Chinese Journal of Aeronautics, (2022), 35(9): 1–18

Chinese Society of Aeronautics and Astronautics
& Beihang University

Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com

REVIEW

6G service-oriented space-air-ground integrated network: A survey

Nan CHENG [a], Jingchao HE [a], Zhisheng YIN [b], Conghao ZHOU [c], Huaqing WU [c], Feng LYU [d], Haibo ZHOU [e,*], Xuemin SHEN [c]

[a] *School of Telecommunications Engineering, Xidian University, Xi'an 710071, China*
[b] *School of Cyber Engineering, Xidian University, Xi'an 710071, China*
[c] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L 3G1, Canada*
[d] *School of Computer Science and Engineering, Central South University, Changsha 410083, China*
[e] *School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China*

Received 12 May 2021; revised 27 May 2021; accepted 17 June 2021
Available online 30 December 2021

KEYWORDS

Mobile Edge Computing (MEC);
Network Function Virtualization (NFV);
Network slicing;
Service-oriented network;
Software Defined Networking (SDN);
Space-Air-Ground Integrated Networks (SAGINs)

Abstract As an indispensable component of the emerging 6G networks, Space-Air-Ground Integrated Networks (SAGINs) are envisioned to provide ubiquitous network connectivity and services by integrating satellite networks, aerial networks, and terrestrial networks. In 6G SAGINs, a wide variety of network services with the features of diverse requirements, complex mobility, and multi-dimensional resources will pose great challenges to service provisioning, which urges the development of service-oriented SAGINs. In this paper, we conduct a comprehensive review of 6G SAGINs from a new perspective of service-oriented network. First, we present the requirements of service-oriented networks, and then propose a service-oriented SAGINs management architecture. Two categories of critical technologies are presented and discussed, i.e., heterogeneous resource orchestration technologies and the cloud-edge synergy technologies, which facilitate the interoperability of different network segments and cooperatively orchestrate heterogeneous resources across different domains, according to the service features and requirements. In addition, the potential future research directions are also presented and discussed.

1. Introduction

With the development of wireless networks, the information services have been evolving, satisfying the requirements of people's life, vertical fields, and the society. From 2G to 4G mobile network, the services mainly focus on the data rate, with the typical services such as Internet surfacing, video streaming, and online gaming. The 5G mobile network also considers network applications and services for vertical areas, such as

* Corresponding author.
E-mail address: haibozhou@nju.edu.cn (H. ZHOU).
Peer review under responsibility of Editorial Committee of CJA.

Production and hosting by Elsevier

https://doi.org/10.1016/j.cja.2021.12.013
1000-9361

industry, medical domain, and transportation. There are three service scenarios classified in 5G, which are enhanced Mobile BroadBand (eMBB), ultra-Reliable and Low-Latency Communications (uRLLC), and massive Machine-Type Communications (mMTC).[1,2] The eMBB scenario is mainly for further evolvement of broadband services, such as high-quality video streaming and big data cloud storage, and thus the technical focus of eMBB is improving the spectrum efficiency and data rate. The uRLLC scenario is mainly for the latency and reliability-critical services such as remote surgery, industry automation, and driving safety. As the name suggests, it focuses on guaranteeing transmission latency and reliability. The mMTC scenario is mainly for the dense machine-type communications, such as body monitoring and smart city, which focuses on energy efficiency and network capacity. Such a scenario classification in 5G network can help identify the services and enabling technologies in a certain scenario, and therefore each service, as long as is classified, can be fulfilled by a set of technologies accordingly. Nevertheless, the 3-fold classification is relatively coarse, and can hardly cover all services and applications accurately.

In the emerging 6G era, the network services will continue to evolve, with the particular features of intelligence, complexity, dynamics, and customization. 6G is envisioned as an Information-Communication-Data technology convergence network, where the big data and Artificial Intelligence (AI) are considered as fundamental components, i.e., native AI. A soaring number of intelligent services, such as networked robots, cognitive Internet of Things (cognitive IoT),[3] self-driving, and digital twins, will require abundant networked AI capability. Meanwhile, the increasing needs of ubiquitous global coverage and high-speed network access have urged the developments of Space-Air-Ground Integrated Networks (SAGINs). With SAGINs, diverse services can be supported anywhere and anytime. However, end-to-end services in SAGINs over large geographical areas could cross many network domains, and thus require the orchestration of multi-dimensional resources. In addition, the complex and heterogeneous mobility in SAGINs makes the service fulfillment more challenging. Moreover, 6G network is envisioned to provide specified services with distinctive and long-tail requirements for each user, further increasing the difficulty in network orchestration and management. Meanwhile, with the increasing customizing capability of the network (e.g., with Software-Defined Networking (SDN) and Network Function Virtualization (NFV) technologies), services deployed by users themselves will emerge and should be paid more attention. Therefore, network service fulfillment, i.e., supporting network services with network capabilities, is a significant yet challenging issue in 6G network. However, the key technologies and methods in 6G network service fulfillment are still under investigation of academia and industry.

As an indispensable characteristic of 6G network, the SAGINs provide seamless global network coverage, which is essential for rural area Internet access, maritime communications, disaster recovery, autonomous driving, etc.[4,5] In future 6G SAGINs, satellites, High-Altitude Platforms (HAPs), and Unmanned Aerial Vehicles (UAVs) will be integrated with terrestrial networks. Compared to conventional terrestrial wireless communications, satellite communications are more cost-effective in service coverage,[6] and can provide broadband network access with multi-beam antennas.[7] In Iridium constellation, 66 satellites with inter-satellite links provide network access to users anywhere and anytime. In recent years, SpaceX's STARLINK introduced the concept of broadband satellite Internet. In the satellite constellation proposed in 2016, the first period of STARLINK (1584 satellites in 24 planes) can provide high-date-rate network access for users in 95% of the area between 58° north latitude and 58° south latitude. HAPs, operating at a height of 17–22 km above Earth's surface in the stratospheric region, can be used in several different scenarios such as broadcast/multicast HDTV signal, high-speed network access, navigation, disaster and emergency detection, remote sensing, and emergency communications.[8] In addition, UAVs have gained attention due to their flexibility and ability of fast response.[9–13] For example, during a special event, such as concerts and sport matches when the cellular network is not sufficient for serving the crowded area, UAVs can be dynamically deployed to assist the terrestrial network to enhance service qualities.[14,15]

Compared with terrestrial networks, on-demand service fulfillment is more challenging in SAGINs. In SAGINs, a network service may include nodes, links, and resources across multiple network domains and segments. Therefore, the service fulfillment is significantly affected by different network protocols, dynamic link conditions, high mobilities, and heterogeneous resources, individual of which renders service fulfillment in SAGINs overwhelmingly challenging. In addition, as SAGINs generally cover much larger areas than conventional terrestrial networks, it is required to handle much more services simultaneously, which further complicates the service fulfillment process. However, the network architecture and key enabling technologies for on-demand service fulfillment in SAGINs have not been studied systematically. In this paper, we present the concept of 6G service-oriented SAGINs from the perspective of network services and applications, and discuss the advantages and enabling technologies. Furthermore, recent research works, technical challenges, and future directions are reviewed and described. This paper could provide insights in research and development of 6G SAGINs, especially addressing the issues of fulfilling complicated and numerous services with dynamic network environment and multi-dimensional network resources.

The remainder of the paper is organized as follows. Section 2 introduces components and applications of 6G SAGINs and main challenges in service-oriented techniques. Section 3 describes the overall architecture of the paper. Sections 4 and 5 discuss related works on two main research issues in service-oriented techniques in 6G SAGINs, where Section 4 focuses on the match between services and heterogeneous resources, and Section 5 investigates cloud-edge synergy. Section 6 presents three potential research directions in future service-oriented SAGINs. Finally, Section 7 concludes the paper.

2. Service-oriented SAGINs applications and requirements

Different from conventional terrestrial networks, SAGINs have distinctive features, in terms of architecture, protocols, resource orchestration, and services. The service-oriented SAGINs are a combination of SAGINs network architecture, service management framework, and on-demand service technologies. In this section, we provide a brief introduction on

SAGIN network architecture and analyze the requirements of service-oriented SAGINs. Then, we give the design of service management framework and the review of enabling technologies in the following sections.

2.1. Introduction on SAGINs network architecture

The SAGINs architecture should support the integration of space networks, aerial networks, and terrestrial networks with significant heterogeneity.[16] Compared with conventional terrestrial base stations, LEO satellites are highly mobile and one satellite may cover an area with a radius of hundreds, even thousands of kilometers. Therefore, it is likely that a massive number of services are required to be fulfilled simultaneously, which poses great challenges to the service capability of the satellite network. UAVs are flexibly deployed and scheduled to assist network services in dedicated areas, such as computing-intensive services in IoT network and massive access services in crowded urban areas. However, the energy of UAVs is generally limited, and the cost and policy of deployment in large-scale should be carefully considered. Notice that different segments are originally designed to operate independently and work with different communication and networking protocols. Therefore, how to design an efficient network architecture suitable for various scenarios and services is a challenging problem, and has attracted much research attention.

There are mainly four basic SAGINs architectures. (A) Hybrid satellite-terrestrial relay network. Due to the influence of bad weather or obstacles, terrestrial relays can be deployed in this architecture to receive satellite signals and offset the masking effect in direct satellite-to-user communications; (B) Satellite-terrestrial backhaul network. In remote areas such as deserts and mountain areas, the deployment of backhaul links is not feasible with unacceptable cost. Satellite networks can be applied to serving as backhaul networks to provide seamless services; (C) Cognitive satellite-terrestrial network. The satellite-to-user link is considered as the primary link that can occupy the spectrum resources anytime. In the opposite, the secondary network (i.e., terrestrial wireless network) can only transmit data in unused spectrum; and (D) Cooperative satellite-terrestrial network. Different from the three architectures above where different network segments work separately, each segment in this cooperative architecture is integrated closely. Satellite networks can not only complement the terrestrial network coverage to achieve ubiquitous communications, but can also enable traffic offloading to relieve the terrestrial traffic burden.

2.2. Application scenarios

Since SAGINs can provide many benefits such as seamless coverage, cost-effective broadband network access, and flexible network management, it enables a series of novel services and scenarios in 6G networks. The distinctive characteristics of the services and scenarios in SAGINs should be clearly identified for efficient service fulfillment. The following describes three specific scenarios.

2.2.1. Automated driving and smart cars

For automated driving, networks can provide road condition information and vehicles security surveillance. However, in remote areas, vehicles (including ships) cannot access the Internet through conventional terrestrial networks due to the lack of network deployment or high access cost. LEO satellites can provide fast data transmission globally with relatively low costs.[6] In addition, a single carrier beam of the satellite has wide coverage, and accessing the satellite network significantly reduces handoffs compared with accessing terrestrial networks, and improves the service continuity. Furthermore, high network loads introduced by automated driving HD-map downloading can be efficiently reduced by caching and broadcasting the contents in UAVs.[17] In addition, through the combination of uploading from on-board sensors and satellite remote imaging, the regional road condition information and the driver's safety situation can be updated in real time.

