Key Technologies of Space-Air-Ground Integrated Network: A Comprehensive Review

Chuanfeng Wei[1,2,3,4](✉), Yuan Zhang[2,3,4], Ruyan Wang[2,3,4], Dapeng Wu[2,3,4], and Zhidu Li[2,3,4]

[1] Macro Net Communication Co., Ltd., CASC, Beijing , China
[2] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
[3] Key Laboratory of Optical Communication and Networks, Chongqing, China
[4] Key Laboratory of Ubiquitous Sensing and Networking, Chongqing, China

Abstract. Facing the urgent needs of wide area intelligent network and global random access, the independent development of terrestrial cellular communication system and satellite communication system will face great challenges in the future. Space-air-ground integrated network is considered to be potential in integrating the space-based network and terrestrial network to realize unified and efficient resource scheduling and network management. In this paper, the architecture, functional requirements, challenges and key technologies of the space-air-ground integrated network are reviewed. It is expected that the paper is able to provide insightful guidelines on the research of the space-air-ground integrated network.

Keywords: Space-air-ground integrated network · Software defined network · Mobile edge computing · Network slicing · Deep reinforcement learning

1 Introduction

With the development of science and technology in conjunction with the continuous expansion of human production and activity space, the emerging network technologies represented by Internet of Things (IoT) will gradually become the main body of the future network demands [1–3]. Compared with the communication needs of ordinary people, IoT communication would greatly expand in terms of space range and communication content so that a variety of IoT devices and services would cover a wider area such as mountains, deserts, oceans, deep underground, air, and space [4]. Owing to the advances in wireless network in

---

This work was supported in part by the National Natural Science Foundation of China under grants 61871062 and 61771082, and in part by Natural Science Foundation of Chongqing under grant cstc2020jcyj-zdxmX0024, and in part by University Innovation Research Group of Chongqing under grant CXQT20017.

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 63–80, 2021.
https://doi.org/10.1007/978-3-030-89814-4_5

recent years, 5G network technology has provided more flexible services, larger capacity, and higher efficiency for new network applications, such as virtual reality, autonomous driving, and smart cities, and has entered the deployment and actual commercial use phases [5]. As for IoT applications, 5G network has specifically devised two significant service scenarios, namely ultra-reliable and low latency communication (uRLLC) and massive machine type communication (mMTC) [6–8].

The technologies promoted by 5G network technology, like narrow band IoT (NB-IoT), beamforming, uplink/downlink decoupling, are able to solve key technical issues in IoT, such as the wide area coverage, energy consumption, and massive connections [9]. However, large-scale 5G network deployment would cause huge costs which consist of the infrastructure cost brought by intensive base station deployment and backhaul network construction, as well as the installation, rent and maintenance cost for optical fiber cable. [10] Meanwhile, it is hard for ground-based networks to cover extremely remote areas, oceans, deep underground, air, and even deep space. Therefore, 5G ground-based network technology is difficult to meet the ubiquitous communication need due to the greatly expanded network space [11]. Furthermore, the demand for multi-dimensional comprehensive information resources in future information services will gradually increase since the efficient operation of services in the fields of national strategic security, disaster prevention and mitigation, aerospace and navigation, education and medical care, environmental monitoring, and traffic management all rely on the comprehensive application of multi-dimensional information such as air, space, and ground [12]. In this context, constructing space-air-ground integrated network, deeply integrated space-based network, air-based network and ground-based network, so as to give full play to the functions from different network dimensions is able to break the barriers of data sharing between independent network systems and realize wide-area full coverage and network interconnection, which will further trigger an unprecedented information revolution.

Space-air-ground integrated network is a large-scale heterogeneous network that can collect and process network resources, which has the characteristics of wide coverage, multiple coexisted protocols, highly dynamic nodes, variable topology and high throughput for providing seamless wireless access services to the world. Space-air-ground integrated network needs to be intelligent to enhance network performance and management efficiency urgently due to its own heterogeneity and other characteristics. To adapt to 5G ecosystem, integrated systems, especially satellite networks, have to provide wireless communication services in a more flexible, agile and cost-effective manner [13].

This paper reviews the basic architecture, network requirements and the faced challenges of space-air-ground integrated network. Additionally, the application and advantages of mobile edge computing (MEC), software defined network (SDN), network slicing and deep reinforcement learning (DRL) in space-air-ground integrated network are discussed in detail. It is expected that the paper is able to provide insightful guidelines on the research of the space-air-ground integrated network.

The reminder of this paper is organized as follows. Section 2 introduces the architecture of the space-air-ground integrated network. In Sect. 3, existing issues and challenges of the space-air-ground integrated network are discussed. In Sect. 4, key technologies related to the space-air-ground integrated network are introduced and discussed in detail. Section 5 finally concludes the paper.

2 Space-Air-Ground Integrated Network

2.1 Network Architecture

Space-air-ground integrated network is based on ground-based network, supplemented and extended by space-based network and air-based network, providing various network applications in a wide area with ubiquitous, intelligent, collaborative and efficient information assurance infrastructure, whose architecture is shown in Fig. 1. Specifically, ground-based network, mainly consisting of ground internet and mobile communication network, is responsible for network services in service-intensive areas. Air-based network is composed of high-altitude communication platforms and unmanned aerial vehicle (UAV) ad hoc networks, which has functions of coverage enhancement, enabling edge service and flexible network reconfiguration. Space-based network is composed of various satellite systems to form space-based backbone networks and space-based access networks, realizing functions such as global coverage, ubiquitous connection and broadband access. Space-air-ground integrated network is able to utilize various comprehensive resources effectively through the in-depth integration of multi-dimensional networks to perform intelligent network control and information processing, so as to cope with network services with different requirements, achieving the objective of network integration, functional service and application customization, where space-based network is the core enabling technology to build the ubiquitous space-air-ground integrated network.

