5G SLICING ASSOCIATION

5G Network Slicing Self-Management White Paper

October, 2020

中国移动 China Mobile | 研究院 CMRI

China unicom 中国联通 创新·与智慧同行

中国电信 CHINA TELECOM

国家电网 STATE GRID 中国电力科学研究院有限公司 CHINA ELECTRIC POWER RESEARCH INSTITUTE

腾讯游戏 Tencent Games

CSM 体奥动力 China Sports Media

DIGITAL DOMAIN 数字王国

GREE 格力

HUAWEI

AsiaInfo 亚信科技

Contents

Introduction

4G changed lives, whereas 5G will change society. 5G will radically reshape mobile communications, and has thus been given top priority in the New Infrastructure initiative proposed by the Chinese government. With this initiative at its foundation, 5G will kick the national economy up a gear. It is already underpinning a wide range of industries, and will go beyond our expectations, empowering more fields. With the finalization of 3GPP R16 standards and the advancement of 5G standalone (SA) network construction by major operators in China, 2020 will see the debut of the 5G B2B market. However, no B2B market can be established without network slicing. Network slicing is one of the key solutions differentiating 5G from 4G, and is a cornerstone for the 5G B2B market.

Now, as we are on the cusp of 5G SA network construction, we are able to configure end to end network slices for services, and have established benchmarks for application in various vertical sectors, benchmarks that prove the feasibility of network slicing. However, the current approach to configuration, which is static, manual, and divided between industries, does not lend itself well to large-scale replication for greater commercial value. In this context, a slice management system which allows for automation and self-management becomes a must.

To this end, 5G Slicing Association (5GSA) has conducted extensive surveying and research on network slicing management to formulate this white paper. This paper is an extension of the Categories and Service Levels of Network Slicing White Paper released in March 2020, and centers on slicing self-management. It analyzes the common requirements on network slicing self-management between industries, with multimedia, electric power, industrial Internet, and healthcare sectors as representatives; it proposes a management solution and architecture; and suggests a roadmap for industry development to keep pace with the advancement of standards and products. We hope this paper can inspire further exploration into network slicing and break through the barriers to commercialization, facilitating the digital upgrade of vertical industries.

Contributors:
China Mobile, China Unicom, China Telecom Research Institute, China Electric Power Research Institute, Tencent Games, China Sports Media (Beijing), Digital Domain Group, Huawei, AsiaInfo Technologies

Individual contributors:
Sun Xiaowen, Shi Xiaonan, Tong Lei, Chen Xuan, Pan Guixin, Li Lisha, Guo Rui, Wang Duyan, Wang Heng, Xia Xu, Mei Chengli, Xing Yanxia, Wang Zhihui, Wu Sai, Meng Sachula, Hu Yue, Li Zhi, Hong Kai, Xu Shili, Du Juan, Lu Yanhui, Zhang Kaifeng, Liuteng, Kang Sheng, Wang Xiaolin, Hua Shuai, Nie Xin, Zheng Yiqi, Zhang Yang, Ren Yaqi, Tian Yuan, Li Shaobin, Ke Dong, Li Bin, Xie Yidong, Yang Xiaohua, Song Xuetao, Guo Jun, Ren Yongzheng, Wang Xiaoming, Ma Xu, Song Jian, Mao Qianyong, Huang Luomeng, Zhu Duozhi, Zeng Gang, Wang Lilei, WangZhigang

02 Concepts, Application, Development, Practice

2.1 Concepts

2.1.1 Network Slicing

Unlike services for individuals, vertical industries require networks that offer deterministic services, with assured latency, jitter, and packet loss rate. Such requirements can hardly be satisfied by the 4G networks, which are essentially just one big pipe with best-effort delivery of network-defined services. Conventional networks need to upgrade to 5G with its differentiated capabilities, to fulfill the varied SLA requirements and to adapt to the ever-increasing number of applications. Network slicing is key to this shift. With network slicing, a physical network can be sliced into logical networks which run on shared infrastructures but target different industries and applications. These logical network slices are provided to tenants together with a certain level of access, allowing them to customize their network slices while assuring SLA needs.

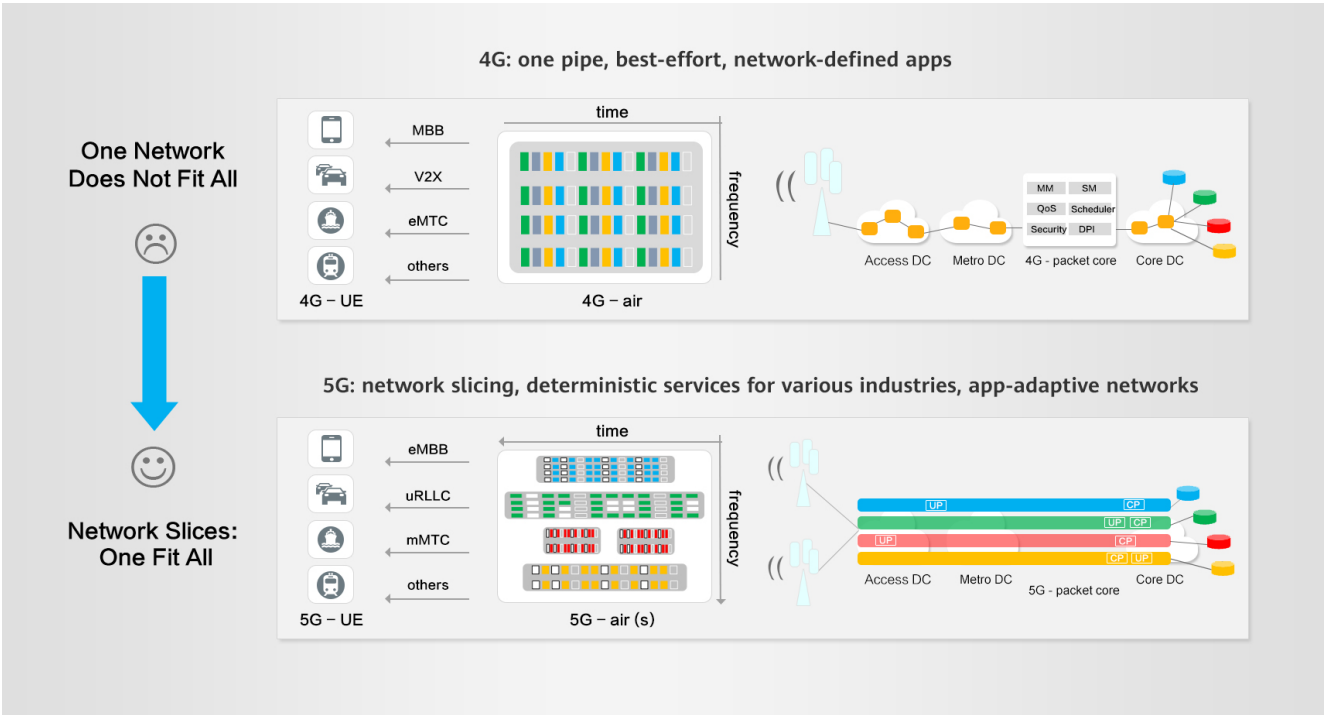


Figure 1 Network slicing

2.1.2 Three-Layer Industry Requirements

Abraham Maslow was an outstanding psychologist. He proposed a pyramid-shaped model for analyzing innate human needs, where individuals must satisfy needs at the lower level of the hierarchy before progressing to those above. This model has become a general-purpose theory used to classify needs. We reviewed one hundred scenarios of 5G service application in ten industries and classified the industry requirements on 5G networks into three categories according to this hierarchy of needs. From bottom to top, these are: deterministic services, security and trustworthiness, and self-management.