2.2.2. Smart city

A smart city is a modern city with versatile sensors and appliances to collect data from citizens, vehicles, buildings, and environments, which should be managed efficiently to provide high quality services.[18] The smart city includes services such as remote treatment, intelligent traffic management, intelligent crime detection, and intelligent power supply. As a supplement, the space network and aerial network in SAGINs can assist terrestrial networks in data collection, on-demand monitoring, and data traffic offloading. For example, based on data collected by meteorological satellites, radars, and other sensors, the water level of rivers near the city can be predicted automatically and instructions can be carried out intelligently, thereby reducing the occurrence of disasters. In addition, UAVs and road side sensors can be utilized to collect data in terms of road status, traffic congestions, and accidents.[19]

2.2.3. Network coverage enhancement

Satellites, UAVs, and HAPs are robust to disasters such as earthquake and flood. Therefore, SAGINs can help disaster rescue and recovery when the terrestrial network infrastructure is damaged in disasters. For instance, UAVs or wreckers equipped with communication equipment can act as backup base stations for network accessing, which helps reduce the disaster loss as much as possible.[20] Besides, analyzing data from satellite images meteorological observatory is potential to predict potential environmental changes and disasters which is beneficial to take precautions. Furthermore, mobile network operators can dispatch UAVs and HAPs as supplement to relieve network traffic congestion in crowded areas.

2.3. Requirements of service-oriented SAGINs

Instead of focusing on network coverage, user access, and data transmission, the service-oriented SAGINs aim at providing guaranteed services, i.e., the goal of network design and management is from the perspective of network services. The network coverage, user access, and data transmission, etc. are considered as components of network service capability, and are scheduled on-demand. To achieve the goal, conventional network-centric and user-centric architecture and technologies are required to evolve to service-centric ones. In the sequel, the requirements of service-oriented SAGINs architecture and technologies are presented and discussed.

2.3.1. Diverse service requirement fulfillment

SAGINs are large-scale networks which cover a variety of scenarios, users, and environments. Therefore, there will be significant diversity in service requirements, which is a critical issue in on-demand service fulfillment. Some initial efforts have been made in defining new service requirements in 6G network. New 6G demands are proposed, i.e., extreme-high-speed and high-capacity communications, extreme coverage extension, extreme-low power consumption and cost reduction, extreme-low latency, extreme-reliable communication, and extreme-massive connectivity & sensing.[21] Considering these new demands, the three 5G service scenarios are reinforced into further-eMBB (feMBB), ultra-mMTC (umMTC), Extremely Reliable and Low Latency Communications (ERLLC), Mobile Broadband Reliable Low Latency Communications (MBRLLC), and massive-uRLLC (muRLLC).[7] Besides, there might be new service demands which cannot be handled by existing methods. For example, with the proliferation of AI, the demand of service intelligence level is essential in 6G era, which has not been defined currently. In future 6G SAGINs, how to meet the diverse service requirements should be investigated.

2.3.2. Dealing with customized services

In 5G era, the network services extend from mobile user services to vertical field services, including industry, automobile, education, government, etc. With the development of 6G network, it is envisioned that the network should support more customized and personalized services with guaranteed quality for any service from any user. This trend stems from multiple new features of the emerging 6G network. First, the network native AI has achieved ubiquitous intelligence across the whole network, which not only enables flexible and automatic network management, but also generates more personalized services due to deeper understanding of user needs. For example, the human bond communications transmit the features of a subject in the way humans perceive it in a holistic way, which involves all the five sense and thus require very personalized service guarantee. Second, with the development of SDN and cloud technologies, the network services and functions are separated from physical network protocols and devices, which facilitates customized service definition and deployment. For instance, most state-of-the-art cloud platforms support the deployment of customized services using software development kit or web APIs, such as Azure, AWS, Huawei Cloud, and so forth. Last, intent-based networking is capable of translating user intents into network configuration and management operations without clearly describing the service requirements by defining the performance indicators, such as data rate and latency. In this way, the intents can be seen as customized service requirements with unlimited small granularity, and thus, is very challenging. Worse still, in SAGINs, the number of users and user categories will be very large, further requiring more customized services.

2.3.3. On-demand network reconfigurability

The conventional network is a vertically integrated system where the control plane and data plane are closely coupled in a networking equipment, such as routers. Although such systems are designed to provide services to most customers, they lack flexibility and ability to innovate, since they are usually very expensive in expertise and time to develop and deploy new network services.[22] However, service-oriented networks should be capable of accommodating customized services with diverse requirements. To achieve this goal, the network is envisioned to reconfigure according to the service demand, which coincides with the concept of network as a service. SDN/ NFV-related technologies are extensively investigated to improve the network reconfigurability by virtualizing network functions as building blocks, which can be further connected and chained to create services. In service-oriented SAGINs, two critical issues in network reconfigurability are worth to be studied. First, due to the large network scale, it is more difficult for the control plane to achieve timely and reliable global coordination, which further degrades the performance of network reconfiguration. Second, since the UAVs are mostly controllable, it is important to reconfigure the positions and resources of the UAVs when composing the service function chain, which is essential to achieve higher service performance.

2.3.4. Multi-dimensional resource and mobility management

The core idea of service-oriented network is to accurately match the network capability and the service demands, in which efficiently allocating the network resources plays the most important role. In SAGINs, the end-to-end service fulfillment usually involves the network configuration and resource orchestration over multiple domains. This introduces more severe challenges in service fulfillment since in SAGINs the resources in different domain share very high heterogeneity. First, the resources are dedicated to specific domain. For example, the cellular systems exploit licensed FDD/TDD bands, V2X communication system owns dedicated 75 MHz band at 5.9 GHz, and satellite communications use very high-frequency bands, such as Ku/Ka bands. It is envisioned that the SAGINs system can achieve higher performance if the resources can be merged and used as a pool, yet the resulting interference could be an important issue to address. Second, the resources have a variety of limitations in different domains, which stem from the fact that the network devices are very distinctive, especially in large-scale SAGINs. For example, although the UAVs can serve as flexible flying base stations, they are usually very limited in energy. For LEO constellations, the capacity of link with laser communication is high, while the on-satellite computing and caching resources are often limited due to the satellite payload. Third, in SAGINs, the network mobility is more complex than the terrestrial network due to the controllable UAV movements and very high-speed satellites. Such a network mobility will bring dynamic service demands and complicates the service fulfillment, and thus should be carefully considered.

3. Service-oriented SAGINs management framework

Each network segment in SAGINs has its own pros and cons in coverage, capacities, and flexibility, as mentioned in Section

2. Through integrating the heterogeneous resources efficiently and economically, networks have the ability to support more versatile services with differentiated Quality of Service (QoS) requirements. However, fully exploiting the SAGINs service capability is significant. We design a novel service-oriented SAGINs management framework which efficiently manages the network services and orchestrates the network resources to achieve on-demand service fulfillment. As shown in Fig. 1, the framework is composed of three layers, i.e., SAGINs physical network layer, service analysis layer, and on-demand service layer. The SAGINs physical network layer is where the physical network operates, and offers a large amount of data containing information about the network status and the services, which are collected by ubiquitous sensing or digital twin technologies. The data is sent to the service analysis layer, which carries out the very important service analysis function. To achieve service-oriented network, it is significant to first clearly describe the diverse and customized service requirements and the implicit service demands in different network contexts. In addition, the SAGINs resources should be sensed and monitored in a very small granularity. For more details of service analysis, some existing research could be referred, such as intent-based networking,[23] digital/cyber twin,[24] and resource monitoring.[25] However, rather than digging into the details of this layer, this paper focuses more on the third layer, i.e., the on-demand service layer. The on-demand service layer takes the extracted information about the network services, resources, and contexts, and orchestrates the network service capability accordingly. Specifically, the network virtualization module considers the service requests and the associated network contexts, and composes the service by connecting individual virtual network functions. The cloud/edge synergy module coordinates the multi-dimensional resources and mobility, and outputs the service capability. Then, the network slicing module matches the service function chain with the network service capability and forms on-demand service slices, providing guaranteed and separated service fulfillment.

In this review paper, we focus on the on-demand service layer and related critical enabling technologies. Two categories of technologies are considered, i.e., heterogeneous resource orchestration technologies and cloud-edge synergy technologies, with the subcategories and organization in this paper shown in Fig. 2.

3.1. Heterogeneous resources orchestration

For the heterogeneous resource orchestration, network slicing, SDN, and NFV techniques have been investigated in SAGINs to achieve deep integration, efficient resource utilization, and flexible network operation.

Network slicing techniques abstract physical network resources of SAGINs into a shared virtual resource pool, to support services with multi-dimensional resource requirements. Existing studies on SAGINs network slicing are mainly concentrated on network architecture and resource allocation. For the network slicing-oriented architecture, the integration of heterogeneous networks is studied to improve the network interoperability and facilitate flexible end-to-end service slices creation over multiple network segments.[17,26–29] Resource allocation mainly focuses on the improvement of network performance such as network pressure reduction and capacity enhancement.[30–33] Based on a flexible and unified architecture, heterogenous resources can be allocated to individual services efficiently, which guarantees the service quality, reduces the network redundancy, and promotes network scalability.