Space-based network contains satellites, constellations and corresponding ground infrastructure, like terrestrial stations and network operation and control centers, in which satellites and constellations have different characteristics and work in different orbits. According to the height above the ground, these satellites can be divided into three categories: geostationary earth orbit (GEO) satellites, medium earth orbit (MEO) satellites and low earth orbit (LEO) satellites. On the other hand, they can also be divided into narrowband and broadband satellite networks according to the channel bandwidth.

- Narrowband satellite network mainly refers to MEO/LEO satellite systems, such as the Iridium and Globalstar satellite systems, which mainly provide voice and low-rate data services for global users.
- Broadband satellite network can transmit large amounts of data due to its wide frequency band, where broadband is a general term in fixed or wireless communications. Broadband satellite system usually uses the Ka frequency

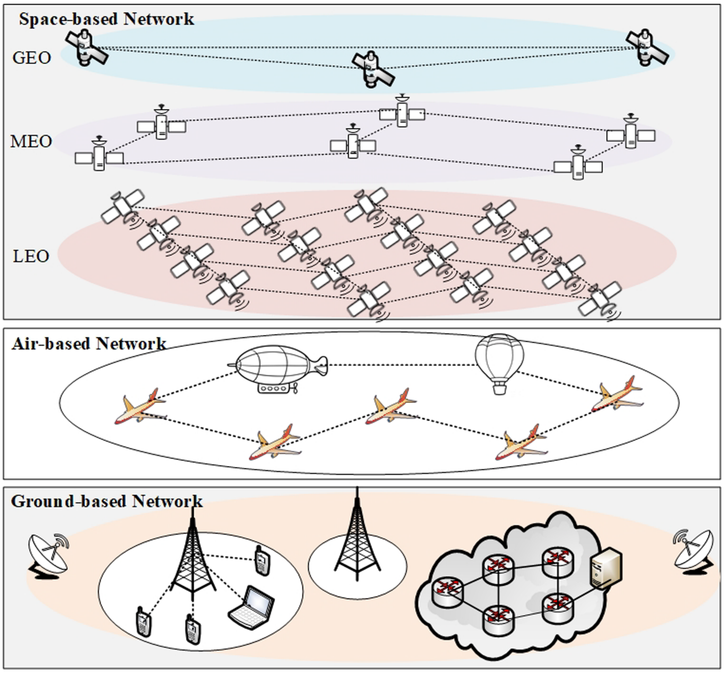

Fig. 1. Space-air-ground integrated network architecture.

band with a data transmission rate of up to 10 Gbps [14] to provide services for high-speed internet access and multimedia information. In addition, the data transmission rate of broadband satellite system is expected to reach 1000 Gbps by 2020 [15].

- Multi-layer satellite network is mainly formed by satellite networks in different orbits [16], which is a kind of practical next-generation satellite network architectures. GEO, MEO and LEO satellite systems could build a typical, tridimensional and multi-layer satellite network through inter-satellite link and inter-layer link (ILL).

Air-based network uses space aircraft as carriers for information acquisition, transmission and processing. UAVs, airships and balloons are the main infrastructures that constitute high-altitude and low-altitude platforms, providing broadband wireless communication complementary to the ground-based network [17]. Compared with the base stations in the ground-based network, the space-based network has the characteristics of low cost, easy deployment and large coverage area to provide high-speed wireless access services in the area covered by it.

Ground-Based Network. Ground-based network is mainly composed of ground mobile cellular networks, mobile ad hoc networks [18], worldwide interpretability for microwave access (WiMAX) network [19], and wireless local area network (WLAN). Ground-based network can provide users with high-speed data transmission, however, whose coverage is limited, especially in rural and remote areas.

Comparison of Various Networks. The space-based network can provide seamless coverage to the global, but owns higher propagation delay. On the other hand, ground-based network infrastructure is vulnerable to natural disasters or man-made damage although it has lower transmission delay. Space-based network has advantages in terms of transmission delay and network coverage, but its capacity and wireless links are limited and unstable, respectively, which should be taken into consideration when deploying such a network [20].

2.2 Network Requirements

Compared with traditional satellite networks, space-air-ground integrated network should own more comprehensive capabilities, not only traditional communication capability, but also computing, artificial intelligence (AI) and security capabilities.

(1) Requirements on communication capability
Communication capability is the basic requirement for the future development of 6G space-air-ground integrated network. Owing to this, space-air-ground integrated network allows terminal travelling speed to exceed 1000 km/h, transmission rate to be greater than 1 Gbps, transmission delay to be less than 10 ms, and spectrum efficiency to be more than 4 times higher than traditional satellite communication systems, respectively.

(2) Requirements on computing capability
In the future, network nodes such as constellation satellites and high-altitude platforms would be an important part of the 6G computing network. It is necessary to provide a high-performance computing platform under resource-constrained conditions for supporting a computing rate of up to tens of trillions per second and the ability to coordinate and migrate computing resource dynamically with the goal of creating a high-performance, low-latency, and large-bandwidth service environment for wide-area business and network management and adapting to the characteristics of rapid changes in network topology and limited satellite load resources.

