SDN switch/GW miniaturization and other schemes. In addition, virtual machine resource pool is an important IaaS platform at present, and edge computing IaaS platform network scheme takes virtual machine network scheme as an important network solution. Considering the introduction of containers in edge nodes, the network scheme will be changed. On the basis of retaining the three-plane isolation in the original hardware networking, it is necessary to consider the compatibility of virtual machines, containers and SDN.

4.3.3. O&M of Edge Computing IaaS

Edge Computing IaaS will be different under different scenarios, and will be decentralized deployed. If considering that all edge nodes adopt independent operation and maintenance management, the management is difficult and O&M costs are high. So, in county/access edge nodes, which are closer to users, only cloud resource pools and network devices should be deployed, while a proper amount of hardware maintainer should be equipped. Centralized O&M ability should be issued to capital/city edge nodes, and provides O&M for themselves and lower-level county/access edge nodes. Since virtual machine is still the main implementation of edge computing IaaS in the early stage, OpenStack is still the major management component of virtual resources. There are many methods of implementing centralized O&M using OpenStack from Open Source communities and industry, including remote Hypervisor, Multi-Region, Cell, and cloud management platform of edge computing IaaS. The first three methods provide unified resource management access portal and centralized identity authentication through centralized deployment of interactive interface and Keystone components in capital/city edge node. Remote Hypervisor method provides only hardware resources and virtual resources in local edge node, such as compute, network and storage devices. Multi-Region and cell provide partial management functions, and hardware resources and virtual resources in local edge node. The forth method, cloud management platform of edge computing IaaS, adds a new cloud management platform at capital/city edge node. This platform provides unified management access portal and interface, API distribution, remote operation links and other functions northward, while managing multiple independent city/county/access edge node southward. In forth method, lightweight OpenStack using minimal resources or fully deployed OpenStack will be chose based on different resource capacity of different edge nodes. Considering the poor conditions of existing networks, and in order to maintain the local management ability and failure recovery ability, central cloud management platform of edge computing IaaS with multiple independent edge cloud resource pools is recommended.

At present, there is a strong demand for fast iteration and dynamic creation of related applications in edge nodes. Containers and Kubernetes have also become one of the important ways to implement IaaS layer in edge computing. Therefore, it is suggested to extend the function of cloud management platform of edge computing IaaS to enable it to manage multiple virtual machine resource pools and container resource pools on edge.

4.3.4. Edge Computation IaaS Layer Acceleration

The edge computing IaaS layer has higher requirements on real-time and network capabilities. Under the hosting of virtual machine scenarios, virtual machines need to be properly interconnected with each other. A universal hard and soft channel that solves the data path between VNF software and accelerated hardware. At present, the more mature solution is to optimize the vSwitch and SR-IOV hard pass-through technology for DPDK.

VNF and edge PaaS/APPs need to communicate with other network devices through the vSwitch. The vSwitch is implemented in software. In order to meet the calculation and forwarding, the CPU has high power consumption overhead. In the hardware acceleration scheme, the atomized functional unit is offloaded to the hardware acceleration card. At present, there are many options for the popular acceleration chip on the market, and the acceleration chip embedded in the network card to form an intelligent network card is the main form of the current acceleration card. In edge computing, the computational acceleration and storage acceleration requirements of new edge services such as AI are more prominent. In the initial stage, the FPGA solution can be flexibly modified and adjusted, and the hardware acceleration and cost performance will be gradually improved in the future.

The acceleration of edge computing IaaS requires management orchestration. Through NFVO and

OpenStack, you can see and select the appropriate acceleration resources for VNF or edge services. OpenStack Cyborg currently offers some features, and it is up to the industry to discuss whether it meets commercial conditions. A universal acceleration API that addresses VNF and accelerated hardware decoupling issues, ensuring that the hardware-providing capabilities are fully abstracted.