For the resource allocation, the most critical issue is to efficiently orchestrate multi-dimensional heterogeneous resources. Some researchers solve this issue based on traditional optimization algorithms, which utilizes network information to

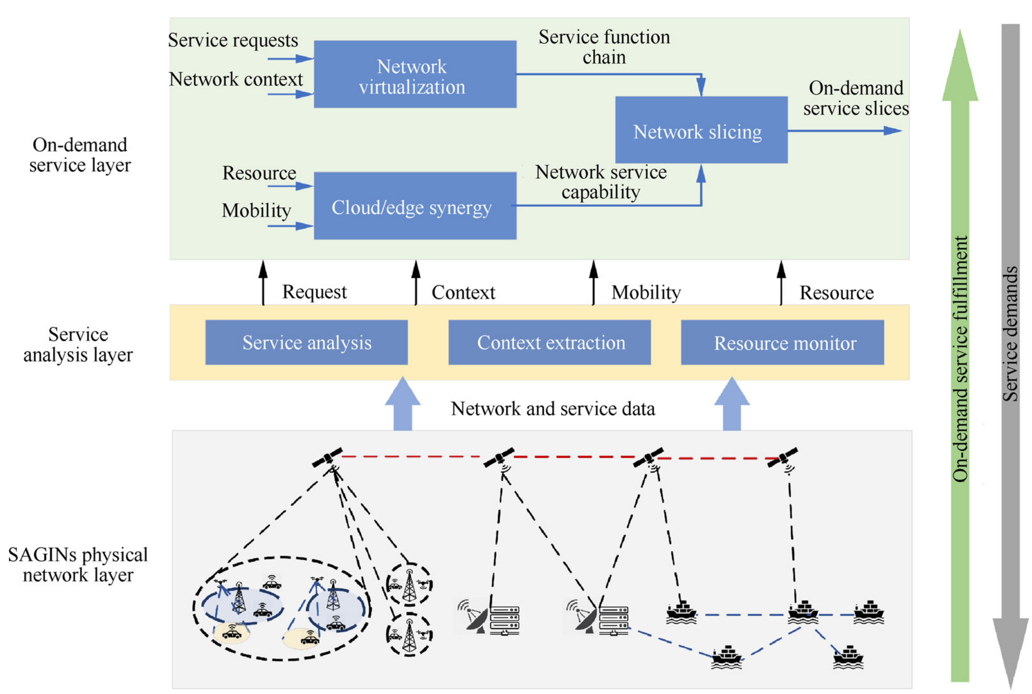

Fig. 1 SAGINs service management framework.

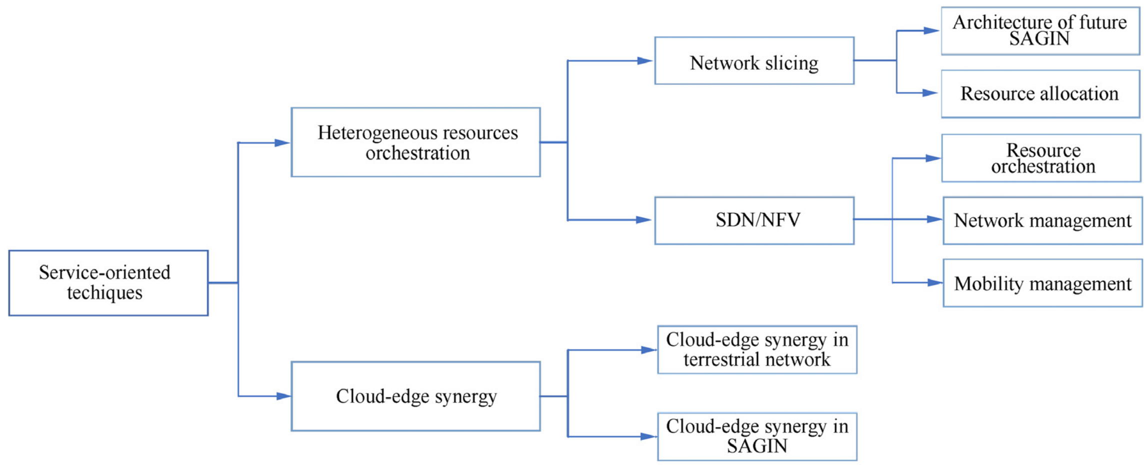

Fig. 2 Service-oriented techniques in 6G SAGINs.

execute efficient decision and control instructions.[34–35] However, it is revealed that the traditional optimization approaches are not feasible in a large-scale network since the computational complexity could be very high.[36,37] Recent works propose to use AI-based resource allocation methods in SAGINs, e.g., Reinforcement Learning (RL)-based methods, which can work when information on network model is unknown.[36,38] Since the vanilla RL methods are not suitable to a relatively large network scale, deep RL is proposed to cope with the problem of scalability by using neural networks.[39,40]

The existing research on SDN/NFV is mainly about three aspects. First, for the resource orchestration in SAGINs, there have been research on SDN-based global network resource orchestration.[10,41–43] In order to satisfy the diversified requirements in the service-oriented SAGINs, these studies take advantage of the global control function of the SDN architecture and calculate multiple transmission paths,[42] combining AI [43,44] and other methods to achieve efficient allocation of various network resources including communication resources, caching resources, and computing resources. Besides, there have also been studies on employing NFV to schedule network resources in specific scenarios,[45–47] e.g., SDN/NFV is analyzed for resource scheduling in the stadium.[48] Second, for the network management in SAGINs, considering the future service-oriented network requirements, SDN demonstrates efficiency on managing heterogeneous networks. Particularly, the characteristics of the separation of SDN control plane and data plane is beneficial to reduce the complexity of the overall network,[49] thereby improving the management efficiency.[50–52] In addition, under the SDN architecture, a NFV-based intelligent gateway is designed in a specific home network to enhance the QoS of Follow Me Services (FMS).[53] Last but not least, mobility management has become a critical issue in SAGINs, since the services bear the complex mobility features of different network segments. Towards the goal of continuous and seamless services, SDN and NFV can be used for mobility management, one idea is to use SDN to predict node movement and link interruption globally and combine mathematical models such as game theory to maximize terminal benefit.[41,54–56] By leveraging NFV techniques, multiple functions of the network can be bundled in a single physical device or a group of adjacent devices, which reduces the mobility management costs.

3.2. Cloud-edge synergy in SAGINs

SAGINs can significantly benefit from the cloud-edge synergy by cooperatively reaping the advantages of both heterogeneous cloud and edge resources. In SAGINs, the terrestrial network can provide a high-speed and stable network access but it is hard to cover remote areas due to the geography and economic reasons. As a complementary, the satellite network could offer seamless coverage such that the cloud servers could be constantly accessed, which provides powerful computing ability, but the delay is relatively large. In addition, edge servers can be deployed at HAPs/UAVs to form a flexible and cost-effective Mobile Edge Computing (MEC) system, offering low-latency services. Confronting the various requirements of the emerging services, it is indispensable to exploit the cloud-edge collaborative computing techniques in SAGINs to better expose network capabilities for on-demand service fulfillment. Through cloud-edge synergy, the network resources can be collaboratively orchestrated according to service requirements and characteristics. However, existing works on cloud-edge synergy in SAGINs have not been conducted well compared with terrestrial networks. There are two main research issues in cloud-edge synergy, the task offloading and the service deployment. Particularly, the task offloading is to divide a complex task into several sub-tasks, which can be processed by several servers cooperatively, and the service deployment is to decide the locations of severs or VNFs to maximize network efficiency. In Section 5, some latest studies containing the above two issues and other miscellaneous aspects will be introduced from the perspective of terrestrial networks[57–60] and SAGINs,[61–63] respectively, and the challenges of future cloud-edge synergy will also be presented.

4. Heterogeneous resource orchestration in SAGINs

In future 6G SAGINs, space, aerial, and terrestrial networks will be deeply integrated. It is forecasted that mobile data traffic will reach 77.5 exabytes per month by 2022,[64] and the number of IoT connected devices worldwide will be 25 billion in 2025.[65] How to utilize the limited but heterogeneous resources efficiently to accommodate the surging data has become an important research direction. Besides, future networks need to support vehicular communications, tactile Internet, smart cities, and other applications mentioned in Sections 1-2, which not only have a higher requirement in individual performance metrics, such as coverage, latency, reliability, throughput, and energy utilization, but also require a mixture of them. Conventional terrestrial networks can hardly fulfill these requirements at a reasonable cost. In this section, we investigate two promising techniques, i.e., network slicing and SDN/NFV, which play an important role in matching between SAGINs services and heterogeneous resources.

4.1. Network slicing

4.1.1. Overview of network slicing

Network slicing is considered as the virtual, modular, and isolated network technique built upon virtualized network resources abstracted from separated physical infrastructures.[29] The network slicing-based SAGINs architecture, as shown in Fig. 3, can be divided into three parts: service instance layer, network slice instance layer, and resource layer.[66] Firstly, resources of physical infrastructures are abstracted into a virtual resource pool by NFV and are defined as resource blocks by the network manager. When service requests of users arrive, they are described in the form of several network indexes (e.g., bandwidth, caching, computing, coverage, power, etc.). Then, the corresponding dedicated virtual resource blocks are mapped into physical resource layer by dedicated resource mapping mechanisms. Finally, a network slice is confirmed and available to support services. In addition, several critical features of the network slicing based system are summarized as follows.

(1) Automation: automation means that the network provides service supports or refines the local network performance automatically without manual interventions. When the service request of users arrives or its QoS updates, the network slicing-based SAGINs orchestrator allocates network resources automatically in the form of Virtual Network Function (VNF) instantiation, scaling up/down, scaling out/in, and cancellation.
(2) Stabilization: considering the highly dynamic network topologies, user mobility, and real-time service requirements in SAGINs, it is very likely to render the optimal slice configuration to sub-optimal or even unqualified. For example, in latency-critical services, medical accidents may occur due to unreliable network performance in remote surgery and unstable transmission latency or rate would affect Quality of Experience (QoE) in remote conference or video streaming services.
(3) Customization: in future service-oriented SAGINs, networks should have ability to support multi-dimensional requirements customized by users in latency, bandwidths, storage, etc. Focusing on differentiated services, networks can be realized based on the separation of control and data planes, which are programmable in policies, operations, and protocols.
(4) Isolation: to guarantee the QoS of different services and different users, there should be isolation between network slices. With network slicing techniques, isolation is assured through the usage of orchestration mechanisms and virtual resources mapping algorithms to avoid mutual interference.[67]

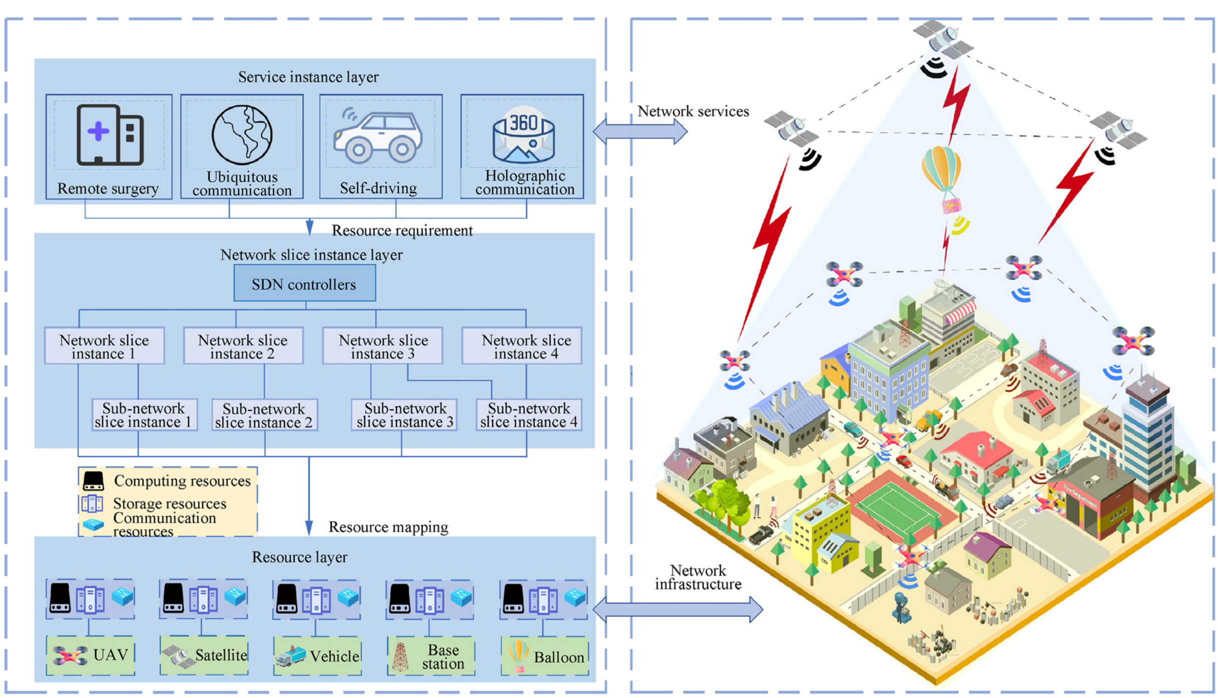

Fig. 3 Network slicing based on SAGINs architecture.

