Engineering 8 (2022) 42–59

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
6G Requirements, Vision, and Enabling Technologies—Review

The SOLIDS 6G Mobile Network Architecture: Driving Forces, Features, and Functional Topology

Guangyi Liu *, Na Li, Juan Deng, Yingying Wang, Junshuai Sun, Yuhong Huang

China Mobile Research Institute, Beijing 100053, China

ARTICLE INFO

Article history:
Received 4 January 2021
Revised 19 May 2021
Accepted 6 July 2021
Available online 30 September 2021

Keywords:
Sixth generation
Network features
Network architecture

ABSTRACT

With the large-scale commercial launch of fifth generation (5G) mobile network, the development of new services and applications catering to the year 2030, along with the deep convergence of information, communication, and data technologies (ICDT), and the lessons and experiences from 5G practice will drive the evolution of the next generation of mobile networks. This article surveys the history and driving forces of the evolution of the mobile network architecture and proposes a logical function architecture for sixth generation (6G) mobile network. The proposed 6G network architecture is termed SOLIDS (related to the following basic features: soft, on-demand fulfillment, lite, native intelligence, digital twin, and native security), which can support self-generation, self-healing, self-evolution, and self-immunity without human involvement and address the primary issues in the legacy 5G network (e.g., high cost, high power consumption, and highly complicated operation and maintenance), significantly well.

1. Introduction

The fifth generation (5G) mobile communication system has been defined by the third generation partnership project (3GPP) to fulfill the requirements of three typical application scenarios beyond 2020, that is, enhanced mobile broadband (eMBB), massive machine-type communication, and ultra-reliability low-latency communication (URLLC) [1]. It is believed that 5G will open a new era of the Internet of Everything and become an enabler of innovation and development in all walks of life.

The 5G system has been commercialized globally since 2019. By the end of 2020, 129 5G networks were launched globally. Early 5G systems aimed to satisfy the requirements of eMBB scenarios (e.g., cloud gaming, high-definition video, augmented reality, and virtual reality). In its later phase, when URLLC capability has been introduced, 5G will enable Industrial Internet and enterprise applications.

The commercialization of 5G will facilitate the application of artificial intelligence (AI), cloud computing, and big data technologies. It has not only profoundly changed the lifestyle of people, but has also accelerated the informatization and digitization of society, which will incorporate the concepts of "digital twin and ubiquitous intelligence" [2,3]. A new "digital twin" world will be built in which each physical entity has a virtual replica, as shown in Fig. 1. A "bridge" will be established between the physical and digital worlds, allowing seamless transmission of information and intelligence, between human and human, human and things, and even things and things. The digital world is the simulation and prediction of a physical entity and accurately reflects and predicts the real state of the physical world. It is also utilized to prevent the risks and disasters of the physical world by intervening in advance with proper operations. It helps to further liberate humans, improve their quality of life, boost production efficiency and social governance, and achieve the vision of "building a new world by digitalization and connecting all things by intelligence."

The digital twin world will cultivate several new application scenarios of mobile networks, such as synaesthesia internet, holographic interaction, digital human, intelligent interaction, super transportation, and precise medical treatment [2,3]. All these scenarios demand higher network capabilities, such as higher data rate, lower latency, preciser positioning, and deterministic quality of service (QoS), as shown in Fig. 2, which will push the 5G mobile network to evolve to the next generation, which is termed as sixth generation (6G) mobile communication system.

A network architecture is the key to providing a framework for assembling different enabler technologies to support targeted services and applications. Therefore, the network architecture is the corner stone for the 6G mobile system. This article discusses the

* Corresponding author.
E-mail address: 13701173393@139.com (G. Liu).

https://doi.org/10.1016/j.eng.2021.07.013
2095-8099/

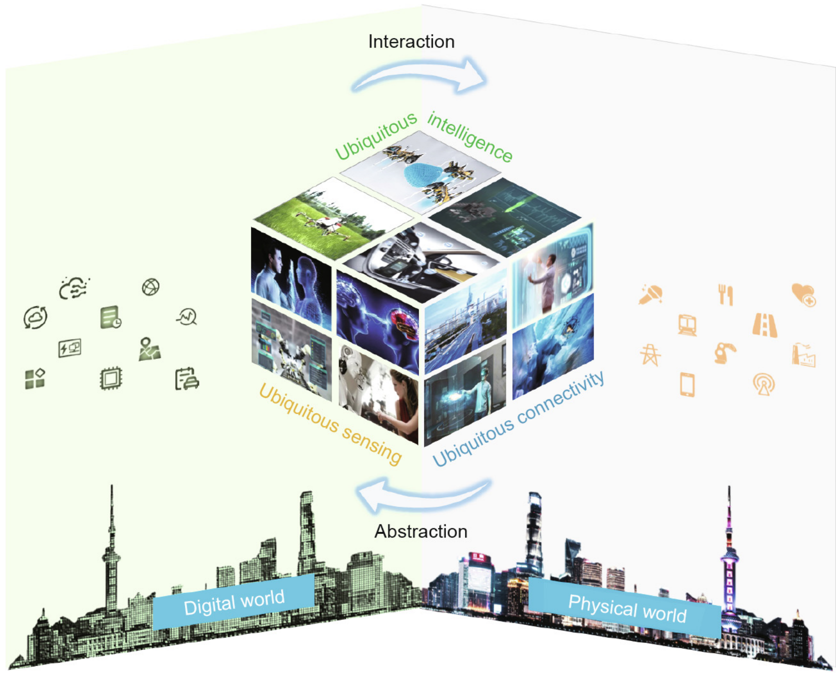

Fig. 1. Digital twin world.

designing of a logical network architecture that can be applied to a 6G mobile system and the potential features of the 6G mobile network. Before the architectural design, an analysis of the driving forces of network evolution is essential.

The first driving force is the demand for new applications and scenarios [4,5] toward 2030. As mentioned above, the commercial launch of 5G has stimulated the imagination and expectations of people for the next generation of mobile networks. On the one hand, new services/applications and use cases will be cultivated to demand higher network performance, such as higher data rate, lower latency, higher reliability, and higher resolution of positioning, which are beyond the capabilities of 5G systems. On the other hand, new network capabilities must be offered, such as native intelligence and security.

In addition, the deep convergence of information, communication, and data technologies (ICDT) [6,7] sheds light on the higher efficiency of the network architecture [8,9]. This is the second driving force of network evolution. Considering that resources such as computing and storage will be extended from the center to the edge, it is necessary to enable the network to be capable of endogenous computing as well as resource perception and control. Edge AI and distributed AI are becoming increasingly popular nowadays [10,11], which drives us to consider such AI deployment methods in network design to support real-time AI applications. Data governance, including data security and compliance, data analysis and application, and data security circulation technology have become trends, which should also be considered in the network design.

The third driving force of network evolution is a consideration of the problems and challenges faced by 5G networks. When designing the 6G mobile network architecture, the mature technologies and valuable philosophy of 5G mobile networks should be inherited, and the lessons from system design, commercial deployment, and operation of the 5G networks should be implemented satisfactorily in the top-level design of the 6G system. Moreover, the problems and challenges faced by the 5G networks should also be addressed to ensure successful and continuous 6G development, such as the rapid growth of capital expenditure (CAPEX) and operational expenditure (OPEX), high power consumption, and the difficulty of operation and maintenance (O&M).

The rest of this article is organized as follows. Section 2 recalls the evolution history of the architecture of mobile networks. Section 3 summarizes the driving forces of innovation in the 6G mobile network architecture. In Section 4, a logical network architecture termed SOLIDS with three layers and four planes is proposed. The word "SOLIDS" is formed from the first letter of each 6G network feature, that is, soft, on-demand fulfillment, lite, native intelligence, digital twin, and native security. These network features are elaborated in Section 5. Finally, Section 6 concludes the paper.

2. History of mobile network architecture evolution

2.1. Evolution of network architecture

Mobile network architecture is constantly changing with the development of mobile communication technologies. The third generation (3G) mobile network adopted a three-layer architecture, as shown in Fig. 3(a) [12]. In the fourth generation (4G) mobile network era, a flat and all-internet protocol architecture was adopted, which was simplified into two layers [13], base stations (termed as eNodeB) and evolved packet core (EPC) network. The major elements of EPC include mobility management entity (MME), home subscriber server (HSS), serving gateway (S-GW), and packet data network gateway (P-GW), as shown in Fig. 3(b). Based on this flat network architecture and the separation of the control plane (CP) and user plane (UP), the latency of the end-to-end data transmission was shortened significantly. In the 5G era, the network architecture introduces information technology (IT) to reform the core network (CN) architecture and leads to a service-based architecture (SBA) [14], with the further separation of the CP and UP, and the incorporation of network slicing, as shown in Fig. 3(c).

To address the diverse service demands from the typical application scenarios beyond 2020, 5G networks have attempted to introduce network slicing [15] based on software-defined network (SDN) [16] and network function virtualization (NFV) [17], which provide a dedicated logical or virtual network with the proper

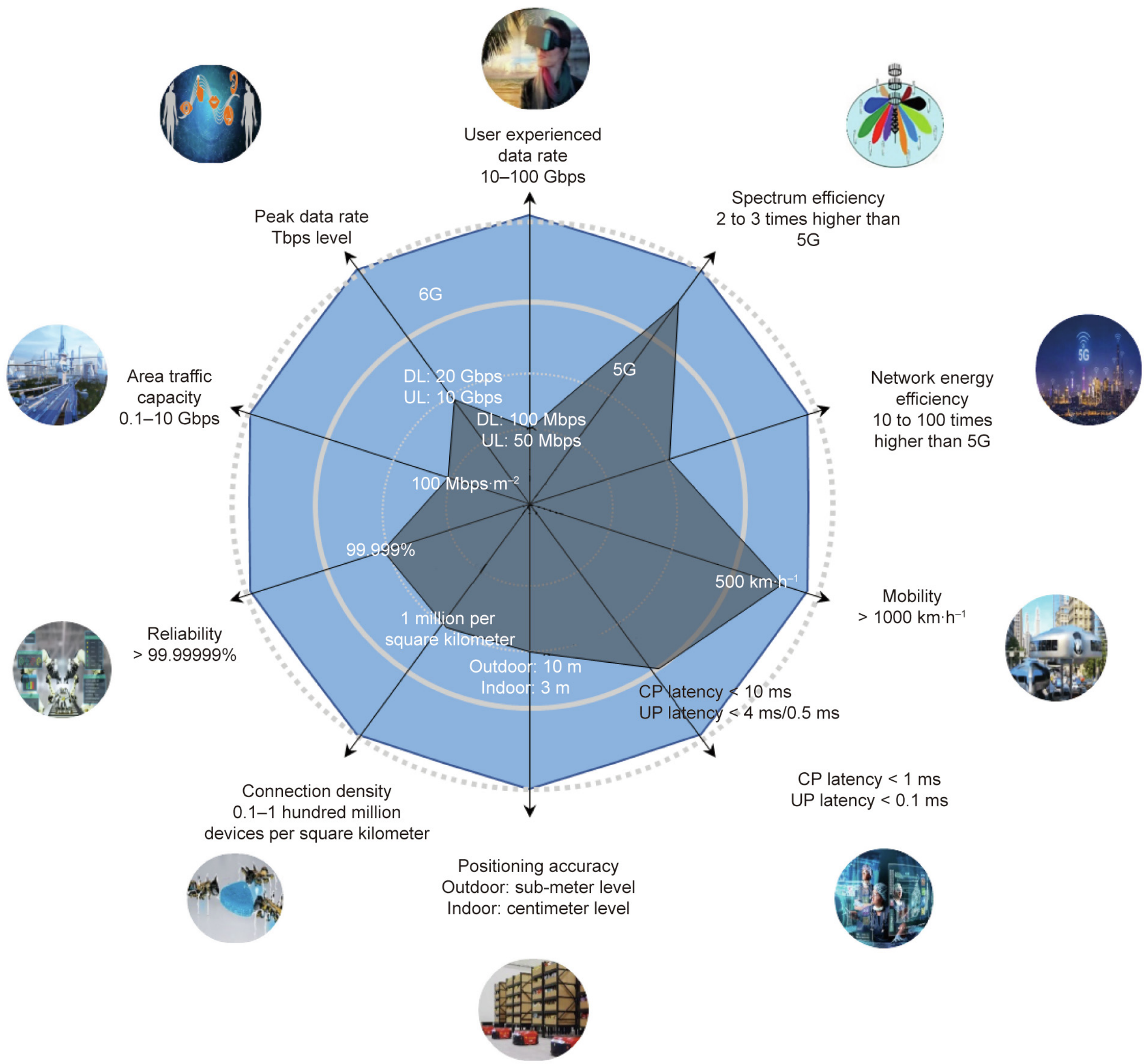

Fig. 2. Comparison of key performance indicators (KPIs) of 5G and 6G systems. DL: downlink; UL: uplink; CP: control plane; UP: user plane; Gbps: gigabits per second; Tbps: terabits per second; Mbps: megabits per second.

function and processing capability to fulfill targeted service requirements such as QoS, privacy, and security. Network slicing is expected to provide elastic capacity expansion, fast function upgrade, and on-demand deployment of network functions for 5G networks. Network slicing is desired to support the diversified service demands from enterprise and industrial application scenarios using a single network.