(3) Requirements on AI capability
AI capability is the core capability of space-air-ground integrated communication system, including five capabilities: perception, learning, reasoning, prediction, and decision-making. Perception capability is able to perceive transmission requirements and security risks through ubiquitous information collection followed by the identification of new services and new threats using self-learning

and self-evolution capabilities. Reasoning capability can be adopted to reason and predict possible future service changes and network events through the big data technology. Eventually, use various technologies comprehensively to make decisions to provide users with personalized services and jointly optimize and schedule system resources.

(4) Requirements on security capability
The security capability of space-air-ground integrated network incorporates the security of data transmission and network behavior with the objective to solve the security and privacy issues caused by the wide participation of various subjects in heterogeneous complex networks. Space-air-ground integrated network needs to form a growing endogenous security mechanism to deal with dynamically changing network security threats.

3 Exiting Issues and Challenges

Space-air-ground integrated network has the characteristics of heterogeneity, highly dynamic of space nodes, highly dynamic of network topology, extremely large space-time spans, limited resources of space nodes and vulnerable attribute of satellite broadcast transmission links, which puts forward higher requirements for the design of network architecture, satellite-to-ground integrated communication standards, and inter-satellite networking protocols.

(1) Highly dynamic of space nodes
The nodes of the traditional ground-based network are relatively fixed, while high-speed relative movements between various satellites and the ground-based network exist in space-air-ground integrated network. In case of considering the access and networking of nodes such as high-altitude platforms, airplanes and low-altitude UAVs, their moving characteristics are usually less regular and the impact will be greater. One of the main effects brought by the highly dynamic of nodes is the serious Doppler frequency offset. At the same time, the communication links are prone to high interruption rate and high symbol error rate. These transmission characteristics put forward higher requirements for the design of space-air-ground integrated communication standards.

(2) Highly dynamic of network topology
In space-air-ground integrated communication system, the network consists of nodes at different levels such as satellites, high-altitude platforms, mid-to-low-altitude suspension/aircraft and ground equipment, and has a tridimensional architecture that is completely different from traditional ground-based cellular communication networks. The movements of nodes in the network will also cause the network topology to exhibit highly dynamic characteristics.

On the one hand, highly dynamic of network topology will cause link changes, making it difficult to transmit data through fixed traffic. On the other hand, the network protocols face the contradiction between asymmetric links, link quality changes and highly reliable transmission feedback control under different changes such as multi-hop and relay, which will cause low transmission efficiency at the application layer and even fail to guarantee data transmission quality.

(3) Extremely large space-time spans
Due to the long distance and fast speed of altitude platform, the links between network nodes far exceed the ground-based network in space-time spans. On the one hand, the transmission signal between nodes has a large attenuation loss and is easily affected by a series of factors such as orbital changes, elevation angles, sun glints, atmospheric scattering, rain attenuation, and blocking, resulting in weak received signals and various interferences, which would be more serious in the high frequency band. On the other hand, the influence of long-distance transmission and space environment will also cause high delay and large jitter in the communication process, resulting the contradiction between timely adjustment and long delay in the feedback adaptation mechanism, which makes it difficult for conventional error control methods to work.

(4) Heterogeneous network interconnection
Space-air-ground integrated communication system is composed of a variety of heterogeneous networks, and the environment and characteristics of each network vary greatly. For the satellite space-based network, the satellite signal and transmission characteristics of different orbital positions are quite different; for the ground-based network, the different communication environment conditions such as air, space and ground are also very different. In order to achieve the integration, compatibility and amalgamation of air, space, and ground-based networks, it is necessary to have uniformity and similarity in the design of waveform systems and communication standards. The aforementioned heterogeneous network interconnection characteristics will inevitably make this goal more difficult.

(5) Limited load resources
The communication satellites are systems with limited power resources. On the one hand, due to the rocket launch capability, the weight and size of communication satellites are limited, which restricts the size of solar panels and the available power resources. On the other hand, the power consumption of the payload is greatly limited by the satellites thermal radiation capability. The larger the area of thermal radiator is, the stronger the heat dissipation capacity is, and the higher power consumption can be supported. However, due to the limitation of the outline size of the satellite fairing and the influence of the satellite antenna, the size of the thermal radiator cannot be increased without limit. As the demand for satellite communications develops toward higher peak rates and more connections, the contradiction between on-board power resource constraints and increased transmission power and on-board processing capabilities will further intensify.

(6) Wide area transmission security
Space-air-ground integrated network uses satellites, high-altitude platforms and other means to achieve wide area coverage of users. However, the wireless channel of satellite communication has the characteristics of openness and broadcasting, which makes the information transmission channel uncontrollable and the wireless link more susceptible to threats such as man-made interference, attacks, eavesdropping and replay. Therefore, the development of space-air-ground integrated

networks needs to solve the transmission security challenges under the conditions of wide area coverage.

4 Key Technology

4.1 Mobile Edge Computing

As one of the important technologies of mobile networks, MEC is playing an increasingly important role in the process of current network development. MEC can be regarded as a cloud service platform running at the edge of the network, which can support the deployment of service processing and resource scheduling functions, improve service performance and users experience and reduce the transmitted data of backhaul links and bandwidth pressure of the core network to a certain extent [21].