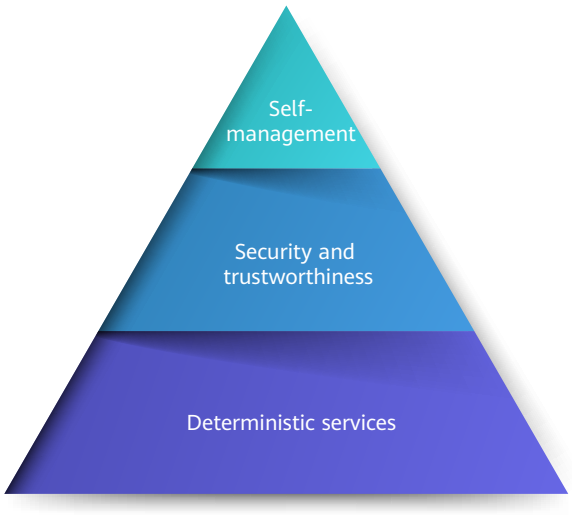


Figure 2 Maslow's model illustrating industry requirements on communication networks

Deterministic services

Different industries, services, and user groups impose different requirements on bandwidth, latency, jitter, and time synchronization, but all of them require a deterministic service experience. To achieve this, the network must be able to adapt to every need with guaranteed QoS or SLA.

Security and trustworthiness

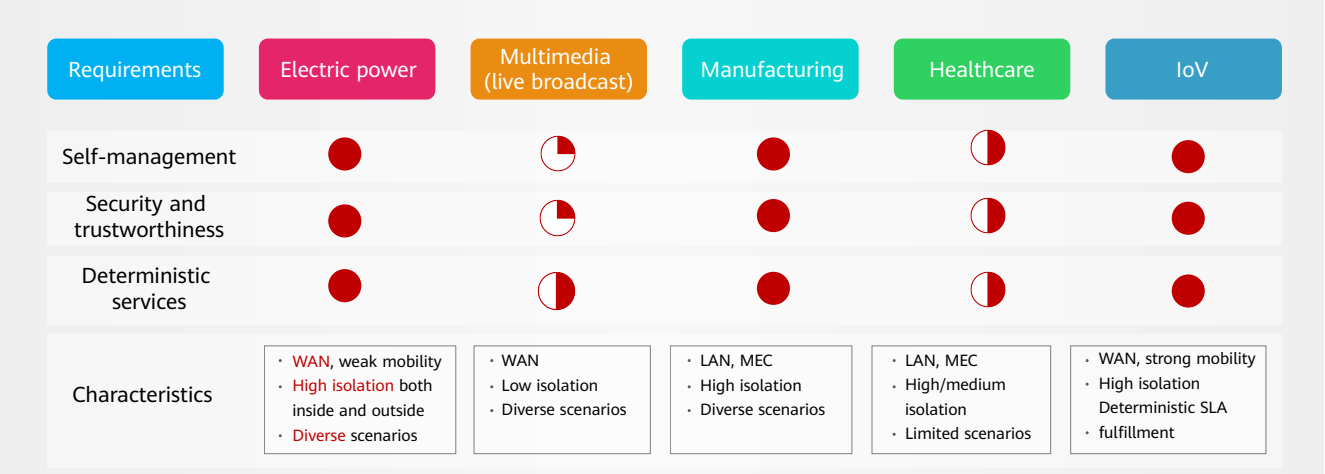
Different industries demand different approaches to protection, such as isolation, data encryption chips, IPSec, integrity check over air interfaces, and secondary authentication, to secure their services at appropriate levels. For example, sectors with predominant WAN coverage, like electric power and public security, require that physical resources be stringently isolated, whereas those with only LAN coverage, such as manufacturing and healthcare, demand that data transmission be restricted to within campuses.

Self-management

As network as a service (NaaS) grows in popularity, industry tenants require more visible management of their networks, and a few of these tenants with ICT professionals expect self-management and self-maintenance, and even self-service and self-operations.

2.2 Application and Development

If we look at how network requirements change with network scale, one rule stands out immediately: industries which require broader network coverage and higher isolation for a larger variety of service scenarios are in more urgent need of network slicing. Electric power and multimedia are two such industries, which will be some of the first to benefit from network slicing.



Typical Slicing Scenario	Requirement Urgency
Slices for power distribution, utilization, and transformation in all scenarios	Strong
Broadcast slices	Medium
Campus intranet slices	Medium
Hospital treatment slices	Weak
Slices for all IoV scenarios	Strong

Figure 3 Typical slicing requirements and applicable scenarios

The development of network slicing will go through three phases.

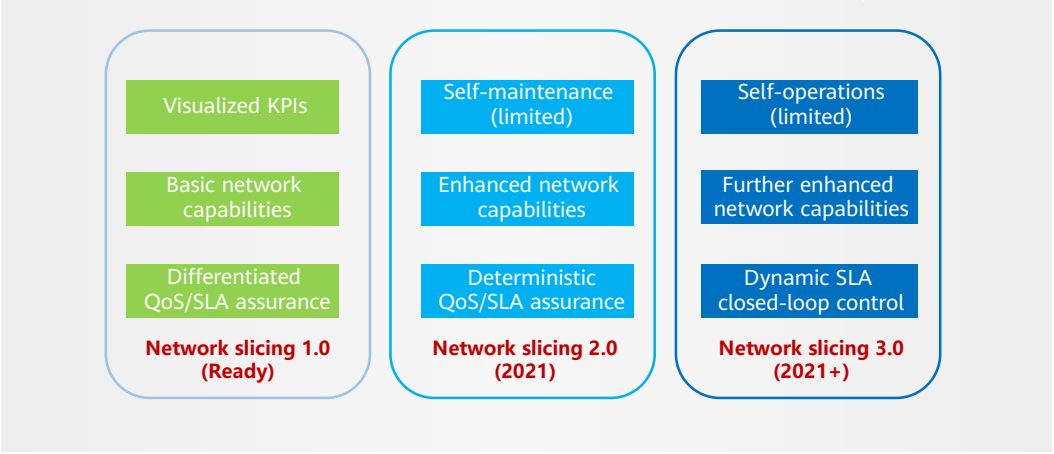


Figure 4 Network slicing industry pace

Phase 1

- The 5G transport and core networks can implement different isolation solutions, through either hardware or software. On the 5G NR side, 5QIs (QoS scheduling mechanism) are mainly used to achieve software-based isolation on WANs. Alternatively, campus-specific 5G NR (including micro base stations and indoor distributed base stations) is used to implement hardware-based isolation on LANs.
- 5QIs are used to for differentiated SLA fulfillment on different slices.
- KPIs covering the entire network can be viewed on an intuitive display.
- From 2020 onwards, E2E network slicing solutions can be delivered commercially for services such as ultra-HD live broadcast and AR advertising, with L0 and L1 slices on public networks and L2 slices on industrial networks. However, for solutions oriented to L3 and L4 slices, solution providers will need to discuss delivery details with customers in depth based on their capabilities.

Phase 2

- The 5G NR side will be enhanced to reserve resource blocks (RBs) so that wireless network resources can be isolated for WANs.
- 5G LAN and TSN technologies will be leveraged to further guarantee differentiated SLA requirements.
- Self-service operations, though limited, will be offered to tenants with the visualized KPIs for tenant networks provided in phase 1.
- Operators can better serve VIP and VVIP customers of industrial networks. VIP services include AR/VR cloud gaming and drone inspection, and VVIP ones have electric power production control, smart industrial campus, and public security.

Phase 3

Adaptive optimization and dynamic closed-loop control of SLA will be achieved through AI and a negative feedback mechanism. Network slicing can better serve industries with high requirements on mobility, roaming, and service continuity, such as IoV. In addition, more comprehensive service capabilities will be provided to industries.

Right now, we are preparing to move towards phase 2 with a series of novel technologies such as RB reservation and 5G LAN. Considering the technology maturity, network requirements and dependency of different industries, and the development plan of the industry chain, we recommend that network slicing be applied in three waves, with multimedia and electric power first, industrial Internet and healthcare second, and IoV last.