In the edge computing IaaS, it is necessary to gradually accelerate the scheduling management function and interface requirements, and promote the development of the community Cyborg project. According to the industry's mainstream accelerated network card product roadmap and business needs, the Acceleration Unloading Solution of Checksum, IPSec, GTP, and GPU will be gradually implemented in the future, conducting tests in the laboratory.

Acceleration technology is an important and difficult area for the edge computing IaaS layer, which needs to be followed up to solve the issue. For example, to promote the accelerated abstraction layer in the DPDK open source maturity, promote VirtIO open source support for the accelerated path. In addition, Kubernetes' acceleration issue is also the focus of future research. In the Kubernetes scenario, how to optimize the use of edge data center GPUs, the process of accelerating machine learning in this environment requires further research.

5. Edge Computing Hardware System

5.1. OTII Server for 5G and Edge computing

In November 2017, China Mobile, in conjunction with China Telecom, China Unicom, China Telecom, Intel and other companies, launched the Open Telecom IT Infrastructure (OTII) for telecommunication applications in ODCC (Open Data Center Committee), the primary goal of which is to form an open and unified server solutions and products that suitable for 5G and edge computing.

OTII project has attracted wide attention. So far, it has received 29 mainstream suppliers' support in the fields of traditional telecommunication equipment, servers, firmware and management systems.

5.1.1. Edge Computing Requirements and Challenges for Servers

1）Edge Data Center Environment

Compared with the core data center, the infrastructure conditions of edge and access central office are quite different. Many aspects cannot meet the deployment and operation requirements of general servers, which brings challenges to edge servers.

Space limitation of rack. Most of the rack deployed in transmission and access central office currently is about 600 mm depth, and very small amount of rack can reach 800 mm depth.

Temperature stability. Because the stability of the refrigeration system in the edge and access central office can not be guaranteed, when the refrigeration system fails, the room temperature may reach more than 45 degrees Celsius.

Load-bearing limitatio. Many edge and access central office are generally below the load-bearing standard of data center.

Other restrictions. Servers deployed in edge and access central office will also face many limitations, such as high requirements for seismic, electromagnetic compatibility noise prevention, and poor air quality.

2）Service Requirements for Server Performance

Different types of edge computing require different performance of servers. Servers need to support certain computing and storage capabilities. In addition, edge computing also has a large number of heterogeneous computing requirements. It is necessary to offload some CPU functions by configuring network cards based on FPGA, ARM or other hardware acceleration schemes to save CPU cores and improve processing efficiency.

3）Operations and Maintenance Management Requirements

OTII edge servers for edge computing services will be distributed in a large number of edge and access central office, so management and maintenance capabilities are needed.

(1) Unified management interface. The server needs a unified management interface to reduce the large amount of adaptation work brought by out-of-band management system, which is to manage the server more effectively.

(2) Efficient operation and maintenance. Edge servers should minimize the requirement of operation and maintenance personnel, make operation and maintenance operation as simple as possible, and improve operation and maintenance efficiency.

(3) Fault diagnosis and self-healing. Server BMC has basic fault diagnosis and reporting capabilities, and provides hardware platform self-healing solutions.

5.1.2. OTII Technical Scheme

In response to the above needs and challenges, OTII project combines the operator Edge and access central office environment and edge computing needs, and carries out a series of research and analysis and program design with industry partners.

1）Configuration

In the aspect of motherboard design, NUMA Balance design will be adopted for the configuration of two CPUs to meet the performance and stability of multi-PCIe device application scenarios. In terms of scalability, it can meet the configuration requirements of most edge scenarios, including storage, PCIe slot expansion and so on. JBOD/JBOF is also considered for edge storage scenarios.

2）Physical morphology and environmental adaptability

In order to meet the environmental requirements of edge computing, the server makes a targeted design scheme:

Server depth is recommended to be 450 mm, up to 470 mm.

Front access of switches, indicator lights, hard disks, cables, etc.

Fans can support hot plugging to ensure online cleaning or replacement.