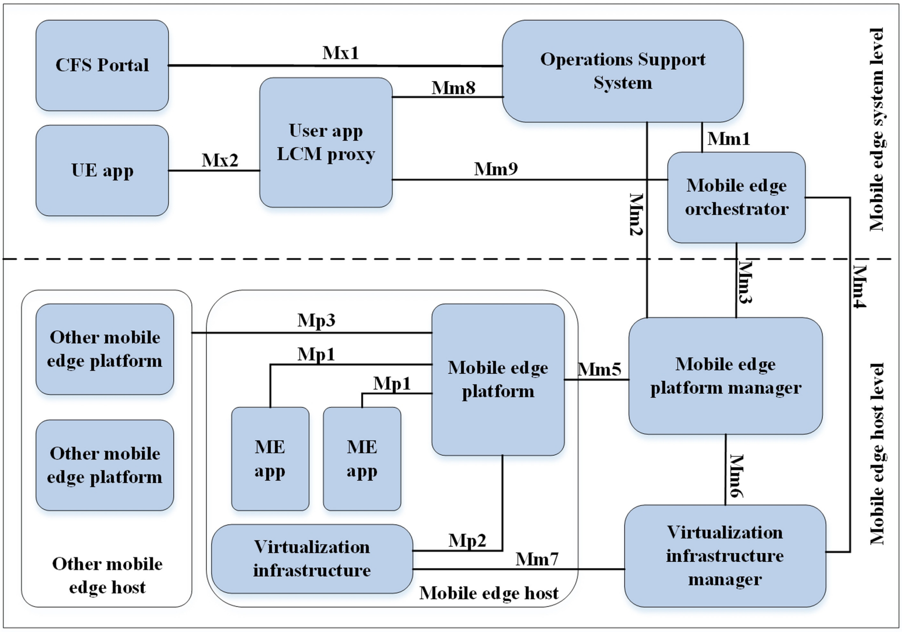

Fig. 2. MEC reference architecture.

An MEC reference system architecture is depicted in Fig. 2, which was proposed by the European Telecommunication Standards Institute (ETSI) [22]. It can be seen from Fig. 2 that MEC can serve users directly through user terminal applications or customer facing service (CFS) portal. Both the user terminal and the CFS portal interact with the MEC system through MEC system-level management.

Mobile edge users can instantiate, terminate or relocate related applications and services through the user application lifecycle management agent managed by the MEC system. Then, the operation support system (OSS) decides whether to approve the related request, and the approved request will be forwarded to the mobile edge orchestrator which is the core function of MEC system-level management and responsible for maintaining an overall view of available computing/caching/network resources and MEC services. The mobile edge orchestrator allocates virtualized MEC resources to the application to be launched based on application requirements, like latency. In addition, the orchestrator can flexibly extend the available resources down/up to the already running applications.

The MEC system-level management is interconnected with the MEC server-level management that constitutes the mobile edge platform and the virtualization platform manager. The former is responsible for the management of application life cycle, application rules and service authorization, traffic rules, etc.; the latter is responsible for allocating, managing and releasing the virtualized computing and caching resources provided by the virtualized infrastructure in the MEC server. The MEC server, an important part of the reference architecture, represents virtualized resources and carries MEC applications that run as virtual machines on the virtualized infrastructure. The advantages of introducing MEC from the perspective of user terminals are to further satisfy users QoS requirements and reduce terminal energy consumption. From the perspective of operators, it is mainly to further reduce core network traffic and improve network scalability and security. The introduction of edge computing can effectively deal with some of the challenges faced in space-air-ground integrated network, which has become an important development trend in the future. These advantages are discussed as follows separately.

(1) Real-time QoS guarantee

With the rapid development of smart terminals, their performance is constantly improved and perfected, but most smart terminals still lack sufficient performance to complete real-time use cases with predefined QoS requirements. Offloading the computing tasks to cloud servers through satellites with computing offloading can effectively meet the computing requirements of terminal devices with limited capacity. However, the rapid development of various emerging service applications, such as Internet of Things, autonomous driving, 4 K/8 K video transmission, etc., has put forward new requirements such as ultra-low latency and high QoS guarantees for space-air-ground integrated network. Through edge computing, the users computing tasks can be directly processed at the edge server without being transmitted to the remote cloud server, which greatly reduces the processing delay of the task and improves the users QoE. Compared with cloud computing, the offloading solution based on edge computing can effectively reduce delay when processing IoT data and guarantee the QoS requirements of users [23].

(2) Energy consumption optimization

Energy consumption is one of the most important parameters when considering mobile devices. Although the processing power of smart terminals is steadily improving, battery life has not improved at the desired rate. With the development of various computationally intensive applications, executing these applications on the device itself will result in very high energy consumption. In this case, although offloading the computing tasks to the cloud server can reduce the computing energy consumption of the mobile devices to process the computing tasks to a certain extent, the transmission energy consumption of the mobile devices for the transmission tasks will also increase due to the long distance between the cloud server and mobile terminals. Therefore, the introduction of edge computing in space-air-ground integrated network and offloading computing tasks to the edge of the network closer to users will help to further reduce the energy consumption of mobile devices. Energy consumption of applications, like face recognition and augmented reality, is analyzed in [24], which shows that compared with cloud servers, offloading tasks to edge servers can effectively reduce the energy consumption of device terminals. Therefore, the introduction of edge computing is very necessary for mobile devices with limited battery energy in space-air-ground integrated network.