4.1.2. Related research

There have been extensive research works in SAGINs, which can be categorized into the investigation of network slicing-based architecture and resource allocation in network slicing.

(1) SAGINs architecture for network slicing

To reduce network complexity and facilitate flexible heterogeneous resource allocation, the SAGINs architecture for network slicing has been investigated.[26,29–31,33,68–71] A novel radio multiservice network adaptive architecture (NORMA) based on SDN/NFV and perception of QoE and QoS is presented, where SDN controller and coordinator are defined for intra-slicing resource allocation and resource competition in physical layer.[30] Based on network slicing in satellite-terrestrial networks, the broadband data transmission of users in remote areas is achieved by Internet access from satellite or terrestrial segments.[26] Ahmed et al. proposed an on-demand satellite-based network slicing framework, named on-demand adaptive network slicing, to support services with highly customizing in the form of an end-to-end function chain.[29] Focusing on QoS-restricted capacity maximization, a study on HAPs constellation is proposed.[33] In order to fulfill real-time processing capability in orbit, a flexible space-based edge computing system architecture and a resource management mechanism are proposed to reduce the delay and improve the throughput.[31] The network backhaul bandwidth pressure can be relieved and comprehensive benefits of the satellite network can be improved. A novel network architecture is proposed from the perspective of business model, where the networks are partitioned between infrastructure providers, mobile virtual network operators, service providers.[69] Network slicing is required to assure mobile virtual network operators to have a complete virtual network. To fulfill diverse content applications with multi-dimensional QoS, a service-oriented hierarchical soft slicing framework is designed, where one slice corresponds to one service and resources are reused opportunistically in inter-slice and intra-slice to obtain the multiplexing gain.[72] The results show that the throughput under the proposed strategy is better than the hard-slicing scheme.

(2) Resource allocation

Since space networks, aerial networks, and terrestrial networks have their own features, how to allocate the distinctive resources to different network slices appropriately is worth investigating. Resource allocation strategies in network slicing can be divided into two main categories: traditional optimization algorithms and AI-based approaches.

Traditional optimization-based resource allocation schemes exploit off-the-shelf optimization methods to orchestrate multi-dimensional resources given an optimization utility and several network constraints. One important research concentration is the maximization of network revenue from the perceptive of the number of users, power, latency, etc. Lyu et al. proposed a service-oriented resource slicing scheme and optimization for Space-Air-Ground Integrated Vehicular Network (SAGIVN) to maximize the system revenue in terms of the average amount of admitted service requests.[73] An online control framework is designed to make decisions on UAV dispatching, request admission, scheduling, and resource slicing for different services. Kasgari et al. proposed a Lyapunov drift-plus-penalty based resource method to minimize the power consumption in providing reliable and low-latency services.[74] For the traffic offloading issue in HAPs, an interference cancellation strategy is proposed based on spectrum sharing and power allocation schemes,[72,75,76] which efficiently saves spectrum resources while achieving the same performance. In these schemes, only unicasting is considered, and there is much room for further performance improvement. For the proactive HAPs broadcasting and vehicle caching, a network sharing and slicing strategy is used to minimize the data rate of the roadside units to guarantee QoS in vehicular networks.[17] In order to handle the traffic in complicated and dynamic network efficiently, a slicing strategy based on the Monte Carlo tree search is proposed in the context of fog radio access networks, which achieves better delay and throughput performance compared with benchmark algorithms.[77] Another research focuses on the total revenue of network operators in Cloud-RAN (C-RAN) from the perspective of user number and idle time.[78] An entity-oriented radio access network sharing strategy is presented in the cellular network to achieve efficient on-demand resource allocation, at the cost of reduction in network flexibility since deep packet inspections and information from Radio Access Network (RAN) are needed.[68] To address this issue, a dynamic and efficient RAN resource allocation mechanism, named application-oriented framework for RAN, is presented,[79] where the resource allocation is allowed in both central controller and applications, and the network revenue is shown to be 40% higher than network reservation scheme[68] and twice as good as Per-Base-station Reservation.[80] The reliability of V2V links from the perspective of Signal to Interference plus Noise Ratio (SINR) outage is analyzed.[81,82] Based on this, considering the resources of roadside units, an online algorithm named JRPSV is proposed,[35] which fulfills ultra-reliability and low-latency services in Vehicle-to-Vehicle (V2V) communications and complies with the queuing rules, and thus acquires asymptotically optimal network capacity. Some other researchers focus on maximizing the sum rate of users. Jiang et al. proposed a dynamic network slicing scheme which timely allocates resources based on the requested Service Level Agreement (SLA) to maximize sum rates of UEs considering user fairness.[83] A mixed integer non-linear programing problem for allocating sub-carriers and power resources in RAN is formulated to solve this optimization problem, and a low complexity resource allocation scheme is proposed, including the heuristic NS assignment and the many-to-one matching sub-carrier allocation algorithm.[34] To provide customized services in vehicular network, the spectrum allocation between BSs to maximize network throughput is transformed into a solvable form and solved through an Alternative Concave Search (ACS) algorithm.[84]

AI techniques have shown superiority in resource allocation in extremely complicated networks.[36,37,43,44] Particularly, a neural network-based resource allocation mechanism has been applied to satellite networks, aerial networks, and terrestrial networks, respectively,[85–87] and an efficient deep learning-based resource allocation in the integrated network of all three segments is presented.[43] A multi-agent RL-based resource allocation algorithm named LESS-DS is proposed to reduce the handoff cost and satisfy user QoS requirement in RAN slicing.[88] In addition, an RL-based slicing allocation scheme is proposed from the perspective of the physical infrastructure provider to maximize long-term network utility.[36] Traditional

RL methods are not applicable in heterogeneous resource allocation (especially in SAGINs) with a large action space, which may result in intolerable convergence time.[37,39] Based on conventional Deep Q-Learning (DQL), discrete normalized advantage functions are introduced into DQL to reduce computation complexity.[38]

4.2. SDN/NFV

4.2.1. Overview of SDN/NFV

To cope with more diverse services and higher QoS requirements in future service-oriented networks, the network hardware equipment may need to be updated frequently since the network infrastructure is only able to provide dedicated services, leading to intolerable costs for network operators. Besides, the integration of space, aerial, and ground segments is challenging due to their differentiation in protocols, resources, and network environment.

SDN/NFV-based network architecture is investigated, which shows a promising solution for building networks with on-demand service in Fig. 4 (redrew from 90).[43,52,89] When the service request, such as Artificial Intelligence of Things (AIoT), from the user is received, it will be described into a specific sequence of VNFs, named Service Function Chain (SFC), which are composed of corresponding virtual resource blocks (containing communication, computation and storage resources). Considering the real-time network states, the SFC can be mapped into physical network links, containing Space-to-Air (S2A) links, Air-to-Space (A2S) links, Air-to-Ground (A2G) links, Ground-to-Air (G2A) links, Space-to-Ground (S2G) links, and Ground-to-Space (G2S) links. Subsequently, these VNFs are embedded into physical resources by network operators, and then services will be supported by the physical network. Particularly, the network controlling and forwarding can be separated into three logical planes by SDN techniques as data plane, control plane, and application plane, where the data transmission can be carried out across planes by interfaces.[54] Due to the programmability of SDN, the network configuration can be adaptive to different requirements of services, thereby simplifying network management and upgrading and improving service diversity and network flexibility.[90] With NFV, heterogeneous network resources can be isolated from each other, which is convenient for the management and transplantation with on-demand services. Besides, network functions can be decoupled by NFV techniques from its dedicated hardware devices to realize the flexible deployment of network functions.[91–93] In SAGINs, SDN/NFV technologies are mainly exploited in three different ways.

SDN/NFV in satellite networks: Considering that satellites have on-board processing and switching capabilities, it is convenient to employ SDN-enabled switches that are responsible for executing commands from the controller. An SDN-based satellite network architecture, named SDSN, is proposed.[43] In the SDSN, the data plane is composed of satellite switches, which are responsible for simple stream-based packet forwarding, the control plane is composed of controllers located in the ground station, and control information is forwarded through Geostationary Earth Orbit (GEO). With SDN techniques, better performance can be achieved with guaranteed diversified user requirements and effective satellite resource utilization;