The separation of the central unit (CU) and distributed unit (DU) was also introduced in the radio access network (RAN) to support flexible deployment. On the one hand, a CU can be deployed centrally to share the cost, such as on a common platform with multi-access edge computing and/or UP function. On the other hand, to improve the intelligence and scalability of wireless networks, a DU can be deployed as close as possible to users to satisfy URLLC service requirements.

2.2. Continuous evolution of 5G systems

The early exploration of the Industrial Internet and enterprise application scenarios proves that the 5G CN appears to be significantly complicated when compared to the network of an over-the-top operator, and its adaptability to diverse service requirements must be further improved. The 5G CN has introduced a heavy burden on the testing and validation of functions or network elements owing to the openness of interfaces between them and the decoupling of software and hardware. The integration and maintenance of the CN has become a challenge for the traditional mobile operators.

Conversely, the major work of network slicing is primarily done by the CN, and considerable effort is still required for the RAN and transport network (TN) to support end-to-end network slicing. To automatically support the management and orchestration of network slicing, further standardization of network slicing is still ongoing in the 3GPP [18].

Moreover, 5G mobile operators are facing the challenges of high cost and high power consumption of base stations, complicated and difficult O&M due to the large-scale deployment and interoperation between 5G and legacy networks (e.g., 4G and 3G).

To address the high power consumption challenge of 5G networks, intelligent power saving was introduced to power off the base station by means of carrier switch off, radio front switch off, timeslot-level switch off, and symbol level switch off. Further, the load balance between 5G and 4G systems is designed to switch off the 5G base station when its load is light.

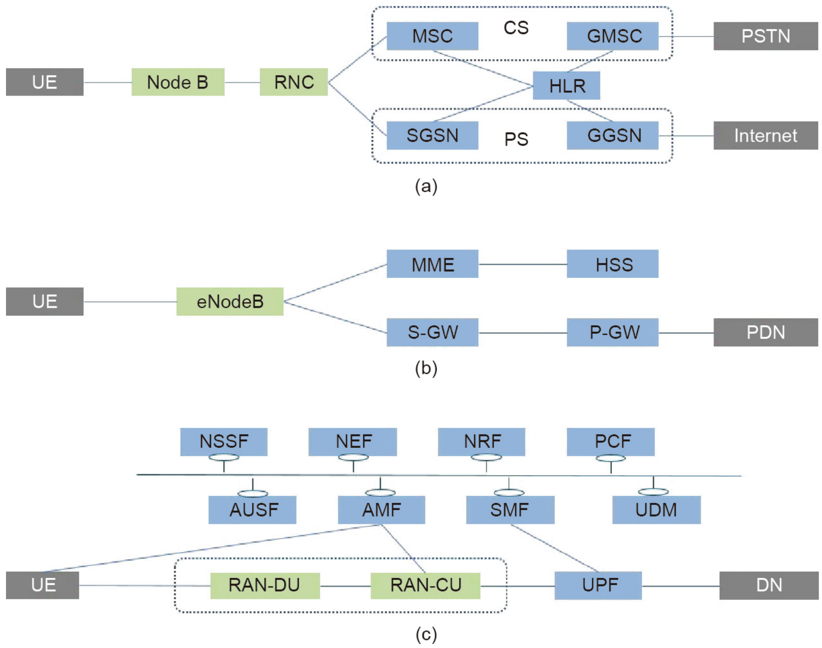

Fig. 3. Network architecture evolution. (a) 3G network architecture [12]; (b) 4G network architecture [13]; (c) 5G network architecture [14]. UE: user equipment; RNC: radio network controller; MSC: mobile switch center; GMSC: gateway MSC; HLR: home location register; SGSN: serving general packet radio service (GPRS) support node; GGSN: gateway GPRS support node; CS: circuit switched; PS: packet switched; PSTN: public switched telephone network; eNodeB: evolve Node B; DN: data network; PDN: packet DN; NSSF: network slice selection function; NEF: network exposure function; NRF: network repository function; PCF: policy control function; AUSF: authentication server function; AMF: access and mobility management function; SMF: session management function; UDM: unified data management; RAN: radio access network; DU: distributed unit; CU: central unit; UPF: user plane function.

To address the low efficiency of O&M, AI is considered in 3GPP and other standardization organizations [19–21]. In 3GPP service and system aspects working group 5, in addition to extending work on self-organizing networks from 4G to 5G, work items related to autonomous networks have been set up and are attracting attention. Among these work items, "autonomous network level" [19] defines network autonomy as the capability of the telecom system to be governed by itself with minimal or no human intervention and describes six levels of application of autonomy capabilities in the network management workflow. The "closed loop communication service assurance" specifies the concept of open and closed control loops, as well as use cases, requirements, and a model for closed-loop communication service assurance. In the "intent driven management for mobile networks" [20], 3GPP has studied concepts, scenarios, and solutions that are able to simplify the management interfaces. "Management data analysis service" [21] brings intelligence and automation to the management system and generates value by processing and analyzing network data, where AI and machine learning (ML) techniques may be utilized. Several mobile operators and manufacturers also work together to study autonomous networks, which target "zero touch" of O&M [22]. In Ref. [23], 39 scenarios were used to identify business-oriented and automation-related challenges faced by operators and vertical industries, the analysis of which derives architectural, functional, and operational requirements. The framework of an autonomous network has also been defined in Ref. [23].

However, owing to the limitations of the current network architecture, the application of AI in O&M can only be implemented case by case and the modifications are added to the network in a "patched" manner. The network efficiency achieved by AI is far behind expectations. A native intelligence-based network architecture is required for further efficiency improvement.

To address the high costs involved in 5G network deployment, several mobile operators and manufacturers work together to advance the open RAN (OpenRAN) project [24], which aims to open the interface between different modules of the base station, and build the base station architecture on the basis of white-box hardware and open-source software.

As a pioneer of OpenRAN, Rakuten Mobile, the fourth largest mobile network provider in Japan, aims to redefine the expectations in the mobile communication industry to provide appealing and convenient services that respond to diverse customer requirements. In February 2019, Rakuten announced the first real-world, commercially friendly, end-to-end trial in a fully virtualized cloud-native mobile network [25]. Rakuten publicly claimed the success of its initiative by demonstrating stable and scalable 4G and 5G services. Rakuten claimed that its strategy yielded a 35%–40% reduction in CAPEX and a 30% reduction in OPEX. For future 5G evolution with the cloud-native architecture, the savings are projected to be as high as 50%.

As the next step in mobile network architecture evolution, 6G network innovation will inherit the experience from the past journey, digest the latest progress on the convergence of ICDT, and achieve a network architecture that supports self-generation, self-healing, self-evolution, and self-immunity.

3. Driving forces of mobile network architecture innovation

The driving force of the 6G network architecture is primarily based on three aspects [26]. The first one is the new requirements brought by the new services and application scenarios, which include higher user experienced data rates, lower CP and UP latencies, higher peak data rates, higher connection density, and the ubiquitous coverage of space, air, ground, and sea. The second is the technology trend of the convergence of ICDT. With the significant progress in cloud computing, big data, and AI technologies, more efficient software and hardware solutions will promote the development of the 6G network architecture in a more efficient and low-cost direction. More specifically, the design of the 6G network must consider new network architectures, stronger network functions, and the use of a general-purpose processor platform based on commercial off-the-shelf components instead of dedicated hardware. Third, the problems and challenges faced by the legacy 4G and 5G networks must be resolved in the 6G network architecture. For example, when compared to the 4G network, both the cost and energy consumption of 5G networks have tripled and the O&M of these networks has become significantly more complicated, which require new thinking and should be solved in the future design of the 6G network.

3.1. New use cases and new scenarios beyond 2030

The rapid deployment and application of 5G will certainly facilitate the application and development of AI, cloud computing, and big data, accelerate the digitization of the entire society, significantly improve the operation, production, and life efficiencies, significantly improve the quality of life, and push the entire society toward a digital world. Every object in the physical world will probably have a digital replica in the digital world, and these digital replicas together constitute a digital world. The physical and digital worlds constitute the digital twin world [2,3]. The digital twin world can predict changes in the physical world in advance and can also intervene in advance to avoid the occurrence of accidents and natural disasters in the physical world. In the digital twin world, several new application scenarios that are closely related to mobile communication will emerge, including synesthesia interconnection, holographic interaction, human digital twin, intelligent interaction, and intelligent industry and agriculture, as shown in Fig. 4 [4].

The comprehensive features of the service and application beyond 2030 can be summarized as follows: ① The service and application requirements will be more diverse than ever; ② three-dimensional coverage of terrestrial, air, and sea is demanded with consistent service continuity; ③ diversified interaction forms and contents enable more immersive user experience; ④ customized and personalized services require on-demand end-to-end network orchestration and configuration; ⑤ integration of communication, computing, and sensing enriches the service types and business cases; ⑥ security has become more important than ever, and the services demand more efficient security support instead of a patched manner of support.

To fulfill these requirements, the 6G network must support on-demand fulfillment and guarantee user-centric experience by customizing the orchestration and parameter configuration of the end-to-end network functions and resources.

3.2. Trend of deep convergence in ICDT

The rapid development of IT has accelerated the popularity of the internet, and various applications have emerged consecutively. With the emergence and rapid development of cloud computing, this process has been accelerated. Large cloud computing companies can quickly deploy large-scale IT service capabilities for computing and storage by using cheap commercial off-the-shelf hardware. Enterprises or individuals can rent the IT service capabilities of cloud computing companies according to service requirements, store their own data in the cloud data center, and call them as required. The required computing power can achieve rapid internet service deployment and application.

Driven by the continuous development of technologies, communication technology is also developing and iterating at a significant speed. The popularity of 4G networks and smartphones has brought unprecedented prosperity to mobile internet services and profoundly changed the daily lives of people. Smart phones have become an important platform for people in their daily lives and are used for various requirements, such as for traveling, shopping, and entertainment, which produce huge volumes of data, including location, trajectory, personal preference, entertainment, and shopping habits. Through the collection and analysis of these user behavior data, internet service providers can obtain user portraits and achieve personalized service provision to a large extent, including accurate content push and convenient service acquisition, which promotes the rapid development of big data application and processing technology. With the rapid rise of Internet of Things (IoT), including the global system for mobile communications, narrow-band IoT [27], and enhanced machine-type communication [28], the objects of communication have been extended from people to things, and the number of connected things has exceeded the number of people. For example, there are approximately 950 million human users in China Mobile, while the number of IoT users is approximately 2 billion. A large number of connections results in huge amounts of data, and the analysis and application of these massive amounts of data promote the

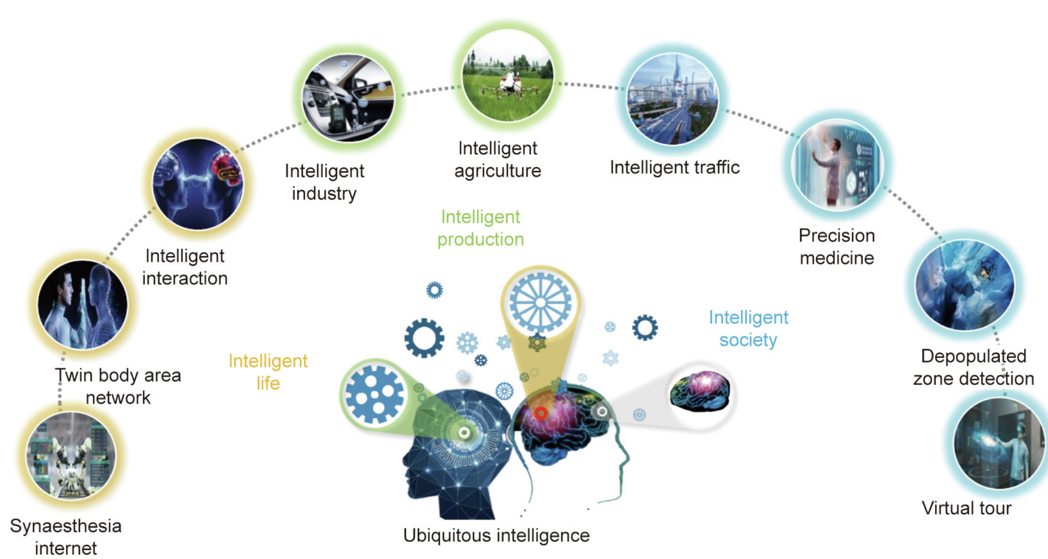

Fig. 4. Typical application scenarios beyond 2030 [4].

informatization and digitization of society. In particular, with the deployment and application of 5G networks, the radio transmission rate reaches gigabits per second (Gbps) or even 10 Gbps, and the radio transmission latency can be shortened to the millisecond level, and the reliability of data transmission is also improved from 99.999% to 99.99999%. All of these will bring the application of mobile communication technology to every corner of society and achieve the interconnection of all things. The digital transformation will bring prosperity to all walks of life, accelerate the digitization of the entire society, and achieve the goal of "5G changing society."