2.3 Practices

Since 2018, operators, industry vendors, and telecom equipment providers in China have jointly carried out 5G network slicing research and pilot projects in an array of industries, and have made remarkable progress.

In the electric power sector:

- China Telecom, the State Grid Corporation of China (SGCC), and Huawei, have completed field tests on network slicing in a series of typical services in Nanjing and Beijing. These services include millisecond-level load control, situational awareness of the power distribution network, 4K ultra-HD video for power transmission lines, intelligent distributed feeder automation, and differential protection. They have also launched the 5G+Smart Grid experimental network in Qingdao, the largest of its kind in China.
- China Mobile, China Southern Power Grid, and Huawei have completed industry's first field test on an electric grid slice for commercial use in Shenzhen, with the focus on differential protection and isolation between the core and bearer networks. They were also the first in the industry to build up a 5G B2B slice management platform. From these practices, China Southern Power Grid acknowledged virtual private networks as a valuable aspect of network slicing, as they are highly secure, reliable, flexible, and can be customized for the power grid.

In the multimedia field:

- China Mobile, MIGU, China Media Group, and Huawei have completed field tests on network slicing in 4K ultra-HD multimedia live broadcasts in Shanghai and Beijing.
- China Unicom and Huawei have taken a leading role, verifying version 1.0 of the 5G HD live broadcast network slice for use and standardizing the configuration of live broadcast services. The two parties have also planned the services for version 2.0 — live broadcast services for TV news, stadium events, and influencer platforms in social media will be guaranteed with differentiated SLA assurance.
- China Mobile, China Unicom, Tencent Games, and Huawei have carried out field tests for AR and cloud gaming in Nanjing, Hangzhou, and Shenzhen. With enhanced bandwidth and reduced latency, players can enjoy a more immersive experience.

In the industrial Internet segment:

- China Unicom, Gree, and Huawei have built up China's first MEC+Smart Manufacturing network with the SA architecture. Paperless inspection and intelligent video analysis have been verified as feasible on the production line.
- China Mobile, Haier, and Huawei have tested the entire network slicing workflow in industrial networks in Qingdao, covering everything from slice purchase to delivery and O&M in smart manufacturing, smart home, smart logistics, and smart campus.

For healthcare, China Mobile, Futian District Health Bureau, the Eighth Affiliated Hospital of Sun Yat-sen University, and Huawei have jointly piloted an MEC+Smart Healthcare medical service community. In this project, the existing public network is leveraged together with network slicing to create what is essentially a private network for healthcare services. With local breakout (LBO) and strengthened bandwidth, health services, such as medical image storage and reconstruction, will be processed faster than ever.

03 Self-Management Requirements

As network slicing pilots sprout up in the B2B field, more and more industry players are acknowledging how crucial network slicing is to achieving their digital transformation goals. As for how to achieve these goals, there is one thing they all agree on: an automated slicing management system will accelerate transformation, and this system must be open to tenants for self-management. As such, understanding the self-management requirements of various industries is key to unleashing the infinite potential of network slicing. This section covers the requirements for four major fields: multimedia, electric power, industrial Internet, and healthcare.

3.1 Stadium HD Live Broadcast

3.1.1 Services

Big sports events, such as the Chinese Super League, usually take place in stadiums, often have long seasons, dense schedules, and are held in different venues with concurrent matches, making live broadcasts challenging. To provide an exciting program for their fans, video and audio need to be collected via lightweight transmission devices during the matches, processed in a remote center and pushed to fans. To address such requirements, the live broadcast network slice needs to support:

- Heavy live broadcast traffic in the uplink direction
  The slice must ensure the quality of local transmission with innate 5G features — high bandwidth, low latency, and wide coverage. Video signals from cameras such as Steadicam and VR cameras, need to be transmitted from various positions over wireless channels, with a delay of less than 1 frame, to the outside broadcast (OB) vans.
- Remote bidirectional control
  The slice must ensure the quality of transmission between venues and the remote production center. Lightweight transmission devices need to be deployed at various locations within these venues, where video signals, multi-track audio streams, sports event data, low-latency voice services, and control messages can be transferred bi-directionally over wireless channels. The remote production center uniformly controls the devices at these venues, using their data to construct programs for audiences.

To underpin the preceding two aspects, the slice must provide:

- Stable bandwidth
  A single channel of HD video (1920 x 1080 pixels) needs to be encoded at a bit rate of 15 Mbit/s. Therefore, the uplink bandwidth of the transmission channel cannot be lower than 15 Mbit/s, and the network latency must be controlled to within 1s. When multiple channels of videos are transmitted, the latency must be restricted to within 40 ms (approximately one frame) to prevent bit errors, display artifacts, or frame loss.
- Disaster recovery for network transmission
  All transmission faults must be broadcasted over the network so that services can be switched over swiftly, and the faults rectified promptly. Service transmission must be consistently stabilized and secured.
- Data security
  Live broadcast signals need to be transmitted over a wired private network which has higher security than public networks. Despite this, the 5G wireless network must still have sound anti-theft to carry copyright-sensitive information for sports events.

3.1.2 Self-Management

	Maintenance	Operations
Visualization	• Both physical and logical resources for transmitting signals can be checked and maintained. The KPIs pertaining to bandwidth, latency, and traffic can be monitored. • Network operating status, data security, and system performance can be monitored. Alarms can be generated for exceptions such as insufficient bandwidth or unacceptable jitter. • The status of SLA fulfillment for use cases, network equipment, and terminals can be monitored.	• All information for industry slices, including signal transmission paths, can be monitored. • SIM cards, slice purchase orders, and service bills can be checked.
Management	• Concurrent signal processing at different sites can be managed. Data configurations can be issued from the remote control center to centrally enable, guarantee, and disable services. • Dialing tests can be performed. • Services can be switched over securely when faults occur. Prompt emergency response measures can be taken.	• Signal transmission for slice subscription, updates, and cancellation can be managed. • SIM cards can be managed; services can be activated, suspended, and resumed for users; user groups can be customized; and slices can be assigned to or withdrawn from users.
Control	• Network access can be controlled. • Service access can be controlled (for example, via a blacklist and whitelist).	Charging policies can be customized.
Capability exposure	Northbound APIs could be exposed, but operators can provide portals for tenants instead.	

HD live broadcast slices have unique requirements that need to be satisfied for self-management:

- Plug-and-play deployment of onsite transmission devices
  The transmission devices can be rapidly deployed and configured at venues, and are easy to troubleshoot. The devices can be monitored and controlled remotely to guarantee efficient signal transmission at venues.
- Service subscription and change
  The transmission service can be subscribed to in advance of a live broadcast (up to 5 hours in advance in extreme cases.) This can be changed up to 30 minutes before the live broadcast starts, and the broadcast can be canceled at any time. This means that the service is still available if a match must be extended or rescheduled due to unexpected factors, such as bad weather.

3.2 Cloud Gaming

3.2.1 Services

Cloud gaming leverages 5G to provide an enhanced gaming experience to players. Game logic is computed by servers on the cloud, as well as rendered images, video streams, and instructions, which are all sent to terminals via the 5G network. Cloud games combine the advantages of console games and mobile games, addressing player demands for quality and convenience; and also breaking free from the hardware limitations imposed by console games and mobile games, allowing games to run on many different types of terminals. Different cloud games have different requirements on network latency, bandwidth, packet loss rate, and stability. To meet these individual requirements, different levels of 5G slicing services need to be provided.

- Assured bandwidth
  - √ Mobile games: 4 Mbit/s for SD, 6 Mbit/s for HD, 10 Mbit/s for ultra HD
  - √ Console games: 25–35 Mbit/s
- Stabilized latency
  The network transmission quality must at least be equivalent to that on fixed networks. The end-to-end network latency must be within 20 ms, and the packet loss rate less than 0.04%.
- Edge computing
  Edge computing nodes are deployed to meet the strict requirements for latency and bandwidth in some cloud games.