For some edge application scenarios, it may be necessary to support running in a wider temperature range (e.g. - 5 degrees to 45 degrees), and it may be necessary to meet the requirements of B-level EMC, earthquake resistance, etc.

3）BIOS, BMC and Hardware Management

OTII project cooperates with server, BMC and FW vendors to develop unified server hardware monitoring and remote management functions, so that the upper management platform can connect with different vendors and servers with different configuration specifications without any difference.

5.1.3. Current progress

Following the first prototype reference design of OTII customization server published in MWC Shanghai 2018, Three Partners have completed the product development of OTII edge server based on Intel® Xeon® Scalable Processors (Cascade Lake)，Which release at MWC 2019 .

In addition, several OTII edge samples have been used in field trials of edge computing. In 2019, we will continue to promote product development and scale out the field trial.

5.2. All-in-One Solution

5.2.1. Application Scenario

In some scenarios, it is more difficult to deploy traditional multiple servers and access switches, such as:

- Some temporary emergency edge computing scenarios require good mobility, rapid deployment, and plug-and-play scenarios;
- Some telecom equipment rooms with few free cabinets or only scattered half-empty cabinets have insufficient space for deploying multiple servers or can be deployed in separate cabinets, which increases the difficulty of deployment;
- In some edge scenarios with small traffic, the server deployment may only need 2 to 3 servers. The deployment of traditional servers and access switches requires equipment to be racked, constructed, and debugged. The deployment time is long and the construction process is complicated. The scenario requires a simplified deployment;
- For some equipment rooms without standard cabinets or without empty cabinets, traditional servers and access switches may not be deployed.

In response to the above scenario, China Mobile launched an Edge Computing All-in-One device deployment solution that supports integrated delivery of hardware (including servers and access switches) and software (including OS and app) for plug-and-play and rapid deployment, and is small in size with high density of integration. This type of server Can be deployed in a standard rack independently.

5.2.2. Technical solutions

In traditional IT data center deployment methods, the server and the access switch are connected through optical fibers. The server virtualization service is divided into three planes: service management, storage, and service traffic. Each plane requires two interfaces for redundancy backup, so each server has at least 6 Interfaces. Suppose there are 8 servers, the number of links between the server and the access switch is 6x8=48, which requires 48 pairs of optical fibers and 96 optical modules. For the edge machine room, this deployment scheme has a large workload, high cost, and many fiber links. There are many fault points. Especially some unattended sites, the faults are big, and the fault location and repair takes a long time, which is not conducive to the fast launch of the business and stable operation.

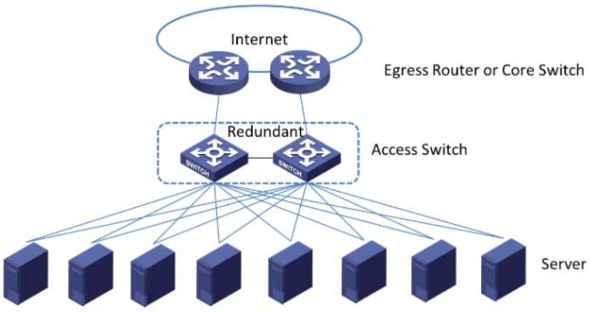


Figure 9 Schematic illustration for IT networking

The All-in-One device integrates the server and the access switch into one cabinet, and the server and the switch are installed in a flexible card form. The device structure is as follows.

The technical solution has the following characteristics:

(1) The server and the access switch are integrated into one chassis. The server and access switch are simplified to the chassis card and can be flexibly inserted and removed.

(2) Ful-lmesh connection between the server card and the access switch card to support link redundancy.

(3) The access switch card supports active/standby backup. The link bandwidth and number of interfaces can be flexibly matched according to service requirements and egress router interface configurations.

(4) The hardware configuration (storage type or computing type) of the server card is flexibly matched according to the service scenario;

(5) The chassis size is adapted to the telecom equipment room requirements, for example, the maximum depth is less than 600 mm.