Big data has become an important strategic resource for enterprises and society and is the new focus of researchers and the industry. With the popularization and application of 4G and 5G systems, we can note that the entire mobile communication network and its applications are always generating massive amounts of data. These data will contain a significant wealth of information of the entire society, and the rapid development of data technology (DT) will enable the application of these data to human life and social governance, such as smart shopping, smart transportation, smart medical care, smart campus, and smart city.

Big data is inseparable from cloud processing [29,30], which is a platform for generating big data. Since 2013, big data has been closely integrated with cloud computing, and it is expected that the relationship between them will be closer in the future. In addition, emerging computing forms such as the IoT and mobile internet will promote a big data revolution, allowing big data marketing to exert greater influence.

Big data is likely to be a new round of technological revolution, similar to computers and the internet. The subsequent emerging technologies such as data mining, ML, and AI may change many algorithms and basic theories in the data world and achieve breakthroughs in science and technology.

With the acceleration of digitalization, every society element will generate a large amount of data, which are obtained from individuals, companies, infrastructure, and so on. Because the ownership of these data is considerably different, problems such as data storage, data management, data sharing, data security and privacy, and data transaction become significantly difficult to handle. Therefore, relevant big data legislation is bound to appear in the future to clarify the ownership of data and the distribution of corresponding benefits. At the same time, big data storage and management platforms will appear to help everyone achieve data storage, management, and transactions.

The large-scale deployment of big data has promoted the development of AI applications. AI is a new technological science that studies and develops theories, methods, technologies, and application systems for simulating, extending, and expanding human intelligence. AI is a branch of computer science. It attempts to understand the essence of intelligence and produces a new intelligent machine that can respond in a manner similar to human intelligence. Since the birth of AI, the related theory and technology have become increasingly mature, and the application field has continued to expand. It is conceivable that the technological products brought by AI in the future will be the "containers" of human wisdom. AI can simulate the information processes of human consciousness and thinking. The research fields of AI primarily include knowledge representation, automatic reasoning and search methods, ML, knowledge acquisition, knowledge processing systems, natural language understanding, computer vision, intelligent robots, and automatic programming. The applications of AI primarily include machine translation, intelligent control, expert systems, robotics, language, image understanding (such as face recognition and license plate recognition), genetic-programming-based robot factories, automatic programming, and aerospace applications. They profoundly affect our daily lives and work.

In the design of 5G networks, the trend of ICDT convergence has emerged. The design of the 5G CN fully introduces advanced IT concepts and achieves network slicing through SDN/NFV and SBA, which provide important support for 5G networks to empower vertical industry applications. Conversely, 5G mobile communication networks usually consist of millions of base stations, routers, CN elements, and other infrastructure devices as well as billions of users. Massive amounts of data, including the operating data of each network element, signaling data generated during communication, event reports, and related information about users moving in the network, are produced. If tags such as time and location are added to these data, it will bring immeasurable value to the automation and intelligence of network O&M. Therefore, based on user location information in the network, operators have begun to study network automation based on big data and AI, such as massive multiple-input multiple-output (MIMO) weight optimization, network anomaly analysis, and user experience analysis and optimization [31]. At the same time, 3GPP has begun to study big data collection in radio networks [32], the automation and intelligence of network O&M [33], and the application of AI in radio resource scheduling [20]. Therefore, it can be observed that in the later stage of the development of 5G standards, the application of DT will be further integrated with communication technology. ICDT integration is becoming a new trend of development, which will further reduce network O&M costs and improve network service capabilities and user experience in 5G networks.

However, we believe that the existing 5G systems cannot perfectly match the deep convergence of ICDT. One example is the network data analytics function (NWDAF). NWDAF is a data analysis network element that automatically perceives and analyzes the network based on the network data, and participates in the entire life cycle of network planning, construction, O&M, and network optimization and operation, resulting in a network that is easy to maintain and control, improving the efficiency of network resources, and improving the user experience. Although the network performance is efficiently improved with NWDAF, the patched integration of AI and the 5G system exposes certain problems. The first is the data security issue and the heavy signaling overhead caused by the reporting of mass measurements. The other is the challenge of low latency because all the data have to be uploaded to and processed on the central analysis unit, such as NWDAF, which may be deployed far away from the data source. Such issues should be considered when designing the 6G system.

For this reason, we believe that the in-depth convergence of ICDT will surely become an important driving force for 6G network design, and cloud native, big data, and AI will perform significantly important roles in future network architecture design.

3.3. Problems and challenges faced by 5G network

Since 2019, 5G networks have been deployed globally on a large scale. These 5G networks, together with cloud computing, big data, and AI, will definitely cultivate several new services and applications, and thus promote digitization of the entire society. With the development of 5G networks and the continuous emergence of new services and applications, 5G networks will inevitably face several new problems and challenges. Some of them may be solved in the evolution of 5G networks; however, it may be difficult to solve certain problems owing to the limitations of the 5G network itself. These problems and challenges will become an important motivation and source of innovation for the 6G network design.

From the perspective of 5G network characteristics and recent developments, 5G networks will face challenges in the following aspects.

3.3.1. The hierarchical protocol stack

In 5G networks, the protocol of the air interface is a hierarchical structure, including a physical layer, medium access control (MAC) layer, radio link control (RLC) layer, and packet data convergence protocol (PDCP) layer. All service data must be processed through all these layers. The process of each layer introduces a specific latency, which leads to a latency bottleneck. For example, the typical latency for an eMBB packet to pass through the air interface is 3 ms. During the study phase of 5G networks, certain UP modifications were proposed to reduce the processing latency, such as the following: ① allowing packet deciphering before reordering at the PDCP layer to reduce the amount of packet processing and latency, which simplifies the implementation; ② removing the packet concatenation function at the RLC layer to enable more offline header computation by decoupling automatic repeat request and concatenation/segmentation; ③ allowing the MAC subheader next to the MAC payload at the MAC layer to address the potential requirement of approximately one symbol time from the uplink grant to uplink transmission time [34–37].

All these modifications aim to reduce the latency by optimizing the processing sequence. Although some of these changes are gainful and have been endorsed in the 5G specification, these changes are still limited to existing layering standards. If the latency of the air interface needs to be further reduced, a straightforward method is to break the traditional layering structure and create a shortcut in the data processing pipeline.

3.3.2. The continuous evolving technologies

The 5G network provides services for vertical applications, which brings diversified requirements for network capability and deployment. To address these requirements, 5G CNs have defined an SBA and network slicing based on SDN and NFV. The 5G system is expected to support end-to-end slicing; however, at the beginning of the standard design, the design and optimization of slicing are considered primarily in the CN and TN. There was no dedicated design for slicing in the RAN in the early releases of 5G networks, and further considerations were contemplated in Release 17. Conversely, the standardization of network slicing involves six major industry organizations, with their respective divisions of work [38]. The slow progress in inter-organization collaboration restricts the commercialization of end-to-end network slicing. Although the early 5G specifications released by 3GPP support functions related to eMBB slicing, the slice management specification is not yet perfect. The International Telecommunication Union and Internet Engineering Task Force have defined the specifications of TN slicing in relation to aspects such as slice identification and management interface but have not defined a coordination mechanism with CN and RAN. The zero-touch network and service management of the European Telecommunications Standards Institute has developed an end-to-end slice management implementation framework and solutions, which are still in the early stages of development. The Telecommunication Management Forum (TMF) is studying the integration of the 3GPP slice management architecture with the existing TMF architecture but has only released the first draft. The Global System for Mobile Communications Association defines the application scenarios of network slicing and common slicing templates and has released slicing parameters and technical requirements for eight scenarios.

Commercial CN equipment that supports network slicing is relatively mature. Owing to the technical difficulty of RAN slicing, there are differences in the implementation by different equipment manufacturers. In general, there are still several technical challenges in the implementation of end-to-end network slicing, which requires the coordination and connection of slices between subdomains, and the complexity is quite high.

To completely adopt the value of network slicing in the future, it is necessary to arrange and manage network slices reasonably in the network slice management domain. Fig. 5 is a schematic diagram of the network slice management domain, which shows the relationship between the slice management domain and the slice network domain. In a commercial network, in several cases, these network slice management functions (NSMFs) do not exist independently, but are embedded into the operation support system (OSS) and business support system (BSS) of operators [39]. Different operators have different specific schemes for network slice management, which require the O&M, IT, government and enterprise departments, and network departments of the operators to participate in the integration of the slice management domain with the BSS/OSS. This will be a challenge for operators, which should be fully considered in future network designs.

3.3.3. The rigid network architecture

3.3.3.1. Deployment analysis of 5G network. For a mobile network, the traffic load at different base station sites is significantly diverse

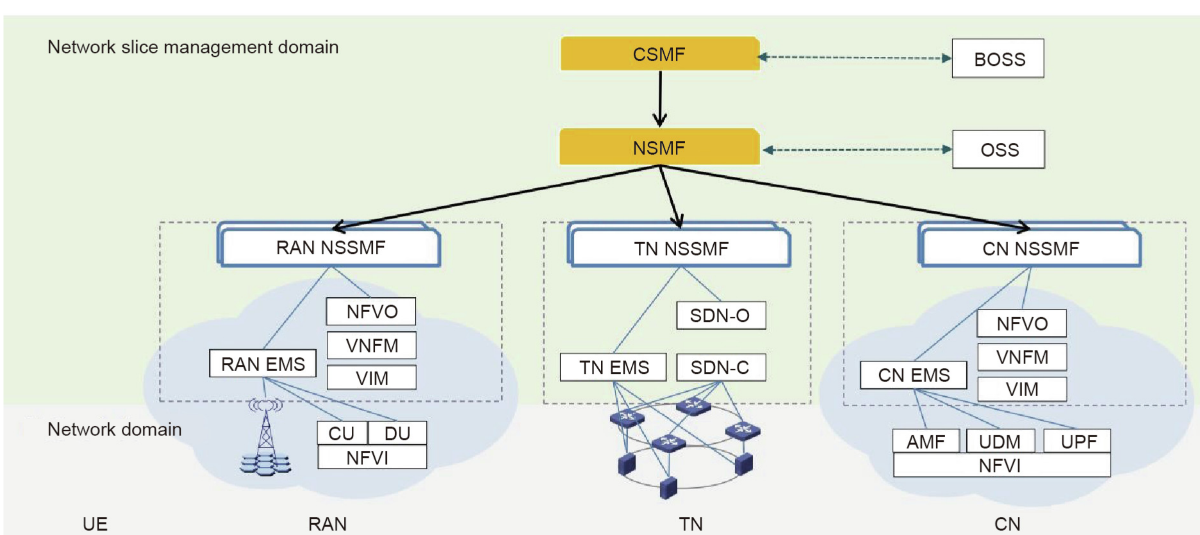

Fig. 5. Diagram of the network slice management. CSMF: communication service management function; NSMF: network slice management function; OSS: operation support system; NFVO: NFV orchestrator; NSSMF: network slice subnet management function; BOSS: business operation support system; EMS: element management system; VNFM: virtual network function manager; VIM: virtualized infrastructure manager; SDN-O: SDN-orchestrator; SDN-C: SDN-controller; NFVI: network functions virtualization infrastructure.

owing to the mobility of the users. Typically, in China, 20% of the sites are heavily loaded, while 80% of the sites have light loads. However, each site is usually configured with full capacity, which leads to a wastage in the hardware processing capability and power consumption. Owing to the larger bandwidth and massive MIMO configuration of each base station, the cost and power consumption of a 5G new radio (NR) base station is approximately three times that of a 4G base station, which further worsens the situation.