3.2.2 Self-management

	Maintenance	Operations
Visualization	• A unified interface is provided, through which maintenance engineers can check whether a user can use 5G services, whether the user is using 4G or 5G services, and whether the user has been assigned a network slice. • Slice usage can be monitored by service or end user usage (including bandwidth, latency, packet loss, duration, and traffic). • The 5G slice usage of end users can be monitored by region or by operator.	• The 5G service usage of end users can be monitored. • End users can subscribe to slicing services. • The slice usage can be monitored by service or by end user. 5G bandwidth and other real-time performance KPIs can be viewed by region.
Management	• Upon a fault, the slicing services can be disabled by region, service, or user in real time. • Network fault notifications can be pushed to engineers in just a few minutes, together with additional information such as troubleshooting contact persons. Maintenance engineers can perform dialing tests and troubleshooting with self-service. • Interfaces and data for dialing tests between terminals and air interfaces are available, facilitating fault identification. • An interface must be provided, through which maintenance engineers can be notified of inadequate 5G KPIs for users. 5G slicing services can be enabled or disabled for individual users at any time.	Faults and their causes can be displayed by region or service.
Control	• 5G slicing services can be subscribed to, enabled, or canceled by service, operator, or region at any time with self-service. • The bandwidth for 5G slicing can be enhanced with self-service. • Alarms can be reported for insufficient bandwidth. The emergency buffer can be expanded automatically upon exceptions. • Edge computing nodes can be expanded, reduced, or deleted with self-service.	Charging policies and results can be checked for each service and tenant.
Capability exposure	Northbound APIs need to be integrated into the management platform of cloud gaming services.	

Cloud gaming slices have unique requirements that need to be satisfied for self-management:

Fault locating

- Network exceptions, including jitter, service interruption, and faults, can be detected and pushed to engineers within several minutes.
- The unexpected network KPIs can be identified and pushed to engineers in real time via well-defined interfaces.
- Slicing services can be controlled remotely by region, operator, or end user. Network slicing can be forcedly terminated by region, operator, or service at any time to respond to faults, and the fault information and impact scope can be easily identified.
- Self-service dialing tests can be conducted through well-planned interfaces between terminals and air interfaces.
- Disabling and enabling of slicing services takes effect immediately.

5G slicing data platform

- This platform needs to provide a standard interface to check whether a user can access 5G services, when and how long the user plays cloud games, the latency, and consumed traffic. The collected data can be used as reference for handling user complaints and charging issues. This interface must be compatible with terminals from any vendor and of any model or OS version.
- This platform can display an overview of slicing service usage and distribution to operations personnel. It also can process the subscription and cancellation requests from end users.

Private network configuration

- On a visualized control platform, tenants can apply for edge computing nodes, expand or reduce node capacity, monitor node operating status, and manage operation rights on these nodes.
- Faulty edge nodes can be deleted automatically, and then system capacity can be expanded accordingly.
- A private network can be started and stopped by region, operator, or service. These configurations take effect immediately.

3.3 Power Grid

3.3.1 Services

To run a smart grid, the 5G network needs to sense and allow access from all devices and personnel involved in power service phases, including power generation, transmission, transformation, distribution, and consumption. The network also needs to stay online at all times, exchanging information generated during each power service phase. To meet these requirements, power grids need to rely on 5G to strengthen both existing and new power grid services, strength which comes from three key features — URLLC, mMTC, and eMBB. With these features, different slices can be tailored to fit different service scenarios on a power grid, as different scenarios impose varied requirements on latency, bandwidth, and reliability. For example, precise load control requires a latency of less than 50 ms, while AR-assisted inspection demands a bandwidth of 20-30 Mbit/s. Additionally, different isolation levels and automatic O&M capabilities can be defined for each slice, further lifting service efficiency.

5G Feature	Typical Power Grid Service	Communications Requirements	Slicing Requirements
URLLC	Differential protection	• The bandwidth must be greater than 2 Mbit/s. • The end-to-end unidirectional latency must be less than 12 ms. • The jitter must be within around 50 us. • High-precision timestamps must be used for information exchange between terminals, and time must be synchronized to within 10 us. • The service reliability must reach 99.999%.	• Isolation These two services are implemented in production zones (zones I and II), which must be strictly isolated from services in management zones (zones III and IV). Dedicated resources must be physically isolated for the two services. • Slicing Hard slicing is applied, offering exclusive resources. • Communications The transmission paths between any two service terminals, which do not take a detour to the central network, and instead stay at the network edge, must always be active. CPE can be used for frequent communications. • Ports Ports can be flexibly selected. Independent Ethernet or RS485 ports can be used.
	Precise load control	• A bandwidth of 48.1 Kbit/s to 1.13 Mbit/s is required between a terminal and the central network for second-level to minute-level load control. • The end-to-end unidirectional latency must be less than 50 ms. • The service reliability must reach 99.999%.	
mMTC	Power consumption data collection	• A bandwidth of 10 kbit/s to 100 kbit/s is required. • The end-to-end unidirectional latency must be controlled to within 0.5s to 5s. • The one-time collection success rate must be at least 97%, and the remote control accuracy at least 99.99%.	• Isolation These two services are implemented in the management zones (zones III and IV). The resources for them need to be logically isolated, and related data needs to be secured through encryption and authentication. • Slicing Soft slicing is applied to create an IoT network, sharing resources with other services. • Communications In areas with high electricity usage, more than ten thousands of connections need to be set up. The transmission paths between service terminals and the central network must always be active. Edge computing can also be considered for these two services. CPE or communications modules can be used if they communicate at 5 minute intervals. • Ports Ethernet, TTL, and RS232 ports can be used.
	Distribution network status monitoring	• Each monitoring node requires a bandwidth of 1 to 10 Kbit/s for 100-ms to second-level monitoring. • The service reliability must reach 99.999%.	
eMBB	Intelligent video analysis	• A bandwidth of 4-10 Mbit/s is required. • The latency must be less than 200 ms. • The service reliability must reach 99.9%. • The required transmission rate varies according to the video definition: usually greater than 2 Mbit/s for an SD video channel and greater than 4 Mbit/s for HD. However, the downlink control signals require less than 10 Kbit/s.	• Isolation These two services are implemented in a management zone (zone III). The resources for them need to be logically isolated, and related data needs to be secured through encryption and authentication. • Slicing The slices are provided with high bandwidth. • Communications The transmission paths between service terminals and the central network must always be active. Edge computing can also be considered for these two services. • Modules and ports Communication modules and Ethernet ports can be used.
	Intelligent inspection	• The data transmission rate must be greater than 2 Mbit/s. • The latency must be less than 300 ms. • The overall service reliability must reach 99.99%, and the reliability for remote drone control must hit 99.999%.	

3.3.2 Self-Management

	Maintenance	Operations
Visualization	• End-to-end KPIs can be displayed by equipment or terminal. • Alarms can be generated for exceptions. • Network KPIs can be monitored.	• All information for industry slices can be monitored. • SIM cards, slice purchase orders, and service bills can be checked.
Management	• Electric power slices can be customized. • New services can be commissioned with self-service before rollout.	• Signal transmission for slice subscription, updates, and cancellation can be managed. • SIM cards can be managed. Services can be activated, suspended, and resumed for users. User groups can be customized. Slices can be assigned to or withdrawn from users.
Control	• Network access can be controlled. • Service access can be controlled (for example, via a blacklist and whitelist).	Charging policies can be customized.
Capability exposure	Northbound APIs need to be integrated into the management platform of power grid services.	

Power grid slices have unique requirements that need to be satisfied for self-management:

- Maintenance

  The service attributes, resource topologies, performance parameters, and operating status of power grid slices need to be displayed in an intuitive interface, so that electric power departments can manage services and adjust service policies efficiently.