(3) Core network traffic scheduling
The core network has limited bandwidth and is vulnerable to congestion. According to the latest forecast report released by Cisco, by 2021, the total amount of global equipment will reach 75 billion, and mobile traffic will exceed 24.3 EB/month [25]. Thus, operators face huge challenges in managing accumulated data traffic with different sizes and characteristics. In the traditional satellite-ground collaborative network, the traffic generated by mobile devices accesses the core network through satellites or other access devices and further accesses the cloud server. Assume that these services can be satisfied at the edge of space-air-ground integrated network, the burden on the core network can be greatly reduced and bandwidth utilization can be optimized. This transformation of the network prevents billions of devices at the edge from consuming the limited bandwidth of the core network which makes that the services that the core network is responsible for become manageable in scale and be simplified. Therefore, the introduction of edge computing can effectively solve the congestion problem of the core network and data center.

(4) Scalability
The number of terminal devices is expected to reach trillions within a few years. Therefore, the scalability issue is one of the major challenges facing space-air-ground integrated network [26]. To support these real-time changing dynamic requirements, the cloud can be scaled accordingly. However, transmitting large amounts of data to the cloud server could cause congestion in the data center. Moreover, it is more difficult for operators to work due to the constantly changing data traffic generated by terminal equipment. In this case, the centralized structure of cloud computing cannot provide a scalable environment for data and applications. Introducing edge computing in space-air-ground integrated network and distributing services and applications in the form of virtual machines (VM) on edge servers for copying them can greatly improve the scalability of the entire system [27]. The corresponding service can be copied to another nearby edge

server and the request can be further processed when the edge server becomes crowded and cannot satisfy the incoming request. In addition, edge computing can preprocess data at the edge of space-air-ground integrated network, which can greatly reduce the traffic forwarded to the cloud server and the scalability burden of the cloud.

(5) Security and reliability
In space-air-ground integrated network, if all data is transmitted back to the main server, the operation process and data are extremely vulnerable to attacks. Distributed edge computing will allocate data processing work among different data centers and devices. Thus, the attacker cannot affect the entire network by attacking one device. If the data is stored and analyzed locally, the security team can easily monitor it, thereby greatly improving the security of the entire system. Furthermore, compared with cloud computing, edge computing provides better reliability. Edge computing servers are deployed closer to users, so the possibility of network interruption is greatly reduced.

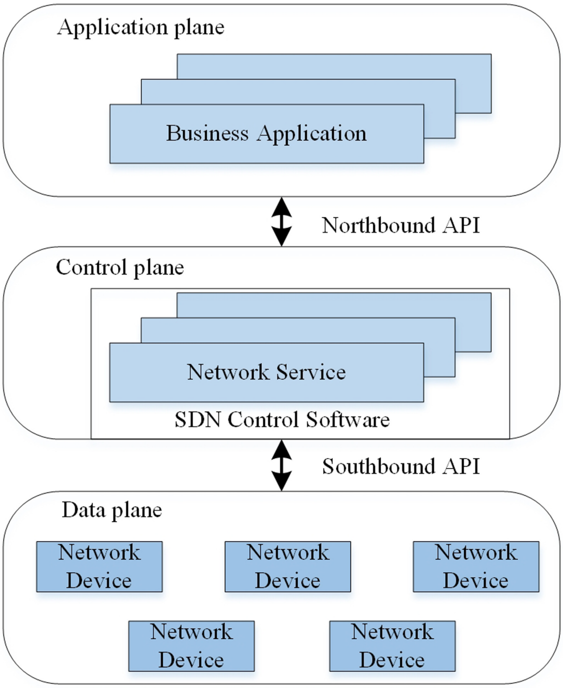

Fig. 3. SDN architecture.

4.2 Software Defined Network

As an emerging network architecture, SDN separates the control plane and data forwarding plane of traditional switching equipment, uses software technology to abstract the underlying network infrastructure, manages and allocates network resources flexibly on demand through the unified and programmable application program interface. The entire SDN architecture is composed of three logical planes: application plane, control plane and data plane, where data transmission and information exchange between logical planes are realized through application program interface, whose architecture is shown in Fig. 3. The devices in SDN have a unified logical structure. Under the action of the controller, these devices only need to execute data forwarding instructions, which can eliminate the network heterogeneity caused by different hardware devices. Using software concepts, SDN can redefine the network architecture to support the new requirements of the ecosystem in the future integrated network. In integrated systems, especially satellite networks, SDN can break through many limitations on network operations and providing end-to-end services, which has become a key factor in promoting technological innovation and network evolution. The introduction of SDN into space-air-ground integrated network can centrally manage heterogeneous networks with different network architectures and communication protocols, thereby reducing network configuration and management overhead, improving network performance and optimizing QoS.

In view of the problems existing in space-air-ground integrated network, the advantages of applying SDN technology to space-air-ground integrated network are as follows:

(1) Flexible routing strategy
In traditional satellite networks, static snapshot routing methods are generally used to ensure the reliability and controllability of the network, but the high dynamics of satellite networks lead to unsatisfactory requirements such as load balancing. Although some dynamic routing methods based on distributed link information collection can meet some of the above requirements, they make it impossible for satellites to easily obtain a global view of the network and can only achieve local optimal routing. In the SDN-based architecture, the control plane obtains a real-time global view of the network state through the network state information uploaded by the data plane and the communication between the controllers to centrally manage the satellite nodes of the data plane, which makes it possible to provide more flexible global routing calculations and routing strategies in the face of highly dynamic network topologies, such as load balancing, multicast path correction, and node failure management. In addition, the control plane with a global view of the network can effectively update the global configuration of the network when the space information network needs to be expanded or updated, allowing newly launched satellites to seamlessly access the existing space information network. On the other hand, the control plane can make timely adjustments to the network configuration, assign neighboring nodes to be responsible for the coverage area and network tasks of the failed node and replace the failed node when a satellite node fails [28].