A straightforward way to reduce the cost of network deployment is to dynamically configure and share the hardware and processing capabilities among different sites. The 5G systems have attempted to share hardware and processing capabilities. To support the flexible deployment of 5G networks, a CU and DU split architecture for a base station was introduced. Several possible functional split options between CUs and DUs have been proposed and discussed, as shown in Fig. 6 [40]. During the discussion, both architectural and specification aspects were considered, including the number of split options to be specified and supported by an open interface, implications of long-term evolution (LTE)/NR tight interworking, granularity of the functional split, and reconfiguration dynamicity of the functional split. It is undeniable that the choice of how to split NR functions in the architecture depends on certain factors related to radio network deployment scenarios, constraints, and intended supported services. However, after multidimensional comparative analysis, only a single higher layer split option, that is option 2 (PDCP/RLC split), was selected.

Based on the CU–DU split architecture, the CU can be deployed centrally and the processing capability can be shared among different sites. However, in the actual deployment, the CU is still combined with DU because the benefit of the centralized CUs is not obvious from the network performance perspective.

In contrast, to save the rent fee of the machine room for the CU and DU, mobile operators combine together the CUs–DUs from several sites, which laid the foundation for sharing the capacity and hardware processing capability among different sites, as shown in Fig. 7. However, owing to the dedicated hardware and software design of the CU and DU, the sharing of capacity and hardware processing capability among different sites is not allowed in the pool. A straightforward method to ensure the feasibility of sharing is to design the CU and DU using a cloud-native approach based on SDN and NFV, which enables the dynamic sharing of the physical hardware and offers capacity on demand for pooled sites. When the load of the pool is light, most of the hardware in the pool can be shut down to save power; when the load becomes heavy, the additional hardware is switched on to support the demand from the corresponding sites.

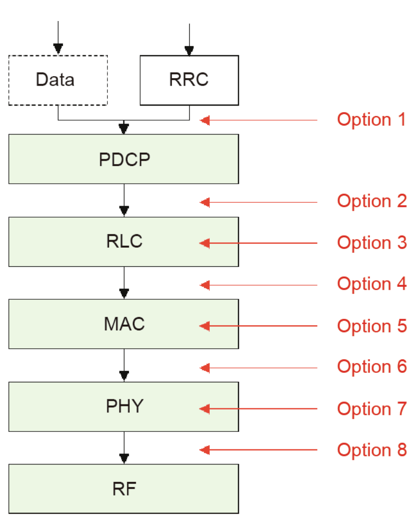

Fig. 6. Function split between DU and CU [40]. RRC: radio resource control; PHY: physical; RF: radio frequency.

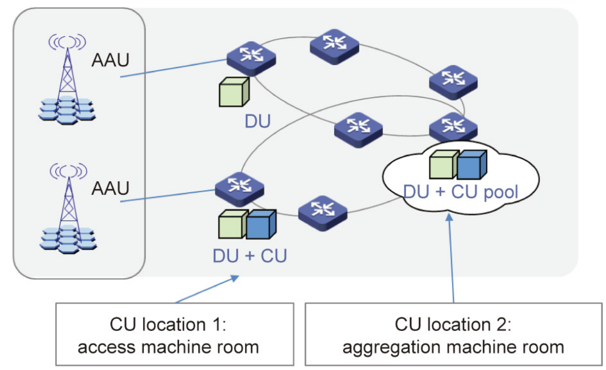

Fig. 7. CU deployment location diagram. AAU: active antenna unit.

3.3.3.2. Power consumption analysis of 5G networks. Power consumption is a tricky problem for 5G base stations and is the biggest burden for mobile operators. In addition to 5G CAPEX, power-intensive OPEX is the most important factor in determining the development of 5G systems. During the period of 2017–2020, the power consumption of base stations was significantly reduced under the concerted efforts of the industry; despite this, the power consumption of 5G base stations is still approximately three times that of 4G base stations.

Currently, the power consumption of a full-load 5G base station may even reach up to 3800 W, while the average power of a 4G base station is approximately 1000 W. Under a normal load, the radio frequency unit consumes most of the power of the 5G base station. The remaining part is consumed by the base station processing because the large bandwidth and high data rate require higher base station processing capacity.

The industry also proposed a variety of base station power-saving solutions. The current solution is to automatically generate power-saving strategies through an AI algorithm based on factors such as base station equipment type, coverage scenario, power-saving target, and turn-off duration. These strategies include cell turn-off, carrier turn-off, RF channel closing, and symbol turn-off. Each strategy has its specific application scenarios and degrees of impact on network quality.

All these efforts on power saving should be fully considered in the 6G network design.

3.3.4. Difficulty in O&M

As the methods of network management and maintenance are still quite traditional, the efficiency of the current network O&M is still considerably low. For 5G networks, several factors affect the complexity of the O&M, such as large-scale deployment of base stations, interoperation between 5G and 4G/3G systems, dynamical spectrum sharing between 5G and 4G networks, hundreds of parameters to configure, SDN/NFV-based CN, network slicing, and diversified service requirements from vertical scenarios. Accidents of network-level shutdowns in Japan and Europe were reported, which significantly degraded the reputation of mobile operators. Currently, operators have begun to study the use of AI and big data to support intelligent network operations. However, because these features have not been fully considered in the early days of 5G network design, it is difficult to effectively support intelligent operation in the current 5G network. Therefore, the automation of

O&M will be an essential direction to approach the low cost and high efficiency of the 6G network.

All these aspects require us to innovate the 6G network design and consider how to solve these fundamental problems to achieve flexible and low-cost network deployment.

4. SOLIDS: Proposed 6G mobile network architecture

In response to the driving forces mentioned in Section 3, the 6G network is expected undergo new changes in the following five aspects to improve the network deployment and cost efficiencies: ① The digital twin technology must be introduced in the network; ② the collaborative management of multiparty data and resources should be considered; ③ the cloud-native-enabled and microservice-based architecture need to be further enhanced; ④ the protocol stack requires simplification; ⑤ further decoupling of signaling and data is required.

Based on these considerations, we propose a functional architecture for the future 6G network, called SOLIDS. These network features are explained in detail in the following section. The proposed network architecture is composed of three layers and four planes, as shown in Fig. 8. The three layers include the resource layer, network function layer, and service and application layer. The resource layer provides the underlying resources such as radio, computation, and storage, and provides corresponding support and services for function generation on the network function layer. The network function layer forms a specific network function or combines one or more network functions to ensure the transmission requirements of services from the service layer. The service and application layer provides corresponding support for the services and applications of the customer and achieves service customization.

At the same time, the new data collection, AI, security, and sharing and cooperation planes are introduced. The data collection plane is responsible for the collection, cleaning, processing, and storage of the data in the end-to-end network, and provides data subscription and update services for other layers and planes. The AI plane provided an AI engine. In combination with the functional requirements of each domain in the network, the AI plane provides the corresponding big data analysis, AI algorithms, model training service, and simulation verification of relevant solutions. The functions of the AI plane can be centralized or distributed, which can be distributed in network elements and terminals to support real-time AI applications or centralized in the cloud to implement complex algorithms by using the mass data on the cloud. At the same time, AI capabilities of the AI plane can be exposed, and the external AI capabilities can be imported in the AI plane through the sharing and cooperation plane to achieve the crowdfunding of AI capabilities. By subscribing data from the data collection plane, models from the AI plane, and the required resources from the resource layer, the security plane provides native security support for the entire network. It can even provide security as a service to customers through the sharing and cooperation plane, as well as expose security capabilities to external partners and achieve crowdfunding of external security capabilities. The sharing and cooperation plane achieves the secure sharing of data, models, and possible network capabilities in the network, as well as their crowdfunding and crowdsourcing.

The 5G CN has been implemented as a three-layer architecture, including the resource, virtualization, and network function layers, owing to the introduction of virtualization and cloudification;

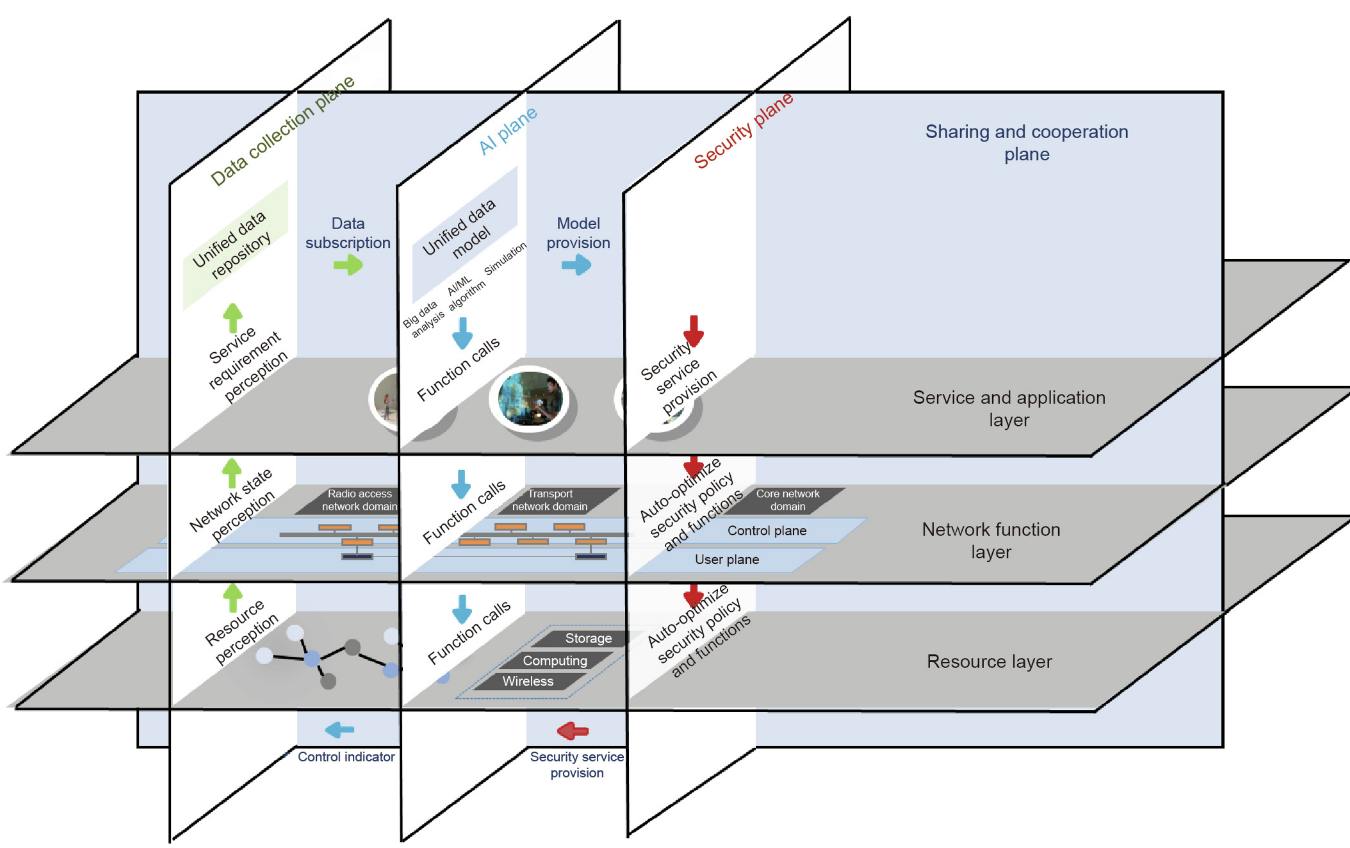

Fig. 8. SOLIDS: the 6G functional network architecture with three layers and four planes.

however, only two planes are considered, the CP and UP. For the 5G RAN, the hardware and software are not decoupled; consequently, there is no definition of layers when compared to the 5G CN. For the 6G network, end-to-end virtualization and cloudification are required; consequently, the concept of three layers is extended to both RAN and CN. Moreover, the data collection and AI planes are introduced in 6G networks to enable native intelligence. It is difficult to provide impeccable security support for networks with the plug-in patch-based security technology; therefore, a dedicated security plane is required.