  In addition, electric power departments expect a certain level of network management rights will be assigned to them by the operator. For example, the electric power departments can select additional network functions when subscribing to slicing services, bring a slice online or offline, or activate and deactivate services.

- Operations

  Power grid operations departments need to take responsibility of power network operations. They need the right to activate and deactivate SIM cards, register slicing services, and check for service and card status. These departments also need basic information for the slices they purchased, and appropriate permissions to modify or update these slices. In addition, power grid customers must be notified of faults on operator networks through alarms, so that prompt measures can be taken to stabilize power services.

The electric power industry expects great things from 5G networks, and the industry is consistently diving into how to better pair 5G network resources with power grid services. The generic slice template (GST) defined by the GSMA is handy. The management parameters in the template can be planned for different power grid scenarios, services, and communications requirements. The key parameters include end-to-end latency, transmission rate, jitter, packet loss rate, security, reliability, area capacity, user density, and coverage area.

3.4 Industrial Internet

3.4.1 Services

5G is borne out of industrial scenarios, including device communications and control, intelligent video analysis and machine vision, as well as AR and cloud robotics. These services require different levels of bandwidth, latency, and jitter. Device communications depend on the strong connectivity and ultra-low latency provided by 5G, as it facilitates data collection from and remote control of sensors in factories, and remote operations on vehicles and drones. As for the higher bandwidth, intelligent video analysis and machine vision can be applied in production line monitoring, product quality inspection, and automated guided vehicles (AGVs). Furthermore, industrial AR relies on high bandwidth and low latency. In smart manufacturing, AR technologies help operators collaborate with machines in a concerted way, and have been proven viable and instructive in task guidance and personnel training. Cloud robotics free robots from their wire shackles, giving them the mobility required for wider application.

Industrial sector requirements on network slicing can be classified into two categories:

- Network slicing can be used to differentiate campus services into intranet and extranet ones. The transmission of data generated for intranet services needs to be restricted to within the campus to keep it secure while providing low latency. The extranet services can be routed to Internet via the public 5G core network.
- The campus services can be further classified based on their SLA/QoS requirements and carried in three slices:

  √ Industrial control, such as crown block control

  √ Production assistance, such as AI-powered machine vision

  √ Non-production activities, such as mobile office and intelligent video analysis for security surveillance

  Among these slices, the industrial control slice has the highest priority.

  In this way, different enterprise services can be isolated and secured.

3.4.2 Self-Management

	Maintenance	Operations
Visualization	• The slice operating status can be monitored. • The location, physical and logical configurations, and services of a slice can be checked. • System performance can be monitored. Alarms can be generated for exceptions. • The status of SLA fulfillment for slices, network equipment, and terminals can be monitored.	• Basic slice information can be displayed, including slice usage, capacity, transmission rate, connections, region, and service density. • Slice users, bills, other charging information, and SIM cards can be checked.
Management	• Slicing tasks, slice users, slice life cycles, and slice instances can be managed.	• Signal transmission for slice subscription, updates, and cancellation can be managed. • SIM cards can be managed; services can be activated, suspended, and resumed for users; user groups can be customized; and slices can be assigned to or withdrawn from users. • The binding between terminals and SIM cards can be managed.
Control	• Network access can be controlled from one or multiple aspects, including region, time, traffic, frequency, and service level. • Service access can be controlled, and the UPFs can be scheduled.	Policies for network access, service levels, security, and charging can be customized.
Capability exposure	Northbound APIs could be exposed, but most industrial enterprises call for a B2B self-service portal from operators.	

Industrial slices have unique requirements that need to be satisfied for self-management:

- Visualization

  In addition to slice operating status, the status of SLA fulfillment for devices and terminals needs to be monitored.

- Management and control

  Network and service access from slice users can be controlled flexibly; the UPFs can be efficiently arranged and selected for different services; and different types of network policies can be configured.

- Capacity exposure

  A self-service operations portal can be provided to Industrial enterprises, or their own campus management platform can be integrated with the capacity exposure platform. Enterprises with strong ICT capabilities may require open northbound APIs.

3.5 Healthcare

3.5.1 Services

The healthcare service scenarios can be classified based on the places where they are mainly provided:

- Inside hospitals

  Mobile medical devices are used in hospitals, where they need to be connected to 5G networks to support mobile nursing, mobile ward rounds, and patient condition monitoring. By harnessing the high bandwidth and low latency of 5G networks, the efficiency and accuracy of services inside hospitals can be significantly improved.

- Between hospitals

  Various telemedicine services are provided between hospitals, for example, patient medical data, either audio or video, can be transferred between hospitals, including results for ultrasounds, imaging, and endoscopy examinations. This requires that medical devices be remotely controlled.

  Thanks to the high bandwidth and low latency of 5G networks, patient files and imaging can be loaded in seconds. Ultra-HD videos can be transmitted in real time without any loss, and multiple doctors can view a patient's data at the same time.

- Outside hospitals

  Dynamic data, such as patient vital signs, can be transferred to hospitals from ambulances during emergency rescue via a stable, secure network.

  Network slicing can be leveraged for scenarios between and outside hospitals.

3.5.2 Self-Management

	Maintenance	Operations
Visualization	• The slice operating status can be monitored. • The location, physical and logical configurations, and services of a slice can be checked. • System performance can be monitored. Alarms can be generated for exceptions. • The status of SLA fulfillment for slices can be monitored.	SIM cards, slice purchase orders, and service bills can be checked.
Management	• Slicing services can be activated, disabled, and restarted. • Network resources can be expanded and reduced to adapt to SLA requirements in real time.	• Signal transmission for slice subscription, updates, provisioning, and cancellation can be managed. • SIM cards can be managed; services can be activated, suspended, and resumed for users; user groups can be customized; and slices can be assigned to or withdrawn from users. • The binding between terminals and SIM cards can be managed.
Control	N/A	N/A
Capability exposure	Northbound APIs could be exposed, but most industrial enterprises call for a B2B self-service portal from operators.	

Healthcare slices have unique requirements that need to be satisfied for self-management:

- Management
  - √ Slice subscription, provisioning, and management

    Hospitals must be able to select and subscribe to slice offerings based on their services, such as remote teaching and consultation. The SLA requirements, maintenance services, and user groups (users and terminals) can be defined along with the subscription. Network policies, such as latency, bandwidth, number of users, traffic, and service start time and duration can be planned for each service, and the configurations can be completed before service provisioning.
  - √ Slice monitoring and adjustment

    Tenants must be able to promptly monitor the operating status and information of slices. Alarms can be initiated, helping tenants identify faults more easily. They can also check for resources and KPIs of slices for troubleshooting. Based on monitoring and troubleshooting results, tenants can activate, stop, or restart slicing services, and adjust network resources, user groups, or SLA requirements.
  - √ Slice users and terminals

    Tenants need to flexibly and efficiently manage users and terminals in each slice. Different user groups can be defined for different services, for example, remote consultation, remote teaching, mobile ward rounds, and medical consumable management, and users and terminals can be allocated to groups based on the services they use. Tenants can associate different network policies with different user groups, and the policies for a group can be batch applied for each group member.
- Capacity exposure

  Hospitals relatively lack ICT experience, so operators do not need to open northbound APIs, but provide a tenant self-service portal instead.