(2) Convenient network configuration
Considering limited resources, small memory and low central processing unit (CPU) processing power in space information network with the increase of the number of space-based network applications, the amount and complexity of tasks that need to be handled by the on-board payload have also increased greatly, which makes the configuration of space information network very difficult compared with ground-based networks. The core idea of the SDN architecture, the separation of control and forwarding simplifies the processing function of the satellite node, eases this problem to a large extent. What the satellite node of the data plane needs to do is to receive and execute various configuration information distributed by the control plane and feedback its own network state information to the control plane. Complex network configuration and control functions, as well as functions like collecting information from the data plane and building a global view of the network, are handed over to the controller in the control plane of the space-ground double backbone to complete.

(3) Better compatibility
SDN architecture has a unified data exchange standard and programming interface, which can manage the entire network equipment in a unified manner when the network is heterogeneous. The flow table in SDN architecture abstracts the two-layer forwarding table and the three-layer routing table, integrates network configuration information at all levels and is able to simultaneously process various protocols coexisting in the space information network, like delay tolerant network (DTN) etc., so as to well solve the problem of heterogeneity of space-based network protocols.

(4) Lower hardware cost
In traditional satellite networks, satellite nodes have to complete complicated processing tasks which are often the most complicated and expensive part. After adopting the SDN-based architecture, the data plane satellite node is just a simple network forwarding device, which simplifies the architecture of the satellite function, effectively reduce the satellite design and production costs, simplify the complexity of satellite management and make space information network more flexible and controllable. In addition, the control plane structure and inter-satellite link forwarding mode of the space-ground double backbone also reduces the number of required terrestrial stations and the investment in infrastructure.

4.3 Network Slicing

Network slicing was first proposed as a key technology in 5G mobile communication networks and one of the important features of 5G network architecture. Applying slicing technology to space-air-ground integrated network can expand the application range of 5G network slicing and improve overall network performance.

The network slicing technology in 5G networks allows operators who share the same infrastructure to configure the network for the slicing and define specific functions. However, for different application scenarios, it has different network bandwidth and node computing processing capabilities, and can be flexibly and dynamically create and repeal slices based on the operators strategy. Applying network slicing technology to space-air-ground integrated network slicing can build mutually independent virtual logical networks on general equipment, provide customized network functions for a variety of different services, meet users QoS requirements, and improve the scalability and the flexibility of network resource. For applications with relatively high delay requirements, the ground-based network is adopted. At the same time, the broadcast application signals can be transmitted through satellites, which greatly improves the utilization of network resources. Space-air-ground integrated network slicing is composed of physical networks and virtual networks that provide different services for different applications, as shown in Fig. 4 [29].

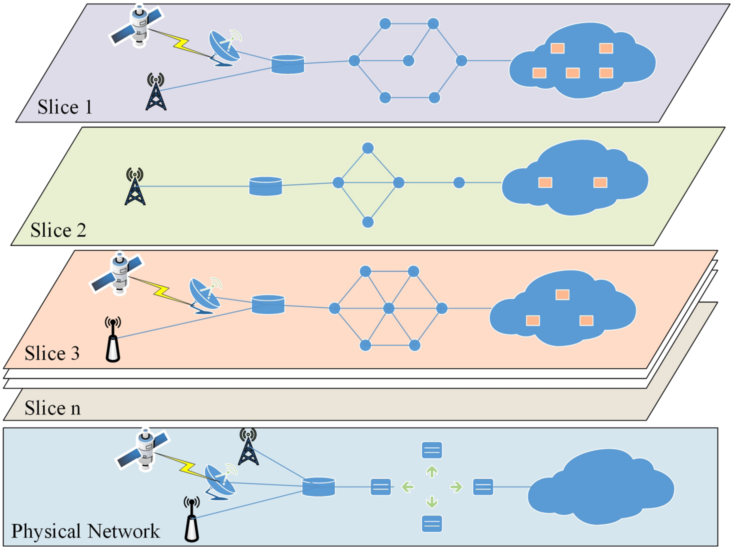

Fig. 4. Network slicing diagram.

The realization of converged network slicing is mainly based on satellite-ground segment virtualization, virtual network embedding (VNE), service function chaining (SFC), etc.

The satellite communication system is composed of space segment, ground segment and user segment. The ground segment includes satellite ground gateway station, ground satellite control centers, tracking, telemetry and command stations. The gateway station is the interface between the satellite system and the ground public network, which can be used by ground users to enter and exit the satellite system to form a link. The ground segment is the management segment of the satellite network, completing the connection between the satellite network and the ground-based network, to allocate resources and charge. Similar to VM, the virtualization of the satellite-ground segment is mainly to virtualize the functions of the satellite-ground segment equipment such as gateways, satellite control centers, tracking, telemetry and command stations, and realize their functions on general equipment rather than specific equipment. Run the same virtual machine on different devices to achieve the same functions. The satellite-ground segment virtualization process firstly converts the satellite signal from the analog signal to digital signal through the receiver, and then transmits it to the cloud server through the network protocol to virtualize the function.

VNE needs to solve the problem of mapping between virtual network functions and physical networks. The computing power required by virtual network nodes and the bandwidth requirements required by virtual network links are mapped to physical network devices through a certain function. How to establish a suitable mapping that meets the virtual network requirements is an important research problem.