4.1. Resource layer

With the integration trend of communication, computing, and sensing, the computing power and storage are also important resources that need to be focused on or managed well in addition to the radio resources. From the perspective of network deployment, the economical use of distributed and centralized computing power must be comprehensively considered, including the computing power-sharing mechanism between the terminals and the network.

4.2. Network function layer

The network function layer provides the most basic network service capability. Through the effective combination of RAN, TN, and CN functions, this layer provides guaranteed data transmission for individual users and customized network slicing service capabilities for vertical industries. Of course, the service ability effect of the network function layer cannot be separated from the support of other layers and planes. The network function layer is required to obtain the accurate demands of users from the service and application layer, intelligently arrange the network functions, and reasonably call the network resources with the help of the AI plane, while simultaneously providing safe service assurance for users with the security policy provided by the security plane.

4.3. Service and application layer

The responsibility of the service and application layer is to generate quantifiable requirements for diversified and unpredictable future services and applications. As the most important reference data, the quantified requirements may be required by every other layer or plane. For example, it may be referenced by the resource layer to arrange reasonable computing power or storage resources for a specific application. On the other hand, the service and application layer requires support from other layers or planes, for example, the security guarantee from the security plane, user information provision from the data collection plane, or analysis service from the AI plane.

4.4. Data collection plane

For the data collection plane, both global and local data collection and processing are required, as shown in Fig. 9. For many future scenarios, the user data are not expected to be uploaded to the network. It is necessary to process and store them at local points. Therefore, both centralized and distributed data processing and storage are required.

4.5. AI plane

From a global perspective, both centralized AI and distributed AI are required. As shown in Fig. 10, the centralized AI platform performs global processing using external and internal data and arranges intelligent capabilities according to the requirements of specific use cases. Then, the results are distributed to the AI platform of the specific execution domain. Thus, the network will be full of pervasive AI capabilities.

For the AI capabilities of domains such as the RAN, TN, CN domains, and even user equipment (UE), it must be deployed as closely as possible to provide real-time AI capability support, including models and algorithms.

In the future, the 6G network should also be able to expose its AI capability to users, similar to how it provides communication capabilities today. In addition, it should be able to provide perception capabilities as a service to users. According to user requests, the network should help in scheduling the AI algorithm and model for execution by the terminals. Thus, the AI capability of a network can be better used.

4.6. Security plane

A native security system consists of three elements: intelligent strategy engine, security capability library, and intelligent security O&M. According to the AI learning model, the intelligent strategy engine can intelligently tune the strategies of network elements and security devices, and then build a security capability library. Based on the security requirements from the applications and services, or the network security requirements from the network functions, the security capability library precisely deploys the security functions to achieve active and deep security defense. The intelligent security O&M function achieves automatic security O&M based on AI and big data.

4.7. Sharing and cooperation plane

As mentioned above, 6G systems will introduce external AI capabilities or external data into the network to provide new services and new capacities, or to further improve the data processing efficiency. Moreover, the AI capabilities and analyzed data within the network can also be exposed to third parties to provide them with services and required support. Such crowdfunding and crowdsourcing behaviors are performed on the sharing and cooperation plane. Note that in addition to the AI capabilities and data, the security capabilities, resources, service and application requirements, and network functionalities must be introduced or exposed in 6G networks. Therefore, the sharing and cooperation plane is related to all the other layers and planes.

5. Features of SOLIDS

Based on the aforementioned 6G network architecture, SOLIDS can support six features, including lite, soft, on-demand fulfillment, native intelligence, digital twin, and native security, as shown in Fig. 11.

5.1. Soft

The first feature of the 6G network is its softness. The future network will be an end-to-end software-definable and cloud-native network, which can help achieve rapid service deployment, rapid iteration of functional software versions, dynamic sharing of resources (e.g., radio spectrum, computing, and storage), and network automation and intelligence. The 6G network should exhibit its softness in the following three aspects.

The network should be user-centric at the service and application layers. Because the network serves for users, the design of the network must fully consider the demands of users and activate network functions in an on-demand manner, so that the network can ‘‘move” with the user.

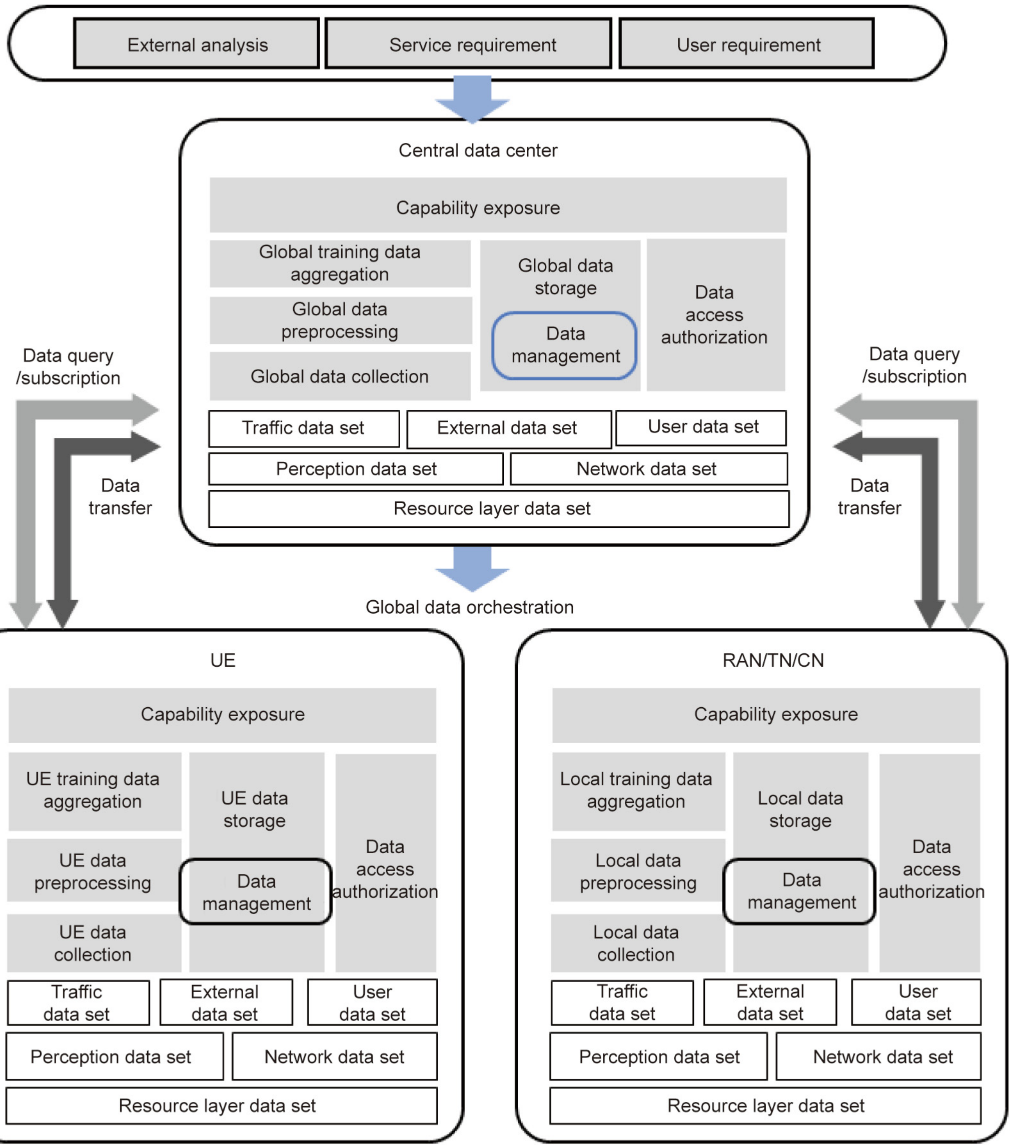

Fig. 9. Logical deployment model of data collection plane.

At the resource layer, the network should have the ability to schedule distributed network resources on demand. To satisfy the requirements of data traffic and transmission latency in new scenarios and applications, the future network will have abilities such as computing and storing within the network, and transforming from a "dumb pipe" to a real "intelligent network." In this context, computing, storage, transport, and other resources will be ubiquitous in the network. Through the introduction of blockchain and other new technologies, on-demand allocation and flexible scheduling of resources among clouds, networks, and edges can be achieved.

At the network function level, the network should be capable of flexibly and independently expanding network element capability, as well as rapidly iterating and evolving software functions. More specifically, the 6G network should be an end-to-end SBA, as shown in Fig. 12. To achieve this, the 6G system will focus on the servitization of the RAN. When compared to the service-based CN, which has basically been completed in 5G networks, the service-based access network must make more changes to the network architecture, such as modifying the layered protocol stack.

5.2. On-demand fulfillment

The second feature of the proposed 6G network is on-demand fulfillment. With the prospect of application scenarios in a digital twin world, a mass of new services and scenarios will emerge in the age of 6G networks, the requirements of which will vary widely. Therefore, the 6G network must have better ability to perceive the behavior, service, intention, and so on of the user, while simultaneously being able to deploy functions, configure parameters and resources based on the requirements of the user, as shown in Fig. 13. Furthermore, the 6G network should provide dynamic services of smaller granularity so that users can freely combine service types and service grades based on their requirements.

To support on-demand fulfillment, a prerequisite is the real-time perception of service and application requirements. Through the cooperation between the service and application layer, the data collection plane, and the AI plane, future service and application requirements can be predicted in advance. Thus, the network can seamlessly switch service mode and service content for users in an on-demand manner when the user requirements change. The

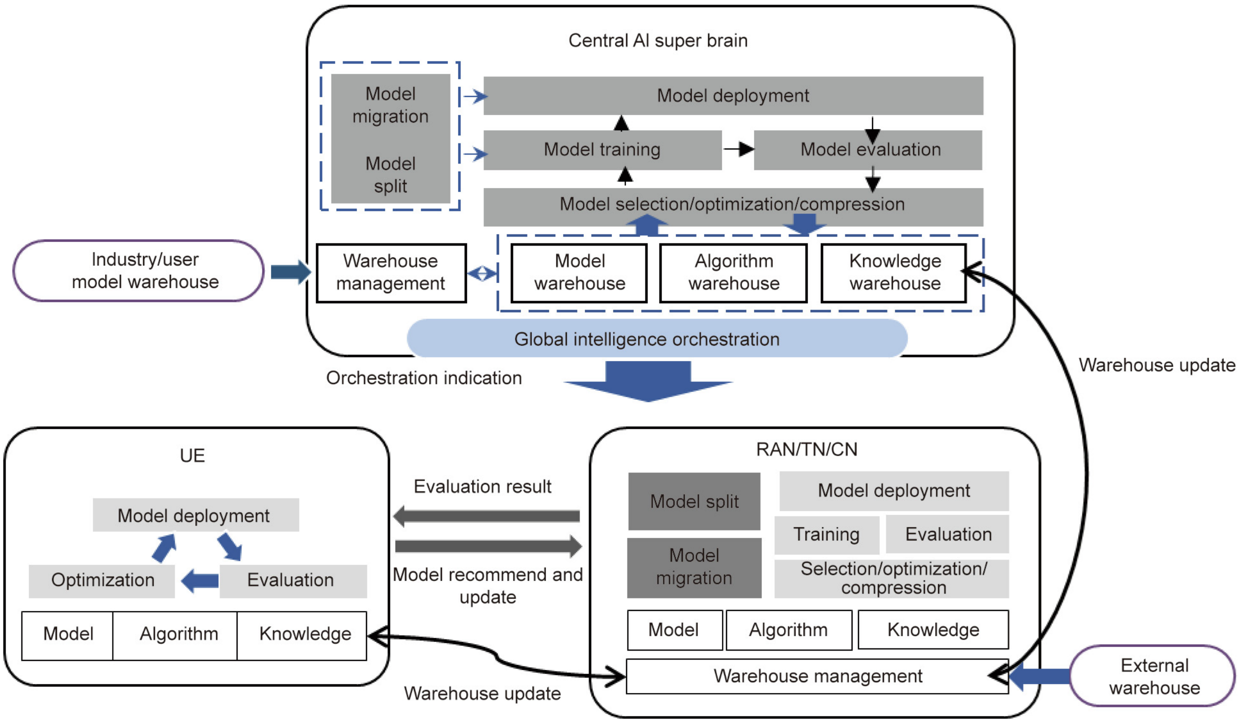

Fig. 10. Logical deployment model of the AI plane.