(6) The power supply and fan support pluggable and support N+1 backup.

(7) All cable interfaces are placed in front for easy installation and operation and maintenance;

(8) Software supports pre-integration and pre-integrates different system software according to business scenarios.

The edge computing All-in-One solution has the following advantages over traditional split server deployments:

(1) Simplify deployment, plug and play, shorten deployment time, and improve service efficiency.

(2) It takes up less space and can be used in small telecom equipment rooms;

(3) Easy to deploy and disassemble, and change the deployment location with business migration;

(4) The connection between the server and the access switch is replaced by hardware PCB traces, eliminating optical modules and fiber fault points, and saving the cost of optical fibers and optical modules.

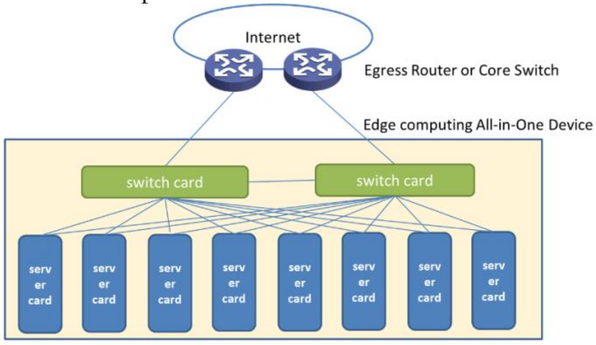


Figure 10 Block diagram of China Moble's All-in-one delivery service

Scalability is better. The server card can replace the number of server cards and hardware configuration according to service requirements. The switch card can also easily upgrade the egress bandwidth.

5.3. Edge Computing Gateway

The Edge Computing Gateway represents On-premise edge computing technology system for vertical industry's Internet transformation. It is committed to extending intelligent network access capabilities to vertical industry sites, relying on quality-guaranteed network connectivity, computing and storage resources to support flexible deployment and operation of multi-ecological services in user site.

The edge computing gateway will cooperate with the edge server and all in one edge devices to integrate the agile and flexible of IT and the reliable and stable of OT. It can apply the network connection, quality assurance, maintenance management and scheduling capabilities of the operator attributes to Vertical industry, providing real-time, reliable, intelligent and ubiquitous end-to-end services.

5.3.1. Location of deployment

The edge computing gateway focuses on the deployment of each vertical field, and belongs to the user terminal network equipment in the operator's network system. For the vertical industry, government, enterprise, home and individual access scenarios, the edge computing gateway can be accessed through the cellular network or through the fixed network.

In terms of management, edge computing gateway and edge data center are also managed by the edge PaaS management platform. There may also be management and service collaboration between the edge computing gateway and the edge data center.

5.3.2. Key Technology

The edge computing gateway includes heterogeneous LAN side interface, flexible customized system resource layer, a lightweight virtualization framework and a WAN side interface layer.