SFC links different virtual network function blocks according to application requirements to construct a complete function link. The traffic and requests in the virtual network fluctuate continuously over time, and algorithms are required to continuously optimize according to the dynamic virtual network environment to balance the stability and effectiveness of the virtual function link. In addition, for some applications that require relatively high latency, it is necessary to allocate resources in advance according to the prediction of the virtual network environment.

The virtualization of the satellite-ground segment, VNE and SFC have jointly laid the foundation for converged network slicing.

4.4 Deep Reinforcement Learning

As one of the most important research directions in the field of machine learning, RL has had a significant impact on the development of artificial intelligence in the past 20 years [30]. RL is a learning process in which the agent can make decisions regularly, observe the results, and then automatically adjust its strategy to achieve the optimal strategy. Although the convergence of this learning process has been proven, it usually takes a lot of time to explore and obtain knowledge of the entire system to ensure convergence to the optimal strategy. Therefore, naive RL is not suitable for large-scale and high-complexity network environment such as space-air-ground integrated network. In order to overcome the limitations of RL, DRL has received more and more attention as a new breakthrough technology. Different from the traditional RL enumerating the mapping relationship between the environmental state and the optimal strategy action through the Q table, DRL uses deep neural network (DNN) to replace the Q table, which can theoretically fit the complex mapping relationship of any characteristic, thereby improving the learning rate and the performance of the RL algorithm. In practice, deep learning has become the theoretical support for emerging industries such as robot control, computer vision, voice recognition, and natural language processing. In the field of communication and networking, deep learning has been used as an emerging tool to effectively solve various problems and challenges [31]. For the future networks represented by the space-air-ground integrated network, it contains a variety of heterogeneous and complex network slices or elements, such as IoT devices, mobile users, UAV base stations, and low-orbit satellite nodes, etc. These heterogeneous network entities need to make various decisions on different space-time scales in a centralized or distributed manner, like network and spectrum selective access, data rate and transmit power control, base station and satellite handover, to achieve the maximization of different network optimization goals including throughput and the minimization of network energy consumption [32]. In a highly dynamic and uncertain network environment, most decision-making problems can be modeled as Markov decision process (MDP). Although MDP can be solved in theory by dynamic programming, heuristic algorithm, and RL technology, considering the large-scale and complex modern networks, technologies like dynamic programming and heuristic algorithm would be difficult to be used due to the large amount of calculation. Therefore, DRL has developed into a core solution to overcome this challenge [33]. Applying DRL method to space-air-ground integrated network has the following advantages.

(1) DRL is able to solve network optimization problems in complex environments. Non-convex and complex problems can be optimized without complete and accurate network information by deploying DRL algorithms on network controllers, like base stations or core network controllers.
(2) DRL allows network entities to establish knowledge of the network environment during the learning process without the need to preset channel models or user mobility modes. For example, in space-air-ground integrated network, the network can monitor user distribution or network environment changes in real time using DRL, and gradually learn the optimal base station selection, channel selection, handover, caching and offloading decisions, without having to be based on abstract or inaccurate environmental model.
(3) DRL greatly improves the response speed, especially in complex problems with large state and action space. Therefore, in a large-scale network represented by space-air-ground integrated network, DRL allows the network controller to dynamically control a large number of mobile users and heterogeneous devices based on relatively real-time environmental information.
(4) Some other problems in space-air-ground integrated network, such as cyber-physical attacks, interference management and data offloading, can be modeled as game theory related problems. DRL has recently been used as an effective tool to solve some complex game theory problems, like finding the Nash equilibrium without complete information.

5 Conclusion

This paper provides an overview on the space-air-ground integrated network. Network capability requirements, challenges, as well as the potential key technologies are discussed. In specific, MEC can reduce the data transmission of backhaul links and the bandwidth pressure on the core network, such that service performance and user experience are improved. SDN is able to manage the network resources centrally, thereby reducing network configuration and management overheads. Network slicing can provide customized network functions for a variety of different services, such that network resource scalability and networking flexibility can be guaranteed. DRL can is suitable for important tasks such as service orchestration, resource scheduling, network access, and mobility management in the future space-air-ground integrated network.