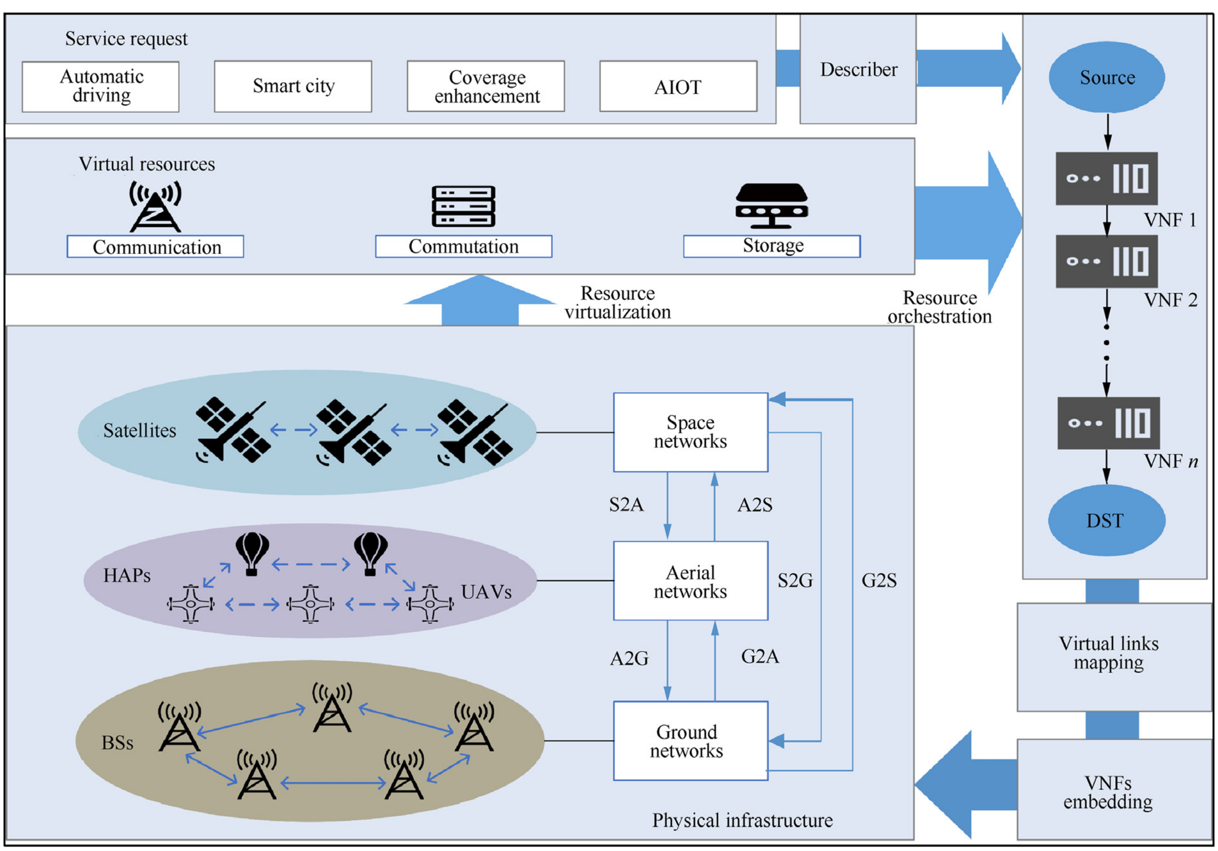

Fig. 4 SDN/NFV-based network architecture in SAGINs.

SDN/NFV in terrestrial networks: With SDN/NFV techniques, the programmability and reconfiguration capabilities of each network component can be improved in terrestrial network. For example, SDN-based architecture in terrestrial networks can make gateway switching and stream-based data forwarding more flexible and intelligent.[42]

SDN/NFV in backhaul networks: In SAGINs, there might be a significant increase in the number of routing paths due to the involvement of space networks and aerial networks in routing. Packet routing path selection and how to avoid congestion or idleness in path selection is a key issue to improve the overall network performance. The central SDN controller can obtain information of the entire network and make a globally optimal path selection according to the overall network status to reduce unnecessary link switching. However, in SAGINs, the information collection itself might be extremely inefficient due to dynamic topology, unstable links, and long propagation delays.

4.2.2. Related research

SDN/NFV has been widely used to enhance the resource orchestration, network management and mobility management in the SAGINs, which can be summarized as follows.

(1) Resource orchestration

(A) SDN-based overall resource orchestration

To improve the resource utilization, Wu et al. proposed an SDN-based satellite network architecture where the data plane is composed of satellite switches, and the control plane is composed of controllers located in the ground station.[43] To guarantee the flexibility, openness, interoperability, and evolvability of the network, a SDN-based aerial backbone network architecture is proposed.[42] Particularly, the segmented routing by SDN is designed to improve the transmission reliability and bandwidth utilization by balancing network traffic load among multiple reliable transmission paths. For the transmission efficiency of both control and data planes, complex network management operations through SDN are transferred to a cloud-based synchronous data network controller, where the computational burden of mobile nodes and drones is reduced and the utilization of wireless channel resources can be improved.[10] Different from the network with independent network segments, an SDN-based SAGIVN is investigated and a hybrid and layered control architecture is designed, where an AI-based solution is proposed to promote effective network slicing, mobility management, and content caching, realizing flexible, reliable, and scalable network resource managements.[43]

(B) NFV-based resource orchestration

In addition to the abovementioned research on the overall network design based on SDN, there are also studies on the use of NFV to implement resource orchestration for specific networks. NFV is used to build a configurable and scalable network and combines bandwidth resources to design a resource allocation strategy for high-density scenarios, e.g., stadiums.[46] An optimal resource allocation problem in multi-domain NFV system is investigated with the competition of multiple NFV service providers that owns different types of resources including virtual computing, storage and network resources, and non-virtual resources.[47]

(2) Network management

Due to the separation of control and data planes and the programmability, SDN networks can achieve more flexible and efficient network management, which exactly conforms to the on-demand network configurability requirement of service-oriented SAGINs.

(A) SDN-based overall network management

Bao et al. proposed an SDN-based satellite network architecture, termed as OpenSAN, which decouples the data and control planes.[49] The data plane is formed by routers and multi-layer satellites around the world, which is responsible for data forwarding for different services. The control plane is composed of GEO satellites, which reduces the number of required ground stations and reduces the complexity; the management plane is responsible for mobility management. Through this layered network, the complexity of the network management can be reduced and the flexibility of the network can be improved to support advanced technologies in future network. Zhang et al. also proposed a SDN-based SAGINs architecture.[86] To improve the overall network management efficiency, a new operationally responsive space (ORS) satellite networking scheme based on SDN is proposed, which adopts a lightweight architecture consisting of a physical layer, a control layer, and an application layer.[50] Specifically, the physical layer defines the standards of network equipment, the control layer maintains the network status centrally by communicating with the physical layer through OpenFlow, and the application layer provides users with network services, such as data transmission protocols, satellite networking management protocols, etc. Secinti et al. proposed an aerial network management protocol based on SDN where each drone is deployed as an SDN switch which runs under the instructions issued by the central controller and calculates the drone's trajectory.[51] Results show that this protocol outperforms benchmark protocols in controlling end-to-end interrupts, and the end-to-end delay is also reduced. In addition, to address the dynamic network topology and stochastic real-time service requirements in satellite networks, a network management based on Viterbi algorithm is proposed, where the service function chain (SFC) mapping process is established as a Markov model and the VNF of the overloaded node is considered.[52] This work reveals that better network management performance can be achieved by migrating resources to other idle nodes.

(B) NFV-based network management

For the network management, the network deployment and intelligent management based on NFV and the NFV architecture will be compatible with legacy networks.[94] NFV can also be integrated into specific networks to reduce the use of special equipment and facilitate network upgrades.[53]

(3) Mobility management

The mobility of network nodes, especially LEO satellites which have high-speed relative movement with earth, results in a dynamic network topology in SAGINs. Due to the constant change of node positions, the connection status between the nodes and networking performance are also changing, which leads to significant issues such as performance degradation, increasing overhead, and network outage.

(A) SDN-based mobility management

SDN-based networks have the ability to collect information and regulate global resources. To achieve effective mobility management, a prototype system based on SDN is designed.[39] The system predicts future events (e.g., link down) through the trajectory of aerial nodes, and then reduces the impact of network events through SDN-based optimization, thereby improving the aerial network performance and availability.[54,55] A seamless handover mechanism based on SDN is proposed, where the ground controller is responsible for the collection of overall network status information and the LEO satellite is only responsible for the exchange in the data plane.[41] Particularly, the user terminal can connect to multiple satellites at the same time, where the link can be changed according to the signal strength to reduce the impact of satellite and user movement. Besides, a satellite resource sharing game model based on the bipartite graph framework is designed, which selects suitable satellites and frequency bands for mobile terminals to maximize the terminal revenue, and designs random access algorithms based on the goal of maximizing user space (i.e., the feasible area of the subsequent access terminals).[56]

(B) NFV-based mobility management

By NFV technologies, the life cycle of network functions and its constituent resources can be managed, which is complementary to SDN. Prados-Garzon et al. proposed a theoretical framework to evaluate the performance of the virtualized Mobility Management Entity computing system (vMME) based on NFV.[95] Particularly, the virtual part hypervisor of the vMME system combines three logical components: the front end, management entity service logic, and State Database (SDB). By predicting the number of given user equipment and MTC equipment, the communication network can be updated in advance to cope with the high mobility. An Evolved virtualization-based Packet Core (EPC) entity is presented, which is divided into groups according to the control traffic where entities that exchange large amounts of traffic are grouped into the same group, i.e., mobility management entity and the home user server entity grouping.[91] In addition, the authentication and authorization processes are executed internally without the need for network data transmission, which further reduces the impact of mobility.

5. Cloud-edge synergy

5.1. Overview

Cloud servers can provide abundant cloud resources accurately according to service requests to guarantee the user QoS. However, a soaring number of time-sensitive services require the rigorous latency, and the cloud-based technologies can hardly ensure it due to the long transmission paths from the users to the cloud servers. To address this issue, the MEC is proposed to decentralize cloud services to the network edge to achieve lower latency,[96,97] where MEC servers can be deployed in access networks such as on base stations.[98,99] Particularly, the MEC can deploy services in advance, which shortens the distance between resources and users.[100] By integrating cloud and edge resources, the energy consumption of terminals, core network traffics, and the network security and scalability can be better addressed. In Fig. 5, a heterogeneous and multi-layered cloud-edge synergy architecture is demonstrated for service-oriented SAGINs. From the perspective of on-demand services, the operations and functions of the cloud-edge synergy architecture is described as follows. The edge servers and controllers can be deployed at BSs, UAVs, LEO satellites, etc. Computation-intensive tasks can be offloaded from user devices to the edge servers for higher energy efficiency and lower computing delay. Furthermore, service caching schemes can be applied to cache the application services and related databases such that the user tasks which require these services can be executed.[101] To enable these functionalities, the SAGINs edge controllers should be carefully deployed to facilitate efficient data collection, and the UAV networking and trajectory should be optimized to provide better link performance and network coverage. The cloud servers can be deployed in terrestrial data centers, which can be accessed even from rural areas via SAGINs global coverage. Through cloud-edge synergy, the service performance can be optimized by coordinating the advantages and disadvantages of both edge and cloud.[102] However, due to the large network scale and high mobility, it is important to design efficient cloud-edge synergy mechanisms for timely and performance guaranteed service fulfillment.

5.2. Related research

Cloud-edge synergy techniques are originally designed for terrestrial networks, and are recently applied by some researchers, into SAGINs. In the following, we begin with some latest research works on terrestrial cloud-edge synergy, and then review the cloud-edge synergy in SAGINs.