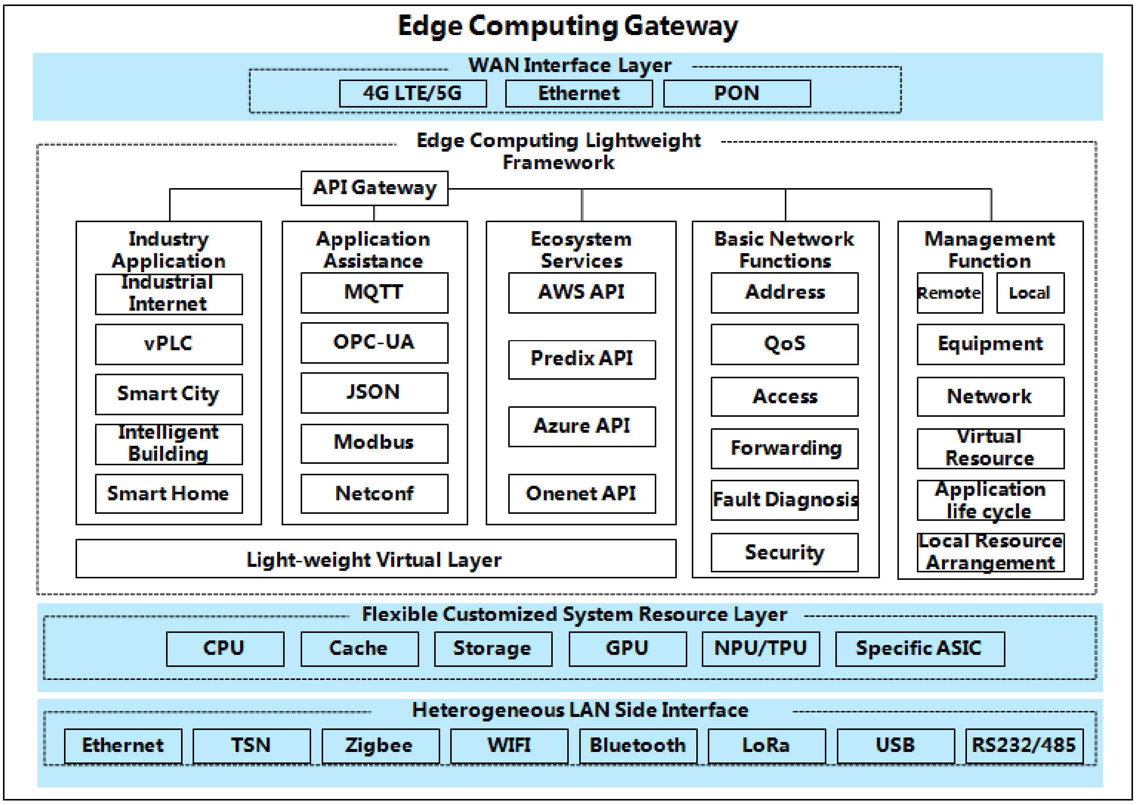


Figure 11 Function block for Edge Computing Intelligence Gateways

1）Architecture

- Heterogeneous LAN Side Interface

The edge computing gateway has a heterogeneous LAN side interface and supports a variety of common protocols, such as Ethernet, TSN, Zigbee, Lora, RS232/485, and so on. Data collection, protocol conversion, etc. are the most basic functions required in the industry, laying the foundation for supporting the industry application.

- Flexible Customized System Resource Layer

The flexible customized system resource layer supports multiple types of processors, caches, and storage resources. Different industries and applications have different requirements for gateway. For example, TSN requires a strong CPU to support the calculation of the traffic scheduling algorithm, and image/ video processing require support of GPU.

- Lightweight Virtualization Framework

In addition to the basic network functions and management functions based on hardware, the edge computing gateway implements industrial applications, ecological services, and application assistance functions based on lightweight virtualization technology.

Considering the number of gateway deployments and the implementation of function, the management of virtualized resources must be lighter than the edge data center, eliminating unnecessary modules. Traditional architectures include the allocation of IaaS resources, management and mutual communication of PaaS, as well as SaaS authentication and deployment. In a single-function scenario, the edge computing gateway may not fully deploy the traditional virtualization architecture or even abandon it.

- WAN side interface layer

The WAN interface supports multiple network access methods include wired and wireless, such as 4G LTE/5G, Ethernet, and PON. It also supports unified application layer protocol of vertical industry to better interface with cloud or edge data centers, such as the industry-wide unified protocol OPC UA and the common protocol in Internet of Things MQTT, etc.

2）Capability

Based on the above architecture, the edge computing gateway can empower the vertical industry.

- Data Collection and Protocol Interworking

The ways to access the Internet of different vertical industries is various. The edge computing gateway can implement normalized conversion between different protocols, solve the problem of information islands in different systems, and improve the efficiency of business data processing.