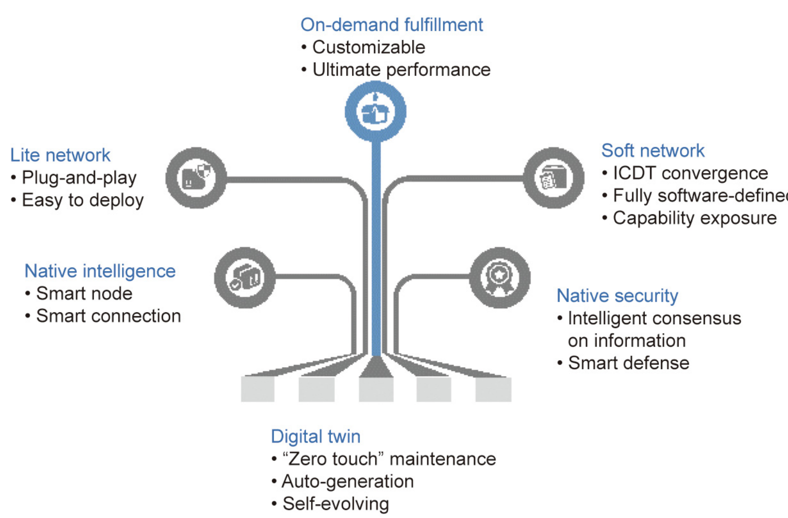

Fig. 11. Six network features.

second prerequisite is the on-demand configuration of radio resources. As a unified infrastructure platform, a distributed resource layer provides radio access, computing, storage, and other resources. In combination with the service and application requirements and the network and resource states as collected data, the network can achieve on-demand resource allocation. The third prerequisite is the on-demand orchestration of the network functions. In the future, these services may have more diversified and personalized requirements for RANs, TNs, and CNs. The network functions combined with the data collected by the perception capability can decompose the demand of the user into RAN, TN, and CN domains, and each domain matches the most appropriate function combination according to its own functional capability.

5.3. Lite

The third feature of the 6G network is the lite network. The 6G network must consider the three-dimensional coverage of air, space, ground, and sea. Managing homogeneous or heterogeneous networks in a unified way to provide users with a consistent experience regardless of handover is a problem that must be considered in the future 6G network [41].

In the traditional method, different networks are designed to support different use cases (e.g., satellite, cellular, cable, and underwater networks). Limited interworking is supported among the different networks to support service continuity. However, service continuity with guaranteed QoS cannot be supported because

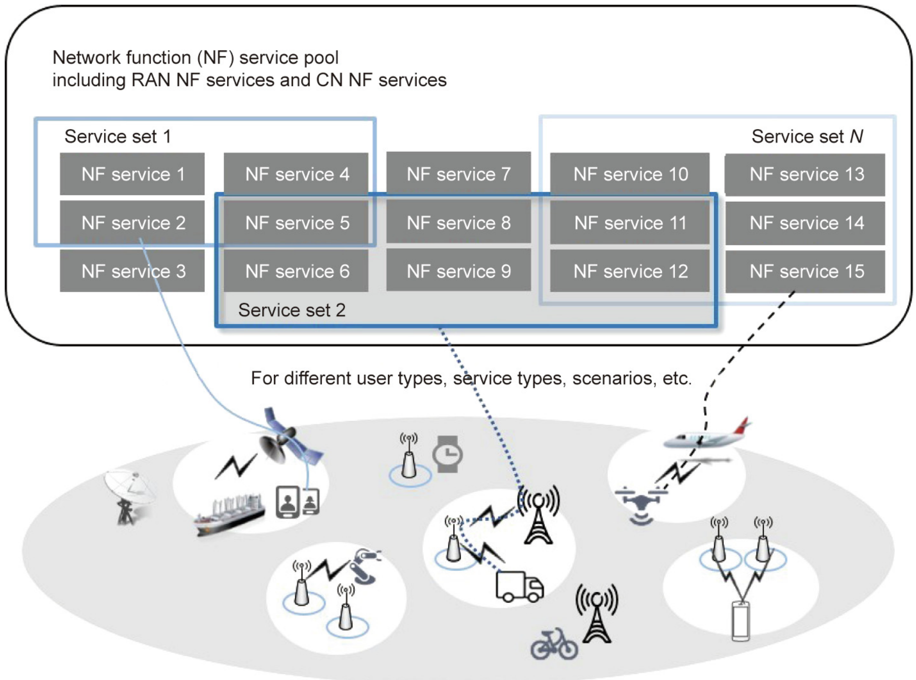

Fig. 12. Soft network.

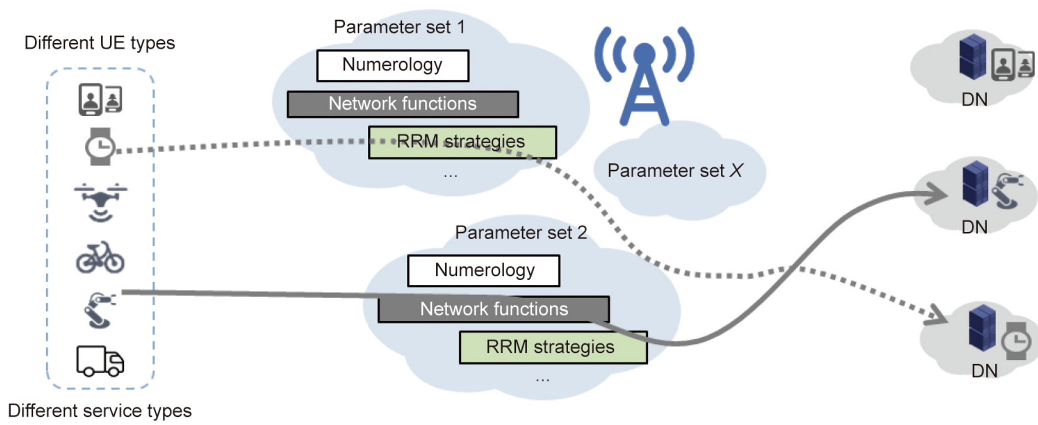

Fig. 13. Example of on-demand fulfillment. RRM: radio resource management.

of the diverse design of different networks on the corresponding protocol and access mechanism. When the user moves out of the desired network, its service can be dropped or degraded.

The network architecture shown in Fig. 14 is an example of a future network architecture in which a unified CN is designed to simplify the network architecture. Through converged radio access technologies (RATs) and unified access mechanisms, one CN is connected to different RATs, and seamless handover between different RATs can be achieved.

The second meaning of the lite network is a light RAN that guarantees reliable mobility management and fast service access through unified signaling and dynamic data access, ensuring the consistency of user service experience, low access latency, and reduced interference between cells and energy consumption of

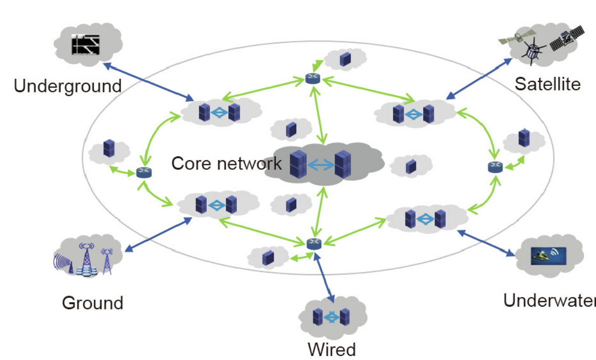

Fig. 14. Unified access CN.

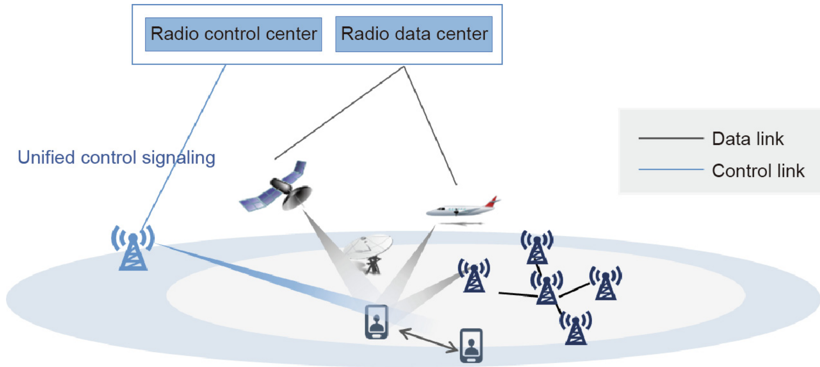

Fig. 15. Unified signaling control.

the radio network, as shown in Fig. 15. The future network may be divided into two layers: One is the wide-area signaling layer, which may work on low-frequency bands such as 700 MHz; and the other is the data access layer, which may work on higher frequency bands with high bandwidth such as 2.3 GHz, 3.5 GHz, millimeter wave, terahertz frequencies, and visible light. The on-demand on–off switching of the data access layer can significantly reduce the number of concurrent serving base stations, which can considerably reduce the cost and power consumption of the network.

5.4. Native intelligence

The fourth feature of the 6G network is native intelligence. AI technology has progressed significantly in recent years. As a tool, it can help operators to improve the O&M efficiency of the network, as well as the efficiency and capability of services. The support vector machine algorithm with different cores was used in Ref. [42] to identify traffic hotspots and predict cell hotspots by using key performance indicators (KPIs) of the mobile network. Ref. [43] proposed a spatiotemporal neural network architecture that can accurately predict the long-term mobile traffic of a region. In a mobile edge computing system, a deep neural network and the digital twin architecture are used to calculate the user association scheme in real time and determine the optimal resource allocation strategy to optimize the energy consumption of the system [44]. Ref. [45] used the density-based spatial clustering of applications with noise algorithm and big DT to track and predict the moving trajectory of users and mine their moving patterns.

In 5G networks, operators have been considering how to use AI to achieve the automatic O&M of the network and attain an intelligent network. However, because the 4G/5G network architecture was not designed with a consideration for AI support, the application of AI in 4G/5G networks is treated in a case-by-case manner, where the data collection, algorithm optimization, and processing are patched to the corresponding network elements or an external processing unit is added as a new network element to the network. For different AI use cases, different modifications to the network may be required, which causes difficulties in the management and operation of the network. Conversely, the data availability is a challenge because the interfaces for data collection have not been predefined in the prior network architecture and protocols, while the current implementation-based data collection server/equipment, such as deep packet inspection or data probing, cannot provide efficient data in time. All these factors determine that the performance and efficiency of AI are far behind expectations.

As shown in Fig. 16, the 6G network architecture will result in ubiquitous and pervasive AI abilities in the network, and provides AI ability on demand, which can be in a distributed or centralized manner, similar to the brain and neural network of the human body. At the same time, through the AI platform, 6G introduces external AI capabilities into the network to provide new services and new capacity, and also introduces external data into the network to further improve the data processing efficiency. In addition, analyzed data and AI capabilities within the network can also be exposed to third parties to provide them with services and required support.

The native intelligence system includes the following six functional entities: AI model management, training data management, AI model evaluation, AI model optimization, data analysis, and knowledge base management, as shown in Fig. 17. The management of AI models includes the selection, generation, storage,

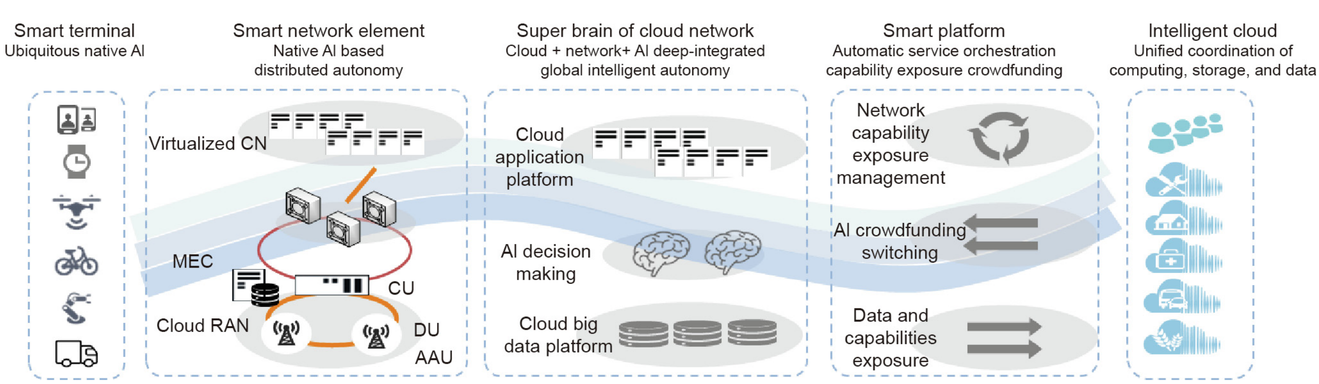

Fig. 16. Native intelligence. MEC: multi-access edge computing.