5.2.1. Cloud-edge synergy in terrestrial network

The most important function of cloud-edge synergy is to deliver network services in different locations according to the service capabilities, in order to achieve service fulfillment in terms of different metrics, such as delay, traffic pressure, security, etc. Most of the work focuses on the optimization of service latency, since the most important motivation of introducing edge computing is to reduce the overall latency. To reduce the task processing latency, Kai et al. proposed a hierarchical computing scheme with pipeline offloading, where the task computing and offloading execute simultaneously.[58] By using the multi-layer servers and the cloud data center, a RL-based computing strategy is designed, which achieves a lower service latency compared with the pure cloud computing and edge computing strategies.[103] In addition, a Federated Learning (FL) framework is designed in MEC systems for the model training of deep RL, based on which, the dynamic resources

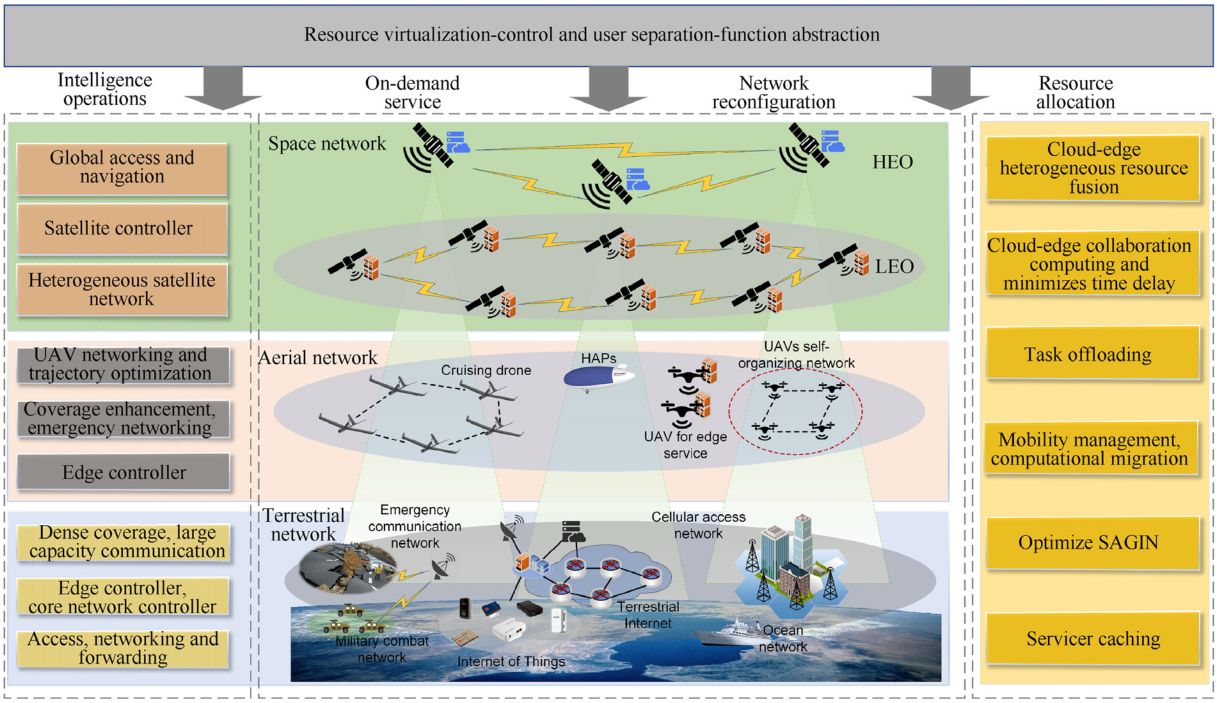

Fig. 5 Cloud-edge synergy architecture in service-oriented SAGINs.

(e.g., communication, storage, and computing) optimization strategy is deployed.[104] Service deployment enables VNF to instantiate in separate nodes such as the edge and cloud servers, which reduces service latency and improves network profit. Some existing research works have been conducted in relatively simple scenarios with only cloud servers[59] or multi-access edge computing servers. [60,105–107] An integrated architecture, which considers both the cloud and the edge and is compliant with the ETSI,[108] is presented to minimize the sum of latency by balancing loads among mobile edge platforms and edge virtualization computing infrastructure, and a new Tabu Search-based algorithm is presented to find near-optimal solutions.[109] Miscellaneous functions of cloud-edge synergy have also been studied. To sufficiently use the full computing capacity of BSs and cloud, a task splitting strategy is proposed to divide the task into several subtasks, and then the divided subtasks are processed in the edge nodes and cloud severs seperatively.[57] Conventional Machine Learning (ML) needs user data to train models, which may result in user privacy leaking. FL enables the ML model to be trained locally and upload the trained model data instead of raw data, which protects user privacy and saves network resources.[110] An asynchronous FL mechanism is proposed to preserve privacy in MEC systems, and the network node can join or leave flexibly, which is suitable in highly dynamic SAGIN scenarios.[111] Shao et al. proposed a learning-based framework with an integration of stochastic simulation, neural network, and genetic algorithm, to execute edge server deployment in BSs to minimize demands-weighted distance. [112] Simulations have shown that the algorithm is better than the Random Fit algorithm, Top-first algorithm, and the K-means algorithm.

5.2.2. Cloud-edge synergy in SAGINs

In SAGINs, the cloud-edge synergy can be further extended since the user can access the edge and cloud through ubiquitous coverage. The existing works mainly focus on two research directions, i.e., edge caching/computing through UAV networks and collaborative cloud-edge task scheduling via SAGINs.

Due to the flexible networking capability, UAV networks can provide on-demand caching and computing resources by controlling the trajectory and orchestrating the resources of UAVs. A UAV caching scheme is studied by Chen et al., where QoE is the main concern when the deployment of UAVs is designed.[113] A ML framework named echo state networks is proposed to predict the users' content request distribution and mobility pattern, based on which the optimal user-UAV association is derived. To make an efficient computation task offloading decision considering the constrained energy at UAVs, the dynamic offloading is formulated as a Constrained Markov Decision Process (CMDP), which is solved by the linear programming. The lowest task processing delay is achieved comparing to the random scheme which executes actions randomly and the greedy scheme that selects BS prior to offload more tasks.[114] Furthermore, a comprehensive scheduling of association control, computation task allocation, transmission power and bandwidth allocation, UAV computation resource, and deployment position is conducted to minimize the maximum computation delay among IoT devices.[115] Considering the migration cost due to the user mobility, a joint task offloading and migration scheme is proposed in RL-based mobility-aware MEC network to achieve the maximum system revenue.[116]

Owing to the seamless coverage and abundant computing/caching resources provided by SAGINs, the cloud-edge synergy can be further enhanced, even in the areas where it is difficult or cost-ineffective to deploy terrestrial BSs. However, the high mobility, heterogeneous resources, and constrained interoperability will make the collaborative cloud-edge task scheduling more challenging in SAGINs. Due to the wide broadcasting and coverage of satellites, MEC services can be provided by satellite networks at a low cost. To provide the low latency ubiquitous services in SAGINs, an approach with joint request dispatching and service placement is proposed, which shows better performance comparing to the greedy algorithm in terms of the service user ratio and average hop count.[63] Zhou et al. proposed an SAGIN-based bidirectional mission offloading framework, which remedies the deficiency of terrestrial networks in coverage and space-aerial networks in power and other resources, to achieve substantial network gain in reliability and cost reduction.[117] The latency and energy optimization problem in SAT-IoT networks is formulated as a dynamic mixed-integer programming problem, where a two-stage optimization is conducted, including the computing and communication resource allocation and the joint optimization of user association and offloading, which are solved by the Lagrange multiplier method and the deep RL-based method, respectively.[62] Learning-based models can adjust resource orchestration and network configuration to adapt to real-time network status. Considering the energy constraint and computation constraint of remote devices, an SAGIN edge/cloud computing architecture is proposed for offloading the computation-intensive applications, where UAVs provide the near-user edge computing and satellites provide the cloud computing.[102] Based on this framework, a learning-based computational offloading method is further proposed, which dynamically learns the optimal offloading strategy to minimize the total task delay, energy consumption, and server usage costs. To make task scheduling decisions online, a deep risk-sensitive RL-based algorithm is proposed to minimize the long-term average delay of IoT services in SAGINs, where the low-latency and extended service lifespan of IoT services can be achieved.[118] Considering a cache-enabled satellite-UAV-vehicle integrated network where the GEO satellite serves as the cloud server and the UAVs are deployed as the edge caching servers, an energy-aware coded caching strategy is proposed to provide more multicast opportunities and significantly reduce the total energy consumption.[119]

Although the cloud-edge synergy in SAGINs can flexibly deploy services in multi-tier networks, considering different server capacity and the dynamic topology, it is challenging to perform the global task scheduling. In addition, the placement of servers also affects the service performance. The location, quantity, and density of cloud/edge servers should be optimized.[103] However, both users and access nodes/servers in SAGINs are dynamic, which may render solution optimal to sub-optimal. Therefore, for cloud-edge synergy-based SAGINs, how to deploy edge cloud servers, whether to offload tasks to the cloud or the edge, and who manages the offloading of tasks are issues that must be resolved in the future.

6. Future research directions

In this section, some potential future research directions in service-oriented SAGINs are discussed.

6.1. Service definition and identification

As discussed in Section 2.3, the 6G service-oriented network faces the challenge of handling diverse, long-tail, and customized services with distinctive requirements. Worse still, since the network environment and contexts also significantly affect the service requirements, the wide variety of network scenarios is a critical issue in on-demand service provisioning. Although some initial works have been conducted to address the dynamic scenario and service, e.g., a graph neural network is proposed to address the scalable radio resource management that is difficult for convention deep neural networks with fixed number of inputs and outputs,[120] a general framework and methodologies for service-oriented network is still missing. Therefore, the service definition and identification should be first addressed in service-oriented networks. Generally speaking, two important functions need to be investigated, i.e., service description framework and service sensing. Service description should provide a generalized framework to represent the features and requirements of the network services, which serves as the theoretical foundation for service identification and on-demand service fulfillment. Currently, the service description mainly exploits classification-based methods, such as the three use cases in 5G network and the model-based service orchestration in off-the-shelf network management systems.[121] As this method lacks accuracy and flexibility to describe 6G customized services, a more general method is thus required. Potential solutions can borrow the ideas from the field of ontology engineering. An ontology is defined as "a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts".[122] By using ontology engineering, it is feasible to explicitly represent the network services as a set of concepts, for example, the environment, network context, requirement, the interrelation, taxonomy, and rules between them accordingly. Currently, the ontology-based network management has been investigated,[123–125] however, application of ontology in defining and fulfilling network services needs to be further studied.