3.6 Comparison

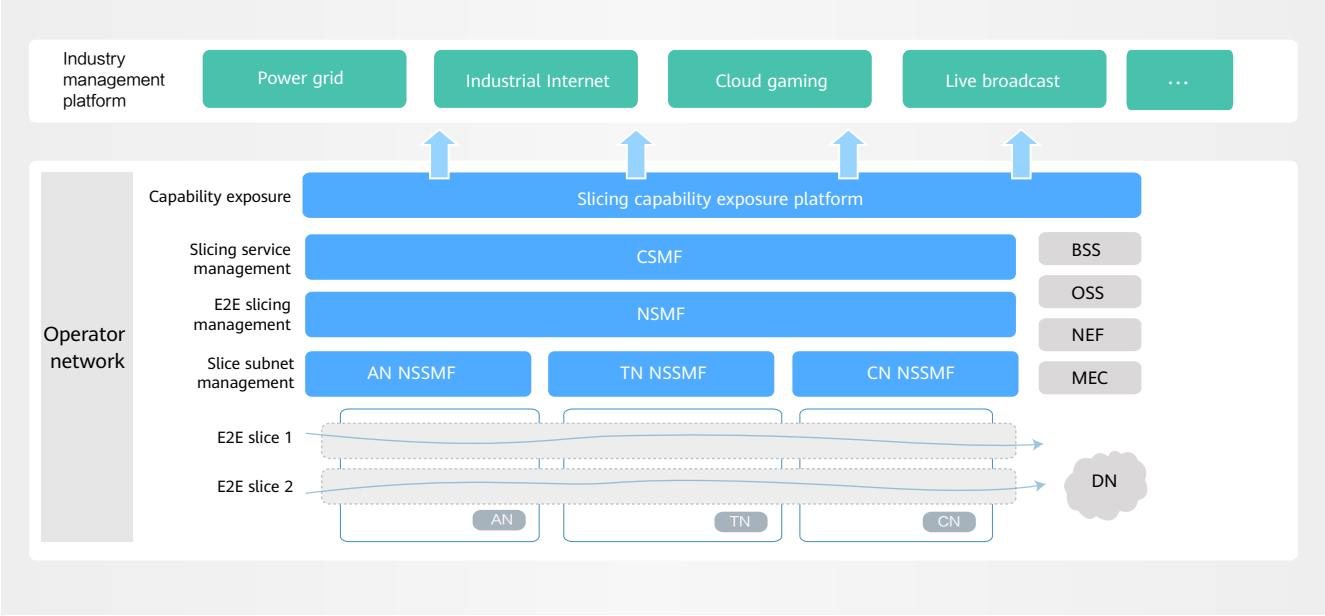
Maintenance/ Operations	Live Broadcast	Cloud Gaming	Power Grid	Industrial Internet	Healthcare
Visualization	◑	◑	●	○	◑
Management	◑	◑	●	◑	○
Control	○	◑	●	◑	○
Capability exposure	◑	◑	●	◑	○

The above table shows that fields possessing strong ICT capabilities, such as power grid, industrial applications, and cloud gaming, require differentiated SLA/QoS and strict isolation, and have higher self-management requirements.

04 5G B2B Network Slicing Management Architecture

We propose an architecture for the network slicing management plane, hoping it can serve as reference for the construction of your network. This architecture is based on the 3GPP-standard 5G network architecture and surveys regarding slice self-management requirements in key verticals.

4.1 Self-Management Architecture



The network slicing management plane consists of four logical layers. From top to bottom, they are:

- Capability exposure platform
  The platform provides both standard APIs and a self-management portal for industry players. It connects to the CSMF, NSMF, BSS, OSS, NEF, MEC, and other such nodes in the southbound direction, integrating and orchestrate their network capabilities.
- Communication Service Management Function (CSMF)
  The CSMF translates service requirements into specific requirements for network slicing, such as the number of users allowed, uplink and downlink transmission rates, latency, and jitter. It also manages service subscription and cancellation.
- Network Slice Management Function (NSMF)
  The NSMF manages and orchestrates network slice instances, and decomposes network slice requirements into network slice subnet requirements.
- Network Slice Subnet Management Function (NSSMF)
  The access network (AN), transport network (TN), and core network (CN) have their respective NSSMFs to manage and orchestrate their slice subnets. Each NSSMF decomposes slice subnet instances on its network into VNF instances and related connections.

In the later part of this paper, we only focus on capability exposure of network slicing, and do not cover the capability exposure from other sides, such as BSS, OSS, NEF, and MEC.

4.2 Key Capabilities

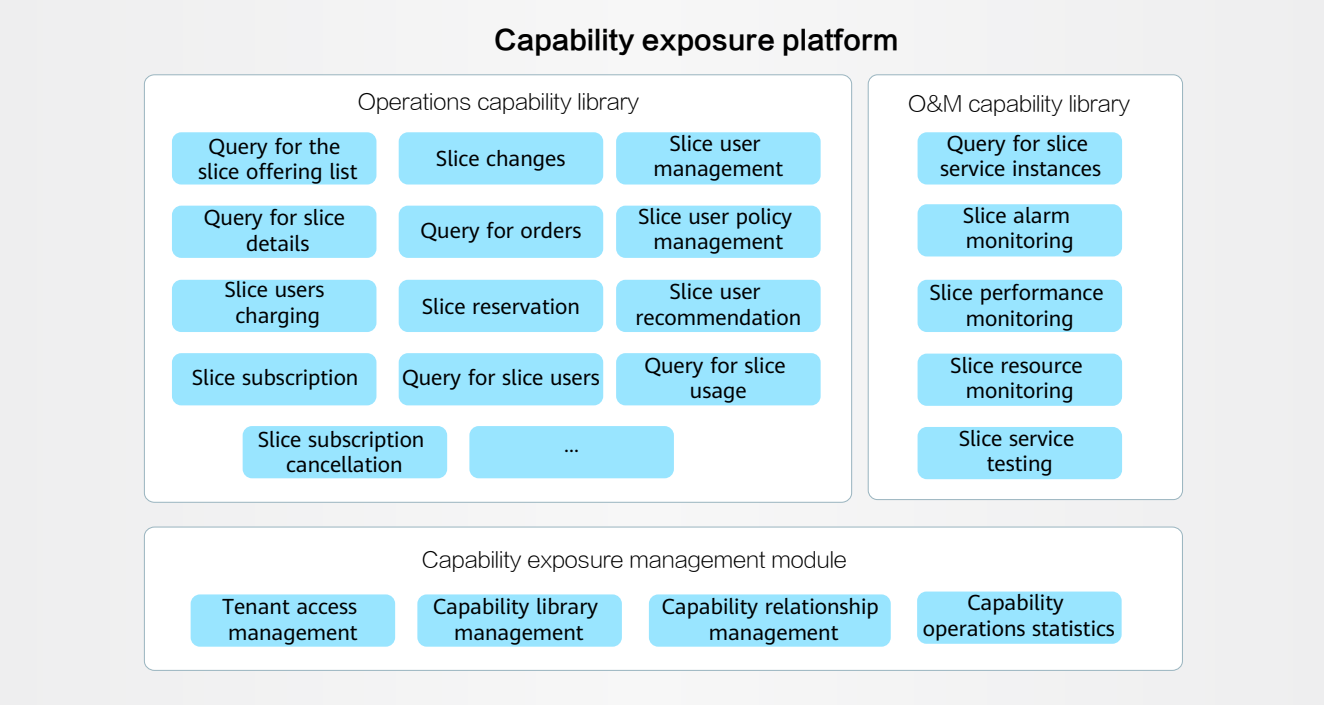


Figure 6 Capability exposure platform

The capability exposure platform consists of three parts:

- Operations capability library
  This library incorporates operations oriented to enterprise self-service, satisfying B2B2X business requirements. It enables tenants to query for and subscribe to slicing offerings and manage slice users.
- O&M capability library
  This library assists tenants in routine O&M. It offers means for tenants to monitor slicing services and SLA fulfillment, such as querying for slice instances, alarms, and KPIs.
- Capability exposure management module
  This module is used to provide access authentication; manage capability libraries, capability relationships, and capability invocation; and analyze operations statistics.