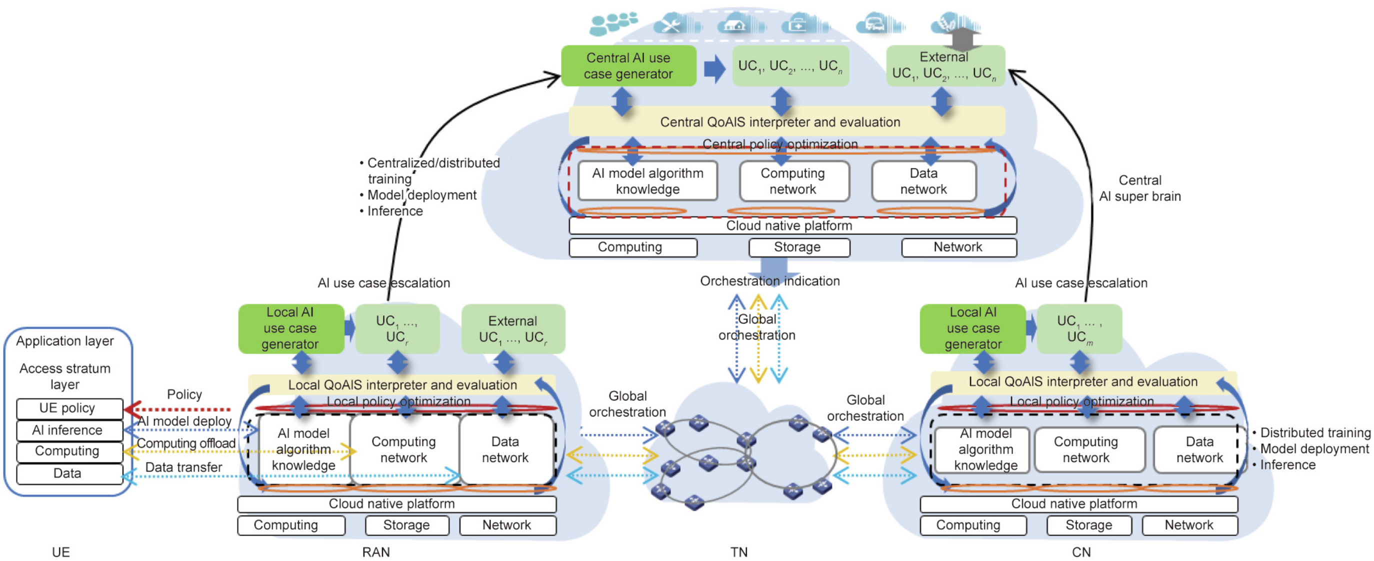

Fig. 17. Logical deployment model of native intelligence. QoAIS: quality of AI service.

update, transfer, and deletion of models. The managed models include various models at the network resource, network function, and service and application layers, which are used to optimize the related entities at these levels. Training data management provides training samples that are required for model training and customizes and preprocesses training data according to the requirements of model performance. AI model evaluation is used to assess the performance of AI models. After each entity optimizes with the reasoning results of the AI model, the performance indicators generated are collected, the evaluation results are calculated and then transmitted to the AI engine. Then, the AI model is optimized accordingly (e.g., change the model structure, change the algorithm, and retraining). Data analysis is the main function that includes AI model training and reasoning based on real-time data. The required data are obtained by data collection. The knowledge warehouse summarizes and abstracts the application results of the AI models, extracts the rules or relational patterns, and the expert experience is externally imported. AI models, training data, and knowledge can be reused and shared by third parties. Meanwhile, the six functions/services of the native intelligence system can also be shared at the network function layer with other parties to provide the required applications and services.

5.5. Digital twin

The fifth feature of a 6G network is the digital twin network [46]. The society beyond 2030 will be a digital twin society as mentioned above, and the digital twin technology can also be applied to the 6G network to achieve full digitalization. Through digital twins, each network entity and user service can be digitized by real-time information collection. Real-time state monitoring, trajectory prediction, and early intervention on possible failures and service drops will be possible; thus, the operational efficiency of the entire network and service efficiency will be improved. It is also possible to verify the effect of the deployment of new features of the network in advance, accelerate the improvement and optimization of new features, achieve rapid and automatic introduction of new functions, and thus the self-evolution of the network.

To develop a digital twin network, obtaining data from all network domains is of vital importance. Based on data processing and the parametric modeling of each network entity and function, we can obtain the digital modeling of the entire network in a virtual space, as shown in Fig. 18. The digital twins of the users may include data such as user trajectory, mobility speed, traffic type, traffic model, data rate, and channel state. Those of the cloud-native RAN may include wireless resource models, hardware resource models, wireless protocol models, wireless channel models, network service states, and the data of KPI and metrics. A similar modeling process and the digital twins can also occur in the TN and CN. Based on the obtained digital twin data, the AI engine predicts the network state using AI algorithms. The optimal state of the physical network will continuously be sought by the network planning entity, verified by simulation, and then mapped onto the network in the real world by enforcing corresponding operations in the management domain.

Conversely, the digital twin of the network can also verify new functionalities, services, and optimization features in the digital domain to avoid any negative impact. A high level of network automation and “zero-touch maintenance” can thus be achieved.

Any network resource object can generate a corresponding digital twin in the digital domain, including the underlying physical resources, network functions, and various applications and services on the upper layer. The generation of digital twins depends on various data acquisition, data processing and storage, and digital twin modeling techniques. To achieve the optimization of resource objects, digital twin networks establish their optimization models and generate their digital plan bodies for future time points based on the collected data and intelligence. Then, by calling the configuration functions, the planning data can be implemented for each resource object. The aforementioned digital twin network functions and functions such as connection management and synchronization optimization between the digital and physical domains are all network functions. The data of the digital twin network include digital twins, digital plans, and intelligent models. Through these functions and data, the digital twin network provides various applications and services in the upward direction and invokes and optimizes various resource objects in the downward direction. Meanwhile, digital twin networks can share the above functions and data with third parties through various sharing technologies. Moreover, it can prevent attacks and tampering through various security technologies.

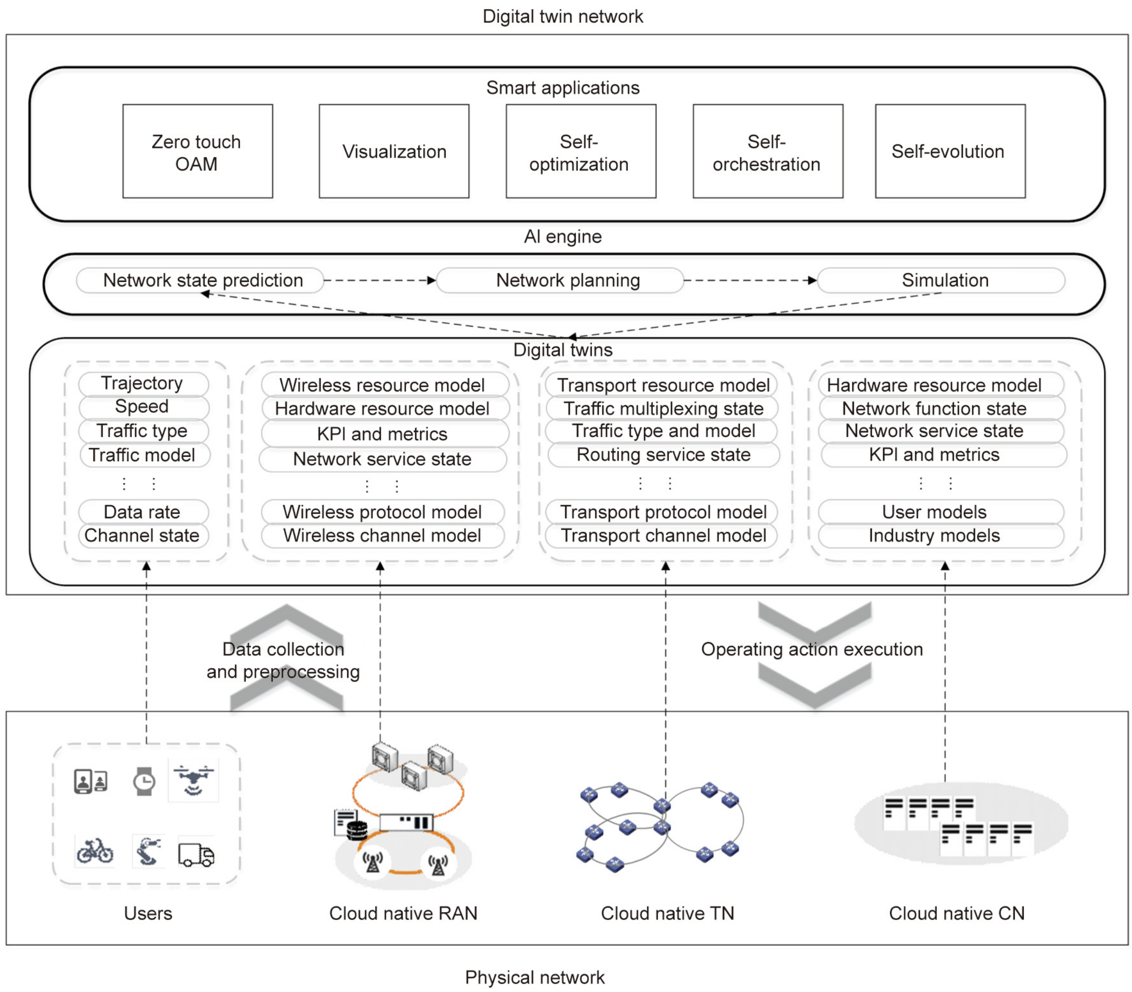

Fig. 18. Digital twin-based autonomy. OAM: operation administration and maintenance.

5.6. Native security

Digital twin society is an inevitable trend in the long-term vision of future social development. Physical objects can all be digitalized and have digital twins, resulting in significant changes in production, life, social governance, and many other aspects.

However, it will also result in more security risks as the changes and operations on digital twins in the virtual space will directly affect its physical counterparts. If the security of a virtual space is threatened, it will directly affect the security of the physical world. It has been predicted that future wars will not be traditional attacks on the physical world, but perhaps on the digital twin world, which could lead directly to the destruction of the physical world. Therefore, security will become particularly important for future society. For 6G networks, digital twins of operator networks, people, factories, enterprises, government infrastructure, and so on will exist in the digital space; consequently, network security is of vital importance.

From the current network design perspective, the design of security is basically in a patched manner and is independent of the network architecture. There is still room for improvement in terms of security and efficiency. The rapid development and progress of cloud computing, big data, and AI technology provides new means and support for the security design of the 6G network. Considering the immune system of the human body as a metaphor, the safety system of the 6G network can be similarly designed, which can actively protect against risk and security attacks, and can also be guaranteed and controlled by external intervention when the former fails.

Therefore, the last feature of a 6G network is its native security as shown in Fig. 19. The network monitors the security status in real time and predicts potential risks. It combines attack resistance with risk prediction to achieve intelligent security, such as risk prediction and active immunity. Intelligent consensus will be formed through the interaction and collaboration of networked intelligent entities and will be used to eliminate interference and provide a high level of security for information and data. Proactive and deep security defenses will be achieved by precisely deploying security features and optimizing security policies based on AI and big DT. Active immunity for network infrastructure, software, and so on will be provided to enhance the security level of the basic platform using trusted computing technology. Through the ubiquitous coordination of the end, edge, network, and cloud, the security status of the entire system will be accurately perceived, and security risks will be appropriately disposed to maximize the security of the 6G network. The network will achieve a comprehensive upgrade from internet security to cyberspace security.