A service needs to be sensed before identified, which involves the information sensing of service features and requirements, network conditions, environments, etc., and the transition of the information from the physical world to the digital world. There are two potential solutions, i.e., joint sensing-communication technique and digital/cyber twin. Sensor fusion, which combines network sensing (the detection of communication environment using radio signals) and other ubiquitously deployed sensors,[126] will provide complete scenario data for 6G service-oriented network.[127] With the network context awareness, the communication services could be fulfilled more efficiently with guaranteed performance. Recently, a joint sensing-communication cooperative sensing UAV network is proposed where UAV equipment can generate a sensing beam and a communication beam orthogonally to enhance the spectrum reuse.[128] Cybertwin is the digital twin in the communication network systems, which is the digital representation of humans, things, and processes in the virtual cyberspace.[24] Cybertwin can offer abundant information required in identifying the network services. For example, the end user can connect its cybertwin with the required service, and the cybertwin will integrate the network context and abundant historical data to evaluate the accurate service requirements for the time being and in the future. However, the research on either technique is still in its infant stage, especially when applied in service identification.

6.2. AI-based networking paradigms

Pervasive network intelligence will be one distinctive feature of 6G network, which not only increases the service diversity, but also more importantly, offers higher capability to handle on-demand services in a more flexible, efficient, adaptive, automatic, and cost-effective way. Therefore, developing advanced AI-based service-fulfillment technologies is of great significance. Specifically, such technologies have the potential to address the following two important issues in service-oriented network, i.e., massive data exploitation by big data techniques and service identification by intent-based networking:

Massive network data mining - The 6G SAGINs generate massive network data from network nodes, operations, and ubiquitous deployed sensors, which if appropriately exploited, will significantly enhance the network and service performance. First, the data contains valuable information for service identification, such as the service type, requirements, user preference, and network contexts. A big data and ML-based network context utilization process is described, which includes context acquisition, context modelling, context reasoning, and context distribution.[129] Second, network data can facilitate automatic service management. The life cycle of services includes the service monitoring, fault detection, heal, and scale functions, which can result in very high cost in time and money if handled by humans. Therefore, through service data monitoring and analysis, the abnormal events can be inferred and coped with in a fully automatic manner, which significantly saves the cost and enhances the service performance. With the fast development of big data and ML, powerful big data tools have been emerging recently, such as distributed storage and process, data analytics, data visualization, etc. However, utilizing big data techniques in service-oriented network needs further investigation.

Intent-based networking - Advances in neural language understanding and deep learning algorithms have facilitated the development of knowledge which can be used to covert user queries in a given language (i.e., intents) into a structured representation that can be processed automated services.[130] Through intent-based techniques, the network service, which might be ambiently expressed by users' intent, can be explicitly and accurately identified. A three-layer intent-based network management system is proposed.[131] Within the architecture, the intent layer is responsible for translating the service from user intents in network layer to key performance indicators, service level agreements, processes, and targets in business layer, which is an intent-based automation process with less human workforce. However, the intent-based techniques have not been thoroughly studied, and their ability to identify the wide variety of SAGINs services remains doubtful and needs further investigation.

6.3. Security

For the proposed service-oriented SAGINs architecture, new networking security issues will be exposed due to its inherent networking features. Considering the open connectivity and dynamic topology, SAGINs are vulnerable to security threats for the random access of multi-mode terminals, including malicious interference, masquerade attack, passive eavesdrop, massage intercept, etc.[92,93] Although the isolation is generally considered between different network slices, the vulnerability could penetrate through tenants with different service requirements, due to different security isolation levels among slices. With the limited resources at satellite and the wide coverage of satellite beams, malicious attacks are easy to hide and difficult to eliminate. Recently, the physical layer security as a lightweight secure approach has shown promising solutions for secure transmission of such massive connections in SAGINs, which is a key-free security approach based on information-theoretic security.[132] However, the security is guaranteed separately between heterogeneous multi-tier network, directing a research issue of compatible security approach in SAGINs, e.g., symbiotic security. In addition, considering the centralized control mode of SDN, the SDN controller charges the control function and status information of the overall network, while the SDN controller is easy to be targeted by hackers, and the inherent security is flawed to the network. Particularly, an attacker can forge many packets that the switch cannot process according to the characteristics of the protocol, forcing the switch to request many processing by the controller, thus the communication and computing resources are occupied. Block-chain technology has the characteristics of decentralization, non-tampering and anonymity, which realizes information recording without a trusted third party, and can guarantee the security of data in the process of collection, transmission, storage, and calculation. Some documents point out that blockchain technology can guarantee the data and privacy security of MEC and SAGINs.[133] Blockchain can provide security for MEC and SAGINs, but may cause delays due to sequential validation requests and consume large amounts of storage resources.[134] The cloud server happens to have computing and storage capabilities, which can solve the computing delay and storage problems caused by blockchain. Therefore, it is promising to use blockchain technology to guarantee the security of SAGINs based on cloud-edge synergy architecture.[135]

7. Conclusions

In this paper, a comprehensive study of 6G SAGINs has been conducted from a service-oriented perspective. The requirements of service-oriented network have been presented based on what the service-oriented SAGINs architecture has been proposed. Two categories of enabling key technologies, heterogeneous resource orchestration technologies and cloud-edge synergy technologies, have been surveyed and discussed. The future research directions are finally discussed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (No. 2020YFB1807700).

References

1. 3GPP. Study on New Radio (NR) access technology physical layer aspects. 2017. Standard No.: TR 38.804.
2. ITU-R M. [IMT-2020.TECH PERF REQ]—minimum requirements related to technical performance for IMT-2020 radio interface(s). 2017. Standard No.: M.2410-0 (11/2017).
3. Wu Q, Ding G, Xu Y, et al. Cognitive Internet of Things: A new paradigm beyond connection. *IEEE Internet Things J* 2014;**1**(2):129–43.
4. Shen X, Cheng N, Zhou H, et al. Space-air-ground Integrated Networks: Review and prospect. *Chin J Internet Things* 2021;**4**(3):3–19.
5. Niu Z, Shen XS, Zhang Q, et al. Space-air-ground integrated vehicular network for connected and automated vehicles: challenges and solutions. *Intell Converged Netw* 2020;**1**(2):142–69.
6. Qu Z, Zhang G, Cao H, et al. LEO satellite constellation for Internet of Things. *IEEE Access* 2017;**5**:18391–401.
7. You XH, Wang CX, Huang J, et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Sci China Inf Sci* 2020;**64**(1):1–74.
8. Mohammed A, Mehmood A, Pavlidou FN, et al. The role of high-altitude platforms (HAPs) in the global wireless connectivity. *Proc IEEE* 2011;**99**(11):1939–53.
9. Xiong F, Li A, Wang H, et al. An SDN-MQTT based communication system for battlefield UAV swarms. *IEEE Commun Mag* 2019;**57**(8):41–7.
10. Yuan ZH, Huang XW, Sun LL, et al. Software defined mobile sensor network for micro UAV swarm. *2016 IEEE International Conference on Control and Robotics Engineering (ICCRE)*; 2016 Apr 2-4; Singapore. Piscataway: IEEE Press; 2016.p.1–4.
11. Cheng N, Quan W, Shi W, et al. A comprehensive simulation platform for space-air-ground integrated network. *IEEE Wirel Commun* 2020;**27**(1):178–85.
12. Shi W, Li J, Cheng N, et al. Multi-drone 3-D trajectory planning and scheduling in drone-assisted radio access networks. *IEEE Trans Veh Technol* 2019;**68**(8):8145–58.
13. Ma T, Zhou HB, Qian B, et al. A large-scale clustering and 3D trajectory optimization approach for UAV swarms. *Sci China Inf Sci* 2021;**64**(4):1–16.
14. Shi W, Li J, Xu W, et al. Multiple drone-cell deployment analyses and optimization in drone assisted radio access networks. *IEEE Access* 2018;**6**:12518–29.
15. Cheng N, Xu W, Shi W, et al. Air-ground integrated mobile edge networks: architecture, challenges, and opportunities. *IEEE Commun Mag* 2018;**56**(8):26–32.
16. Qiu J, Grace D, Ding G, et al. Air-ground heterogeneous networks for 5G and beyond via integrating high and low altitude platforms. *IEEE Wirel Commun* 2019;**26**(6):140–8.
17. Zhang S, Quan W, Li J, et al. Air-ground integrated vehicular network slicing with content pushing and caching. *IEEE J Sel Areas Commun* 2018;**36**(9):2114–27.
18. Connected vehicles in smart cities: The future of transportation [Internet]. Available form: https://interestingengineering.com/.
19. Ding G, Wu Q, Zhang L, et al. An amateur drone surveillance system based on the cognitive Internet of Things. *IEEE Commun Mag* 2018;**56**(1):29–35.
20. Casoni M, Grazia CA, Klapez M, et al. Integration of satellite and LTE for disaster recovery. *IEEE Commun Mag* 2015;**53**(3):47–53.
21. Nakamura T. 5G evolution and 6G. *2020 IEEE symposium on VLSI technology*; 2020 Jun 16-19; Honolulu, USA. Piscataway: IEEE Press; 2020. p. 1–5.
22. Zilberman N, Watts PM, Rotsos C, et al. Reconfigurable network systems and software-defined networking. *Proc IEEE* 2015;**103**(7):1102–24.
23. Pang L, Yang C, Chen D, et al. A survey on intent-driven networks. *IEEE Access* 2020;**8**:22862–73.
24. Yu Q, Ren J, Fu Y, et al. Cybertwin: An origin of next generation network architecture. *IEEE Wirel Commun* 2019;**26**(6):111–7.
25. He HL, Ren PY, Du QH, et al. Traffic-aware overload control scheme in 5G ultra-dense M2M networks. *Trans Emerging Tel Tech* 2017;**28**(9):e3146.
26. de Cola T, Bisio I. QoS optimisation of eMBB services in converged 5G-satellite networks. *IEEE Trans Veh Technol* 2020;**69**(10):12098–110.
27. Höyhtyä M, Ojanperä T, Mäkelä J, et al. Integrated 5G satellite-terrestrial systems: Use cases for road safety and autonomous ships. 23rd *Ka and Broadband Communications Conference*; 2017. p. 1-9
28. Hendaoui S, Mannai A, Zangar N. Cognitive CQI/5QI based scheme for software defined 5G hybrid satellite-terrestrial network: Slicing for ultra reliability and video congestion offloading. *2020 international symposium on networks, computers and communications (ISNCC)*; 2020 Oct 20-22; Montreal, Canada. Piscataway: IEEE Press; 2020. p.1–7.
29. Ahmed T, Alleg A, Ferrus R, et al. On-demand network slicing using SDN/NFV-enabled satellite ground segment systems. *2018 4th IEEE conference on network softwarization and workshops (NetSoft)*; 2018 Jun 25-29; Montreal, Canada. Piscataway: IEEE Press; 2018.p.242–6.
30. Pavon IL, Sánchez JR, Colazzo A. 5G NORMA network architecture—intermediate report. EU H2020 5G NORMA. *Tech Rep Deliverable D* 2017.
31. Cao SZ, Wei JY, Han H, et al. Space edge cloud enabling network slicing for 5G satellite network. *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*; 2019 Jun 24-28; Tangier, Morocco. Piscataway: IEEE Press; 2019.p.787–792.
32. NGMA Alliance, Description of network slicing concept, 2016.
33. Dong F, Han H, Gong X, et al. A constellation design methodology based on QoS and user demand in high-altitude platform broadband networks. *IEEE Trans Multimed* 2016;**18**(12):2384–97.
34. Zhao ZL, Bai J, Tie ZZ, et al. Network slicing technique assisted resource allocation in small-cell networks. 2019 IEEE 19th International Conference on Communication Technology (ICCT); 2019 Oct 16-19; Xi'an, China. Piscataway: IEEE Press; 2019.p.843–8.
35. Chen Y, Wang Y, Liu M, et al. Network slicing enabled resource management for service-oriented ultra-reliable and low-latency vehicular networks. *IEEE Trans Veh Technol* 2020;**69**(7):7847–62.
36. Li R, Zhao Z, Sun Q, et al. Deep reinforcement learning for resource management in network slicing. *IEEE Access* 2018;**6**:74429–41.
37. Shen X, Gao J, Wu W, et al. AI-assisted network-slicing based next-generation wireless networks. *IEEE Open J Veh Technol* 2020;**1**:45–66.
38. Sciancalepore V, Costa-Perez X, Banchs A. RL-NSB: Reinforcement learning-based 5G network slice broker. *IEEE/ACM Trans Netw* 2019;**27**(4):1543–57.
39. Qi C, Hua Y, Li R, et al. Deep reinforcement learning with discrete normalized advantage functions for resource management in network slicing. *IEEE Commun Lett* 2019;**23**(8):1337–41.
40. Chen X, Zhang H, Wu C, et al. Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning. *IEEE Internet Things J* 2019;**6**(3):4005–18.
41. Yang B, Wu Y, Chu X, et al. Seamless handover in software-defined satellite networking. *IEEE Commun Lett* 2016;**20**(9):1768–71.
42. Chen K, Zhao S, Lv N, et al. Segment routing based traffic scheduling for the software-defined airborne backbone network. *IEEE Access* 2019;**7**:106162–78.