No.	Component	Capability	Remarks
1	Operations capability library	Query for the slice offering list	Enterprises can check the slice offering list and historical offering snapshots for all business scenarios.
2		Query for slice details	Enterprises can check detailed descriptions, prices, and SLA parameters of slice offerings.
3		Slice users charging	Enterprises can check charging types for slice users, and customize charging policies for B2B2C services.
4		Slice subscription	Enterprises can subscribe to slices online.
5		Slice subscription cancellation	Enterprises can unsubscribe from slices through self-service.
6		Slice changes	Enterprises can change slice parameters, capacity, and geographical scopes.
7		Query for orders	Enterprises can check the content and status of orders.
8		Slice reservation	Enterprises can reserve slices based on service plans.
9		Query for slice users	Enterprises can check the list of users who have subscribed to the slicing service and check service changes.
10		Slice user management	Enterprises can process the subscription and cancellation requests from users, and manage users via blacklisting and whitelisting.
11		Slice user policy management	Enterprises can adjust service policies for slice users.
12		Slice user recommendation	Slice user recommendations can be pushed to enterprises.
13		Query for slice usage	Enterprises can check slice usage (for example, traffic usage) of users in real time.
14	O&M capability library	Query for slice service instances	Enterprises can check information about service instances of provisioned slices.
15		Slice alarm monitoring	Enterprises can subscribe to service alarms to efficiently monitor services and locate faults.
16		Slice performance monitoring	Enterprises can subscribe to slice performance statistics.
17		Slice resource monitoring	Enterprises can check slice resource topologies.
18		Slice service testing	Enterprises can test services and analyze test results.
19	Capability exposure management module	Tenant access management	Access is protected by authentication, and enterprise service security can be managed.
20		Capability library management	CSMF and NSMF capabilities can be integrated and exposed via interfaces.
21		Capability relationship management	Capabilities can be orchestrated, and their relationships can be managed.
22		Capability operations statistics	Capability usage statistics can be collected and analyzed.

The 3GPP specifications define a logic module for exposing the capabilities of the CSMF. In practice, this module can be either co-located with or deployed independently from the CSMF.

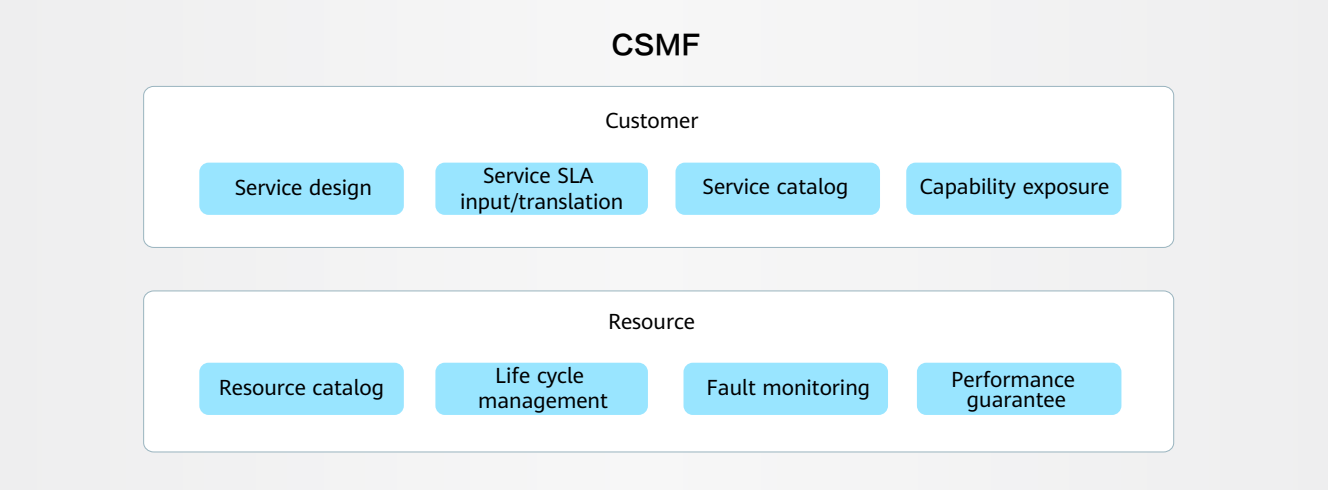


Figure 7 CSMF capability model

The NSMF orchestrates and guarantees services of slicing instances from end to end, and splits service requirements and passes them onto the NSSMFs on the AN, TN, and CN.

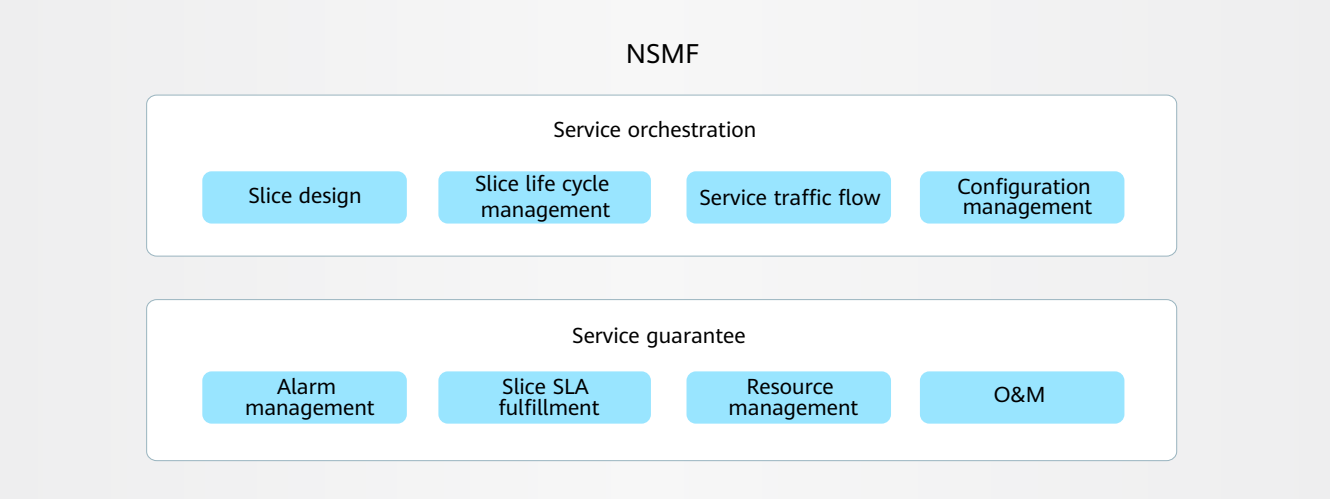


Figure 8 NSMF capability model

05 Suggestions for B2B Self-Management Development

5.1 Development Roadmap

Without any doubt, network slicing is the ideal choice to fast track B2B services, for it creates secure, reliable, manageable, and controllable private lines or networks for verticals. However, a network slice traverses a diverse range of entities within operator networks, including industrial terminals, wireless access devices, transport network entities, core network entities, NSMF, CSMF, billing systems, and more. In addition, verticals' requirements for slicing are consistently diversifying, though at different levels of urgency. This diversity ultimately leads to complex slice networking, not to mention end-to-end slicing management. The road to self-management is full of twists and turns.