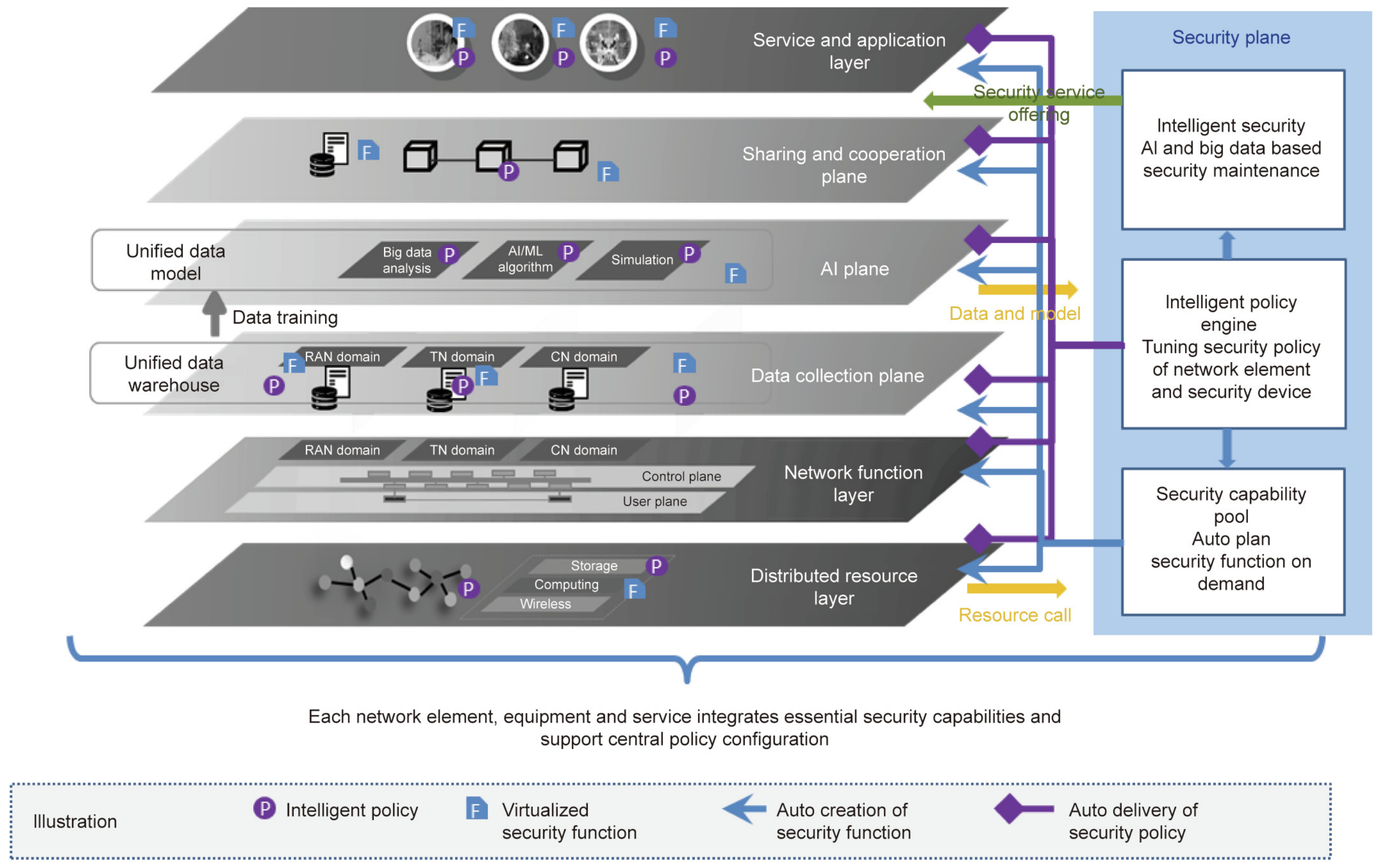

Fig. 19. Logical deployment model of native security.

6. Conclusions

The rapid commercialization of 5G networks is accelerating the development of big data, cloud computing, and AI, which leads to the deep convergence of ICDT. ICDT convergence highlights a cloud-native and software-definable mobile network. The design of the 6G mobile network will consider the problems and challenges faced by 5G networks, trend of ICDT convergence, and requirements of new services and applications driven by the digital twin world beyond 2030. This article recalled the evolution history of the mobile network architecture, proposed a functional architecture of a 6G mobile network with three layers and four planes, and identified its intrinsic network features, that is, soft, on-demand fulfillment, lite, native intelligence, digital twin, and native security. The detailed mechanisms, procedures, protocols, and interface design will be further studied in future works.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (2020YFB1806800).

Compliance with ethics guidelines

Guangyi Liu, Na Li, Juan Deng, Yingying Wang, Junshuai Sun, and Yuhong Huang declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] [5G vision and requirements]. Report. Beijing: IMT-2020 (5G) Promotion Group; 2014 May. Chinese.
[2] Liu G, Jin J, Wang Q, Dong J, Lou M, Huang Y, et al. Vision and requirements of 6G: digital twin and ubiquitous intelligence. Mob Commun 2020;44(6):3–9. Chinese.
[3] Liu G, Huang Y, Li N, Dong J, Jin J, Wang Q, et al. Vision, requirements and network architecture of 6G mobile network beyond 2030. China Commun 2020;17(9):92–104.
[4] [Vision and requirements for 2030+ (version 2.0)]. Report. Beijing: China Mobile Research Institute; 2020 Nov. Chinese.
[5] Nakamura T. 5G evolution and 6G. In: Proceedings of 2020 International Symposium on VLSI Design, Automation and Test; 2020 Aug 10–13; Hsinchu, China. Piscataway: IEEE; 2020. p. 1–5.
[6] [Technology trends for 2030+ white paper]. Report. Beijing: China Mobile Research Institute; 2020 Nov. Chinese.
[7] Saad W, Bennis M, Chen M. A vision of 6G wireless systems: applications, trends, technologies, and open research problems. IEEE Netw 2020;34(3):134–42.
[8] Li R, Toy M, Trossen D, Humphreys D, Sarian V, Bhandari S, et al. Network 2030: a blueprint of technology, applications and market drivers towards the year 2030 and beyond. Report. Focus Group Technologies on Network 2030; 2019 Jul.
[9] Weissberger A. ITU-R: future technology trends for the evolution of IMT towards 2030 and beyond (including 6G) [Internet]. IEEE ComSoc Technology Blog; 2020 Oct 16 [cited 2020 Nov 11]. Available from: https://techblog.comsoc.org/2020/10/16/itu-r-future-technology-trends-for-the-evolution-of-imt-towards-2030-and-beyond-including-6g/.
[10] [Network architecture outlook for 2030+ white paper]. Report. Beijing: China Mobile Research Institute; 2020 Nov. Chinese.
[11] Viswanathan H, Mogensen P. Communications in the 6G era. IEEE Access 2020;8:57063–74.
[12] Dong H, Zheng D, Wang R. The beyond 3G network structure and system applications. Mob Commun 2004;S2:26–8. Chinese.
[13] Li Z, Wang X, Huang Y, Liu G, Liu J, Tang H, editors. [Technology and standards of TD-LTE]. Beijing: Posts & Telecom Press; 2013. Chinese.
[14] Liu G, Fang M, Guan H, Li Y, Sun C. 5G mobile communication: a fully connected world. Telecommun Sci 2019;9:166. Chinese.
[15] TR 23.740: Study on enhancement of network slicing. 3GPP standard. France: 3GPP; 2018.
[16] Software-defined networking: the new norm for networks. Report. Open Networking Foundation; 2012 Apr.
[17] GSNFV-MAN 001-V1.1.1: Network functions virtualization (NFV); management and orchestration. ETSI standard. France: ETSI; 2014.
[18] New WID enhancement on management aspects of 5G service-level agreement [presentation]. In: 3GPP TSG-SA5 Meeting 87e; 2020 Feb 24–Mar 4; online; 2020.

[19] New WID on autonomous network levels [presentation]. In: 3GPP TSG-SA5 Meeting 88e; 2020 May 25–Jun 3; online; 2020.
[20] New WID self-organizing networks (SON) for 5G networks [presentation]. In: 3GPP TSG-SA5 Meeting 85; 2019 Sep 17–20; Newport Beach, CA, USA; 2019.
[21] Huawei. SP-180899: New WID on intent driven management service for mobile networks [presentation]. In: 3GPP TSG-SA5 Meeting 81; 2018 Sep 12–14; Gold Coast, QLD, Australia; 2018.
[22] GS ZSM 002: Zero-touch network and service management (ZSM); reference architecture (V1.1.1). ETSI standard. France: ETSI; 2019.
[23] GS ZSM 001: Zero-touch network and service management (ZSM); requirements based on documented scenarios (V1.1.1). ETSI standard. France: ETSI; 2019.
[24] Our community [Internet]. Palo Alto: Telecom Infra Project; c2021 [cited 2020 Nov 11]. Available from: https://telecominfraproject.com/.
[25] Omar S. Rakuten Cloud platform is a blueprint for the future [Internet]. Cisco Blogs; 2019 Mar 22 [cited 2020 Nov 11]. Available from: https://blogs.cisco.com/sp/rakuten-cloud-platform-is-a-blueprint-for-the-future.
[26] [Vision and requirements for 2030+ white paper]. Report. Beijing: China Mobile Research Institute; 2019 Nov. Chinese.
[27] Xie Y. The evolution of NB-IoT standard system and development of Internet of Things industry. Chin J Internet Things 2018;2(1):76–87. Chinese.
[28] Zhang L, Li F, Hu Z, Zhao Y, Cao G, Hu Y. Study on key technology and deployment scheme of LTE-eMTC. Des Tech Posts Telecommun 2018;7:1–5. Chinese.
[29] Mayer-Schönberger V, Cukier K, editors. Big data: a revolution that will transform how we live, work, and think. New York: Houghton Mifflin Harcourt; 2013.
[30] Zhou T, Pan Z, Cheng X. Developing tendency prediction of big data in 2019 from CCF TFBD. Big Data Res 2019;5(1):109–15. Chinese.
[31] Y.3173: Framework for evaluating intelligence levels of future networks including IMT-2020 (Study Group 13). ITU-T standard. Geneva: ITU; 2020.
[32] TR 37.816: Study on RAN-centric data collection and utilization for LTE and NR. 3GPP standard. France: 3GPP; 2019.
[33] TR 23.791: Study of enablers for network automation for 5G. 3GPP standard. France: 3GPP; 2019.
[34] Way forward on U-plane modifications to LTE baseline [presentation]. In: 3GPP TSG-RAN2 95bis; 2016 Oct 10–14; Kaohsiung, China; 2016.
[35] TS 38.321: 5G new radio-medium access control (MAC) protocol specification. 3GPP standard. France: 3GPP; 2021.
[36] TS 38.322: 5G new radio-radio link control (RLC) protocol specification. 3GPP standard. France: 3GPP; 2021.
[37] TS 38.323: 5G new radio-packet data convergence protocol (PDCP) specification. 3GPP standard. France: 3GPP; 2021.
[38] Wang W, Lu G, Chen X. Key technologies and challenges of commercial 5G core network. ZTE Tech J 2020;26(3):9–16. Chinese.
[39] Nie H, Zhao H, Mao C. Deployment issues and suggestions of 5G core network. ZTE Tech J 2020;26(3):3–8. Chinese.
[40] TR 38.801: Study on new radio access technology-radio access architecture and interfaces. 3GPP standard. France: 3GPP; 2017.
[41] Liu C, Lu L, Wang S, Hu Y. Prospects for a multi-access air–space-terrestrial integrated 6G network architecture. Mob Commun 2020;44(6):116–20. Chinese.
[42] Zhou L, Chen X. SVM hotspot identification for cellular networks. In: Proceedings of IEEE 5th International Conference on Computer and Communications; 2019 Dec 6–9; Chengdu, China. Piscataway: IEEE; 2019. p. 1103–7.
[43] Zhang C, Patras P. Long-term mobile traffic forecasting using deep spatio-temporal neural networks. 2017. arXiv:1712.08083.
[44] Dong R, She C, Hardjawana W, Li Y, Vucetic B. Deep learning for hybrid 5G services in mobile edge computing systems: learn from a digital twin. IEEE Trans Wirel Commun 2019;18(10):4692–707.
[45] Zhao Z, Zhang P, Huang H, Zhang X. User mobility modeling based on mobile traffic data collected in real cellular networks. In: Proceedings of 11th International Conference on Signal Processing and Communication Systems; 2017 Dec 13–15; Surfers Paradise, Australia. Piscataway: IEEE; 2017. p. 1–6.
[46] Y.DTN-ReqArch; Requirements and architecture of digital twin network. ITU-T standard. Geneva: ITU; 2021.