- The Formation of a Deterministic Network

It supports deterministic network technologies such as TSN, optimizes forwarding scheduling mechanism's optimization, and introduces new technologies such as high-precision time synchronization and time slot preemption to enhance the determinism of network delay.

- Multi-ecosystem

It can carry different third-party PaaS and application ecosystems, such as Amazon Greengrass, Microsoft Azure, etc., and achieve business isolation through lightweight virtualization technology, so that different ecosystems of the same node do not interfere with each other.

- Functional Modular Design

It can package the basic functions that software can implement in a service manner, and support on-demand remote deployment, thus achieving low-cost software customization needs. Modular design concept provides flexible cutting and expansion capabilities.

- Localized Artificial Intelligence

The edge computing gateway can have some intelligence, such as executing the artificial intelligence matching algorithm, and processing the service request locally. Meanwhile, it can feed back to the cloud for data backup and iterative training of the model.

6. Building an Industrial Ecology

6.1. Establishment of Open Lab

China Mobile established the Edge Computing Open Laboratory on October 30, 2018, which is committed to providing an industry cooperation platform, merging the advantages of edge computing in various industries, and promoting the prosperity of the edge computing ecosystem. Currently, there are 34 partners. The edge computing ecology is fragmented, and each industry explores it in their own. To solve the problems in the current field, the open laboratory has formulated the following specific targets.

1）**Free and Open Participation. Cross-industry Cooperation.**

Open Laboratory Welcomes IT/CT/OT Cross-border Cooperation and Promotes Collaborative Innovation of Industry, University and Research

2）**Easy-access Lab Resources. Shared Research Outcomes.**

Open laboratories will provide a platform for collaborative integrated research and development. In open laboratories, systems, capabilities and results will be fully open.

3）**Use-case-driven &Application-centric.**

Laboratory will focus on demand guidance and practical application, thus enabling vertical industries to promote commercial use.

The open laboratory internally coordinates China Mobile's various industry research institutes, provincial companies and professional companies, externally condenses partners in various fields, standardization organizations, industry alliances, etc. It has s three working groups, namely the overall technical group, product integration group and application promotion group.

The overall technical team will focus on the overall architecture of edge computing, industry standards and open source projects. It will build China Mobile's edge computing certification system in the future. The product integration group is mainly responsible for the research and planning of edge computing platforms and hardware, which will become the most important aspect of industry promotion. The application promotion group is responsible for the establishment and promotion of edge computing solutions, which initially focus on the test bed. Thus, technology, integration and application jointly promote and cooperate with each other to create a complete edge computing promotion strategy.

6.2. Product of the Open Lab

The development of edge computing requires a unified end-to-end full-stack system. Combined with China Mobile's edge computing technology architecture, the open lab will use platform and hardware development as its basic strategy. The combination of platform and hardware will build the edge computing service capability oriented to full connection and full service, and lay a good foundation for condensing edge computing industry resources.

On the platform side, the open lab will build Sigma platform that supports edge cloud construction based on the 3-level implementation options. Sigma will provide management, network and industrial API. Meantime, the open lab will also provide centralized PaaS management and lightweight PaaS implementation choices.

In terms of hardware, the open lab will accelerate the development of edge hardware ecosystem by investigate customized server, all-in-one solutions and edge computing intelligent gateways.

6.3. Resource and Capabilities

The open lab currently has full-stack service capability, which can be used by partners for technical research and application deployment. In terms of access capacity, the laboratory can provide 4G/5G wireless access and broadband wired access network capability. In terms of hardware, the laboratory can provide OTII server and embedded gateway equipment. In terms of basic resources, the lab can provide IaaS capabilities including virtual machine/container.

In terms of API, the Sigma platform released by the laboratory currently has 6 classes and more than 30 kinds of network API capabilities for 5G, and can provide scheduling capabilities for multiple applications interworking. The open lab will continue to improve and upgrade the capabilities available to serve the development of edge computing technologies and applications.