We suggest a three phase roadmap for self-management services that considers:

- Industry requirements
- Complexity of network services
- Development progress of standards and specifications by 3GPP and the CCSA E2E slicing team
- Integration complexity, which increases from single-vendor, cross-domain, to multi-vendor plus cross-domain integration
- Progress of tests and certification by the IMT-2020 5G application promotion team

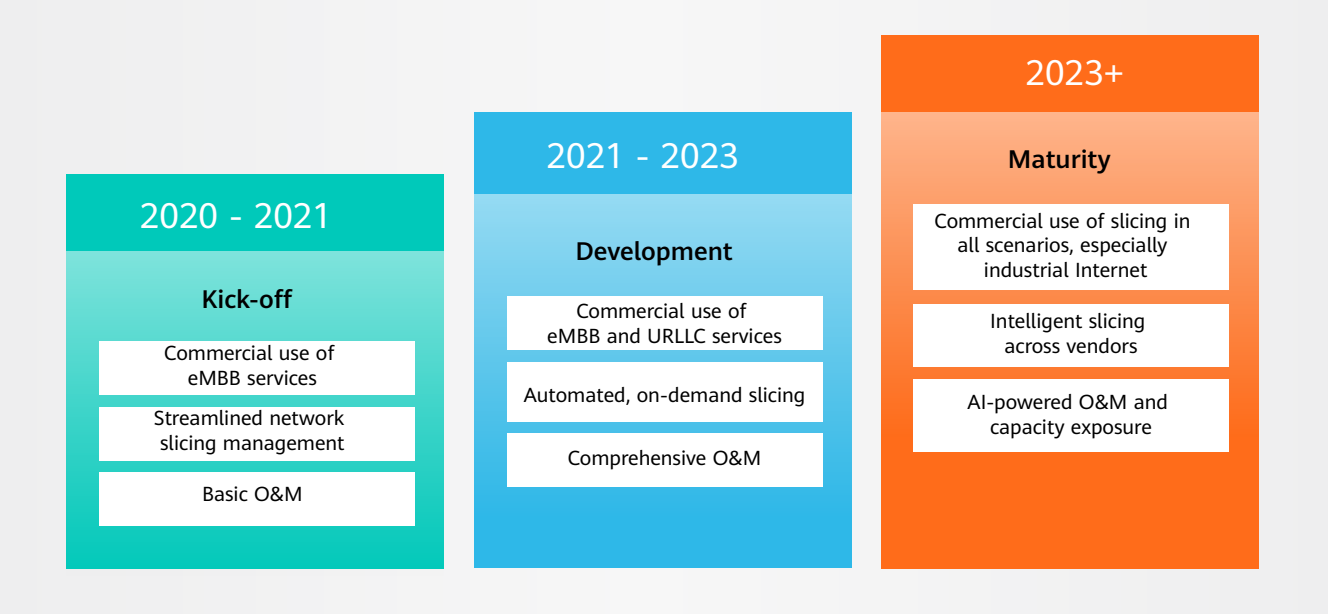


Figure 9 B2B self-management roadmap

Phase 1

- Target industries
  Industries that require high bandwidth, such as multimedia live broadcast
- Technologies
  Streamline the network slicing workflow based on the existing technologies and standards. Create eMBB slices, offering reduced latency.
- Management
  Provide basic capabilities for monitoring, automatic deployment, and capacity exposure, for example:
  √ KPI monitoring
  √ Online slice subscription
  √ Open interfaces for fetching KPI data

Phase 2

- Target industries
  Industries that require high bandwidth and low latency, such as electric power and energy
- Technologies
  Extend slicing across a vendor's domains as slicing standards advance. In addition to eMBB slices, develop URLLC slices.
- Management
  Provide comprehensive capabilities for monitoring, automatic deployment, and capacity exposure, for example:
  √ KPI monitoring from different aspects
  √ On-demand slice deployment and self-service for tenants
  √ Online slice subscription and open user management interfaces

Phase 3

- Target industries
  Industries with complicated requirements, such as manufacturing, and industries requiring cross-region operations
- Technologies
  The specifications and standards for interconnection interfaces will mature in this phase. Based on this, further extend slicing across domains and vendors, and adapt slicing to mMTC services.
- Management
  Leverage AI to provide intelligent O&M, and expose capabilities thoroughly, for example:
  √ Slice self-healing
  √ Integrate capabilities into tenant IT systems.

5.2 Suggestions

- Industry players need to get engaged in slicing development as early as possible to align initial network design with service requirements.

  Currently, there is no mature standard for an industry slice template. Though positioned as a general-purpose slice template, the GST defined by the GSMA does not allow for fine-grained parameter settings. The GST needs to be enriched and refined. We suggest that industry players join forces with operators and vendors to conduct surveys in different service scenarios under each industry, and define specific parameters for each service scenario to work out a more adaptive, scalable slice template for all industries.

- A well-rounded industry chain is required, for example, terminals must come with slicing support.

  Customized terminals are widely used with industry slices, while consumer terminals can only be used with a few slice-based services, such as cloud gaming. R&D of customized terminals for industry network slicing is still in its infancy, and 5G SA-enabled consumer terminals are inadequate. To ensure that network slicing services can be provided to industry users during the early commercial use of SA networks, we recommend that industry partners actively participate in the formulation of terminal specifications to promote the development of customized industry terminals.

- Operators need to improve network operations efficiency to accelerate B2B services.

  E2E SLA-oriented O&M mechanism.

  √ Currently, industry service requirements cannot be accurately mapped to specific network requirements, such as bandwidth, latency, and resources. In most cases, O&M personnel manually deploy the network and adjust services based on previous O&M experience, resulting in a network that cannot fully satisfy SLA needs. Whether a service can meet user requirements can only be determined via service tests or after the service is delivered. Apart from this, risks of faults which are irrelevant to network performance cannot be identified. To address these problems, we need to shift from conventional O&M to SLA-driven O&M, eliminating hidden risks caused by experience-based operations and fulfilling E2E SLA requirements of industry users.

  √ Intelligent closed-loop management

  Conventional network management is a static method for ensuring services stay operational. However, network slicing requires that O&M approaches be dynamically adjusted for different users, durations, and services, sometimes even for real-time service updates. Operators need to flexibly adapt slice architectures, topologies, and resources to cope with the varying service requests from users, maximizing resource utilization. In addition, they need to automate and apply AI to network operations, offering intelligent closed-loop management for all services.

- Industry players need to continuously improve self-O&M and self-adaptability and build an integrated self-management system.

  Once E2E network slicing management matures, industry players excelling at network operations will choose to integrate their operations system into that of their operator. Current industry operations systems mainly interact with the operator system via a capability exposure platform. However, the capabilities for exposure and the APIs involved have not been clearly defined yet. Industry players need to further their cooperation with operators to clarify the capabilities and APIs they require, and improve self-management. Additionally, coordination and connection between the BSS and OSS need to be enhanced to facilitate network slicing management.

06 Summary and Prospects

Network slicing is a novel technology coming about with 5G. It opens a broad B2B market for operators, who recognize the infinite potential in network slicing, and has dived into slicing technologies together with industry partners, vendors, and integrators. They have verified network slicing services as feasible and unlocked a series of new applications in different fields, including electric power, healthcare, live broadcast, industrial Internet, and cloud gaming.

To extract greater commercial value from network slicing, we need to strengthen our collaboration to explore new business and operations models for network slicing.

Since we do not have previous experience in all aspects of network slicing business, no matter the products, pricing policies, sales approaches, or delivery, it is rather difficult to reconstruct the present network structure to facilitate E2E network slicing management. As such, we need to standardize every aspect of slicing business models as soon as possible through joint effort, and polish these models through extensive practice. For example, network service providers may consider providing a self-service portal or designating API invocation counts as value-added services for industry customers.

Operators open their operations capabilities by interworking their slicing capability exposure platform with the management platforms of industry customers. With these opened capabilities, industry customers can reach the top level of Maslow's Model as described earlier in this paper, that is, self-management, or even self-maintenance and self-operations. A combination of the visualized, manageable, and controllable slicing services and conventional industry services will mobilize more and more business and operations models, ultimately reshaping society as a whole.

The development of 5G network slicing is just kicking off, and its commercialization will advance gradually. The speed of this will be affected by the construction progress of 5G SA networks, terminal capabilities, slicing management capabilities, and industries' understanding of network slicing. To ease this process, we need to unite under the 5GSA, to explore, define, and standardize the delivery and business models of network slicing, as well as plan and promote slicing application in various industries. Together, we can build a prosperous B2B market and lift industries to new heights.