References

1. Chettri, L., Bera, R.: A comprehensive survey on internet of things (IoT) toward 5G wireless systems. IEEE Internet Things J. **7**(1), 16–32 (2020)
2. Xiong, J., et al.: A personalized privacy protection framework for mobile crowdsensing in IoT. IEEE Trans. Ind. Inf. **16**(6), 4231–4241 (2020)
3. Xiong, J., et al.: Enhancing privacy and availability for data clustering in intelligent electrical service of IoT. IEEE Internet Things J. **6**(2), 1530–1540 (2019)
4. Cirillo, F., Gmez, D., Diez, L., Elicegui Maestro, I., Gilbert, T.B.J., Akhavan, R.: Smart city IoT services creation through large-scale collaboration. IEEE Internet Things J. **7**(6), 5267–5275 (2020)
5. Huang, S., Zeng, Z., Ota, K., Dong, M., Wang, T., Xiong, N.: An intelligent collaboration trust interconnections system for mobile information control in ubiquitous 5G networks. IEEE Trans. Netw. Sci. Eng. **8**, 347–365 (2020). https://doi.org/10.1109/TNSE.2020.3038454
6. Li, J., Zhang, X.: Deep Reinforcement Learning-Based Joint Scheduling of eMBB and URLLC in 5G Networks. IEEE Wireless Communications Letters **9**(9), 1543–1546 (2020)
7. Xiong, J., Zhao, M., Bhuiyan, M.Z.A., Chen, L., Tian, Y.: An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT. IEEE Trans. Ind. Inf. **17**(2), 922–933 (2021). https://doi.org/10.1109/TII.2019.2957130
8. Pokhrel, S.R., Ding, J., Park, J., Park, O.-S., Choi, J.: Towards enabling critical mMTC: a review of urllc within mMTC. IEEE Access **8**, 131796–131813 (2020)
9. Wan, S., Hu, J., Chen, C., Jolfaei, A., Mumtaz, S., Pei, Q.: Fair-hierarchical scheduling for diversified services in space, air and ground for 6G-dense internet of things. IEEE Trans. Netw. Sci. Eng, (2020). https://doi.org/10.1109/TNSE.2020.3035616
10. Ye, J., Dang, S., Shihada, B., Alouini, M.S.: Space-air-ground integrated networks: outage performance analysis. IEEE Trans. Wirel. Commun. **19**(12), 7897–7912 (2020)
11. Niu, Z., Shen, X.S., Zhang, Q., Tang, Y.: Space-air-ground integrated vehicular network for connected and automated vehicles: challenges and solutions. Intell. Converged Netw. **1**(2), 142–169 (2020)
12. Sharma, S.R.S., Vishwakarma, N., Madhukumar, A.: HAPS-based relaying for integrated space-air-ground networks with hybrid FSO/RF communication: a performance analysis. IEEE Trans. Aerosp. Electron. Syst. **57**(3), 1581–1599 (2021). https://doi.org/10.1109/TAES.2021.3050663
13. Jiang, C., Li, Z.: Decreasing big data application latency in satellite link by caching and peer selection. IEEE Trans. Netw. Sci. Eng. **7**(4), 2555–2565 (2020)
14. Farserotu, J., Prasad, R.: A survey of future broadband multimedia satellite systems, issues and trends. IEEE Commun. Mag. **38**(3), 128–133 (2000)

15. Gharanjik, A., M.R., B.S., Arapoglou, P., Ottersten, B.: Multiple gateway transmit diversity in Q/V band feeder links. IEEE Trans. Commun. **63**(3), 916–926 (2015)
16. Nishiyama, H., Tada, Y., Kato, N., et al.: Toward optimized traffic distribution for efficient network capacity utilization in two-layered satellite. IEEE Trans. Veh. Technol. **62**(3), 1303–1313 (2013)
17. Chandrasekharan, S., Gomez, K., Al-Houran, A., et al.: Designing and implementing future aerial communication networks. IEEE Commun. Mag. **54**(5), 26–34 (2016)
18. Anjum, S., Noor, R.M., Anisi, M.H.: Review on MANET based communication for search and rescue operations. Wirel. Pers. Commun. **94**(1), 31–52 (2017)
19. Rooyen, M.V., Odendaal, J.W., Joubert, J.: High-gain directional antenna for WLAN and WiMAX applications. IEEE Antennas Wirel. Propag. Lett. **16**, 492–495 (2017)
20. Zhang, N., Zhang, S., Yang, P., et al.: Software defined space-air-ground integrated vehicular networks: challenges and solutions. IEEE Commun. Mag. **55**(7), 101–109 (2017)
21. Taleb, T., Samdanis, K., Mada, B., et al.: On multi-access edge computing: a survey of the emerging 5G network edge architecture and orchestration. IEEE Commun. Surv. Tutorials **19**(3), 1657–1681 (2017)
22. Xie, R., Lian, X., Jia, Q.: Survey on computation offloading in mobile edge computing. J. Commun. **39**(11), 138–155 (2018)
23. Shukla, R.M., Munir, A.: A computation offloading scheme leveraging parameter tuning for real-time IoT devices. In: IEEE International Symposium Nanoelectronic Information Systems pp. 208–209 (2016)
24. Ha, K., Pillai, P., Lewis, G., et al.: The impact of mobile multimedia applications on data center consolidation. In: IEEE International Conference on Cloud Engineering, pp. 166–176 (2013)
25. CISCO: Cisco visual networking index: global mobile data traffic forecast update 2016–2021 white paper (2017)
26. Wang, W., Tong, Y., Li, L., et al.: Near optimal timing and frequency offset estimation for 5G integrated LEO satellite communication system. IEEE Access **7**, 113298–113310 (2019)
27. Vaquero, L.M., Rodero-merino, L.: Finding your way in the fog: towards a comprehensive definition of fog computing. Comput. Commun. Rev. **44**(5), 27–32 (2014)
28. Ren, C., et al.: Enhancing traffic engineering performance and flow manageability in hybrid SDN. In: IEEE GLOBECOM (2016)
29. Ordonez-Lucena, J., et al.: Network slicing for 5G with SDN/NFV: concepts, architectures, and challenges. IEEE Commun. Mag. **55**(5), 80–87 (2017)
30. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. The MIT Press, Cambridge (1998)
31. Luong, N., Hoang, D., Gong, S., et al.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutorials **21**(4), 3133–3174 (2019)
32. Liu, J., Shi, Y., et al.: Space-air-ground integrated network: a survey. IEEE Commun. Surv. Tutorials **20**(4), 2714–2741 (2018)
33. Kato, N., Fadlullah, Z.M., Tang, F., et al.: Optimizing space-air-ground integrated networks by artificial intelligence. IEEE Wirel. Commun. **26**(4), 140–147 (2019)