6.4. Application and Testbed

The application field of edge computing is very extensive. Considering the most possibility of commercialization and development potential, the open laboratory initially lays out four areas of smart city, intelligent manufacturing, live games and vehicle interconnection.

At present, the Open Laboratory has carried out a total of 15 test bed projects with representative partners in various fields, including 4 smart cities testbeds, 6 smart manufacturing testbeds, 4 live streaming and gaming testbeds and 1 vehicle interconnection testbeds. The first batch of test bed projects integrates PAAS resources of many vertical partners, covering new technologies such as high definition video processing, vPLC, artificial intelligence, TSN, etc. and involves many scenarios such as intelligent building, intelligent construction, flexible manufacturing, CDN, cloud game and vehicle interconnection. The test bed project will serve as an important reference to form several industry solutions in various fields in the future and promote the commercial deployment of edge computing

Table 2 List of Testbed Project in China Mobile Edge Computing Open Lab

Number	Name	Area	Cooperation
1	Smart City Video Networking Service Platform Based on Mobile Edge Computing	Smart City	Inspur
2	Smart City Video Networking Service Platform Based on Mobile Edge Computing	Smart City	Tridium
3	Seven-layer full-scale Experimental Intelligent Construction Pilot Based on Edge Computing Service Architecture	Smart City	China Construction Group
4	Application of Edge Intelligence in Smart City	Smart City	Alibaba
5	Digital Production Line Based on TSN and vPLC	Intelligent Manufacturing	Huawei
6	Industrial Flexible Manufacturing Based on Wise-PaaS	Intelligent Manufacturing	Advantech
7	Smart Factory Testbed	Intelligent Manufacturing	Ericsson
8	Smart Test Eye: An Automatic Detection Scheme for Intelligent Manufacturing	Intelligent Manufacturing	Lenovo
9	Industrial Vision Application Based on OpenIL	Intelligent Manufacturing	NXP
10	New Network of Industrial Internet Testbed (opcua&tsn)	Intelligent Manufacturing	CertusNet
11	Application of Edge Computing in CDN	Live and Games	Tencent
12	5G Fast Game Based on Edge Computing	Live and Games	China Mobile
13	MEC Based 8K 360° VR Video Broadcasting	Live and Games	Intel
14	12K VR Panoramic Video On Demand based on Edge Cloud	Live and Games	China Mobile
15	Application of Edge Computing in Vehicle Interconnection	Vehicle Interconnection	Baidu

7. Vision and Future Work

Edge computing is yet in a pre-mature stage, facing many problems. These include incompatible technical system, lack of standardization, uncertainty of collaboration model and fragmented ecosystem. China Mobile endeavors to continuously contribute to the industry and establish full-stack capability, build-up open collaboration platform and accelerate the prosperity of applications.

In October 2018, with 14 distinguished fellow and industrial leaders, China Mobile release it vision on “Collaborative Promotion of Edge Computing Technology and Industry Development”. In this vision, China Mobile proposes the following.

(1) Establish platform for cross-industry collaboration

(2) Clarify technical architecture and improve standardization work to build up full-stack capabilities

(3) Encourage Use-case-centric innovation

(4) Build-up open industrial ecosystem for joint development

China Mobile actively reacts to the above promotion. In MWC201, China Mobile released the “Pioneer 300” action on edge computing. This action delivers a clear strategic roadmap to the industry. In this action, China Mobile announced the enablement target in 2019 on resource, platform and ecosystem aspects:

(1) Evaluates 100 edge-computing-ready sites

(2) Exposes 100 edge capability APIs

(3) Develops 100 partners in edge computing open lab

As a leading telecommunication service provider, China Mobile will continue to promote edge computing technology and industry development under open and collaborative principles, creating new opportunities for cross-field innovation and transformation of vertical industries.

China Mobile Edge Computing Open Laboratory