43. Wu H, Chen J, Zhou C, et al. Resource management in space-air-ground integrated vehicular networks: SDN control and AI algorithm design. *IEEE Wirel Commun* 2020;**27**(6):52–60.
44. Faraci G, Grasso C, Schembra G. Reinforcement-learning for management of a 5G network slice extension with UAVs. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*; 2019 Apr 29 - May 2; Paris, France. Piscataway: IEEE Press; 2019.p.732–7.
45. Gil Herrera J, Botero JF. Resource allocation in NFV: A comprehensive survey. *IEEE Trans Netw Serv Manag* 2016;**13**(3):518–32.
46. Pedreno-Manresa J-J, Khodashenas PS, Siddiqui MS, et al. On the need of joint bandwidth and NFV resource orchestration: a realistic 5G access network use case. *IEEE Commun Lett* 2018;**22**(1):145–8.
47. Pham TM, Chu HN. Multi-provider and multi-domain resource orchestration in network functions virtualization. *IEEE Access* 2019;**7**:86920–31.
48. Sharma P, Tyagi S. Enhancing network security: An SDN (Software Defined Networking) method. *Int J Eng Technol Innov* 2020;**9**(5):1359–63.
49. Bao J, Zhao B, Yu W, et al. OpenSAN: A software-defined satellite network architecture. *Proc ACM SIGCOMM Comput Commun* 2015;**44**(4):347–8.
50. Feng J, Jiang L, Shen Y, et al. A scheme for software defined ORS satellite networking. *2014 IEEE Fourth International Conference on Big Data and Cloud Computing*; 2014 Dec 3-5; Sydney, Australia. Piscataway: IEEE Press; 2014.p.716–21.
51. Secinti G, Darian PB, Canberk B, et al. SDNs in the sky: Robust end-to-end connectivity for aerial vehicular networks. *IEEE Commun Mag* 2018;**56**(1):16–21.
52. Pan CS, Shi JJ, Yang L, et al. Satellite network load balancing strategy for SDN/NFV collaborative deployment. *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*; 2019 Aug 19-23; Leicester, UK. Piscataway: IEEE Press; 2019.p.1406–11.
53. Jawad N, Salih M, Ali K, et al. Smart television services using NFV/SDN network management. *IEEE Trans Broadcast* 2019;**65**(2):404–13.
54. Iqbal H, Ma J, Stranc K, et al. A software-defined networking architecture for aerial network optimization. *2016 IEEE NetSoft Conference and Workshops (NetSoft)*; 2016 Jun 6-10; Seoul, Korea (South). Piscataway: IEEE Press; 2016.p.151–5.
55. Ding G, Wu F, Wu Q, et al. Robust online spectrum prediction with incomplete and corrupted historical observations. *IEEE Trans Veh Technol* 2017;**66**(9):8022–36.
56. Wu Y, Hu G, Jin F, et al. A satellite handover strategy based on the potential game in LEO satellite networks. *IEEE Access* 2019;**7**:133641–52.
57. Ren J, Yu G, He Y, et al. Collaborative cloud and edge computing for latency minimization. *IEEE Trans Veh Technol* 2019;**68**(5):5031–44.
58. Kai C, Zhou H, Yi Y, et al. Collaborative cloud-edge-end task offloading in mobile-edge computing networks with limited communication capability. *IEEE Trans Cogn Commun Netw* 2021;**7**(2):624–34.
59. Bari F, Chowdhury SR, Ahmed R, et al. Orchestrating virtualized network functions. *IEEE Trans Netw Serv Manage* 2016;**13**(4):725–39.
60. van den Bossche R, Vanmechelen K, Broeckhove J. Cost-optimal scheduling in hybrid IaaS clouds for deadline constrained workloads. *2010 IEEE 3rd international conference on cloud computing; 2010 Jul 5-10; Miami, USA*. Piscataway: IEEE Press; 2010. p. 228–35.
61. Khalid O, Khan MUS, Huang Y, et al. EvacSys: A cloud-based service for emergency evacuation. *IEEE Cloud Comput* 2016;**3**(1):60–8.
62. Cui G, Li X, Xu L, et al. Latency and energy optimization for MEC enhanced SAT-IoT networks. *IEEE Access* 2020;**8**:55915–26.
63. Li C, Zhang Y, Hao X, et al. Jointly optimized request dispatching and service placement for MEC in LEO network. *China Commun* 2020;**17**(8):199–208.
64. Cisco. Cisco visual networking index: Global mobile data traffic forecast update[Internet]. Available from: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.
65. Statista, Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2030[Internet]. Available:https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/.
66. Afolabi I, Taleb T, Samdanis K, et al. Network slicing and softwarization: a survey on principles, enabling technologies, and solutions. *IEEE Commun Surv Tutor* 2018;**20**(3):2429–53.
67. Jawad N, Salih M, Ali K, et al. Network slicing to enable scalability and flexibility in 5G mobile networks. *IEEE Commun Mag* 2017;**55**(5):72–9.
68. Mahindra R, Khojastepour MA, Zhang HH, et al. Radio access network sharing in cellular networks. *2013 21st IEEE International Conference on Network Protocols (ICNP); 2013 Oct 7-10; Goettingen, Germany*. Piscataway: IEEE Press; 2013. p. 1–10.
69. Bor-Yaliniz I, Yanikomeroglu H. The new frontier in RAN heterogeneity: Multi-tier drone-cells. *IEEE Commun Mag* 2016;**54**(11):48–55.
70. Jou BT, Vidal O, Cahill J, et al. Architecture options for satellite integration into 5G networks. *2018 European Conference on Networks and Communications (EuCNC); 2018 Jun 18-21; Ljubljana, Slovenia*. Piscataway: IEEE Press; 2018. p. 398–9.
71. Zhang S, Luo H, Li J, et al. Hierarchical soft slicing to meet multi-dimensional QoS demand in cache-enabled vehicular networks. *IEEE Trans Wirel Commun* 2020;**19**(3):2150–62.
72. Foo YC, Lim WL, Tafazolli R, et al. Other-cell interference and reverse link capacity of high altitude platform station CDMA system. *Electron Lett* 2000;**36**(22):1881.
73. Lyu F, Yang P, Wu HQ, et al. Service-oriented dynamic resource slicing and optimization for space-air-ground integrated vehicular networks. *IEEE Trans Intell Transp Syst* 2021. https://doi.org/10.1109/TITS.2021.3070542.
74. Kasgari ATZ, Saad W. Stochastic optimization and control framework for 5G network slicing with effective isolation. *2018 52nd annual conference on information sciences and systems (CISS); 2018 Mar 21-23; Princeton, USA*. Piscataway: IEEE Press; 2018. p. 1–6.
75. Grac D, Thornton J, Chen GH, et al. Improving the system capacity of broadband services using multiple high-altitude platforms. *IEEE Trans Wirel Commun* 2005;**4**(2):700–9.
76. Foo YC, Lim WL, Tafazolli R. Centralized downlink call admission control for high altitude platform station UMTS with onboard power resource sharing. *Proceedings IEEE 56th vehicular technology conference; 2002 Sep 24-28; Vancouver, Canada*. Piscataway: IEEE Press; 2002. p. 549–53.
77. Xiong K, Leng S, Hu J, et al. Smart network slicing for vehicular fog-RANs. *IEEE Trans Veh Technol* 2019;**68**(4):3075–85.
78. Tang J, Shim B, Quek TQS. Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB. *IEEE J Sel Areas Commun* 2019;**37**(4):881–95.
79. He J, Song W. AppRAN: Application-oriented radio access network sharing in mobile networks. *2015 IEEE international conference on communications (ICC); 2015 Jun 8-12; London, UK*. Piscataway: IEEE Press; 2015. p. 3788–94.
80. Kokku R, Mahindra R, Zhang H, et al. NVS: A substrate for virtualizing wireless resources in cellular networks. *IEEE/ACM Trans Netw* 2012;**20**(5):1333–46.

81. Sun W, Strom EG, Brannstrom F, et al. Radio resource management for D2D-based V2V communication. *IEEE Trans Veh Technol* 2016;**65**(8):6636–50.
82. Sun W, Yuan D, Strom EG, et al. Cluster-based radio resource management for D2D-supported safety-critical V2X communications. *IEEE Trans Wirel Commun* 2016;**15**(4):2756–69.
83. Jiang M, Condoluci M, Mahmoodi T. Network slicing management & prioritization in 5G mobile systems. *Proc. European wireless 2016, 22th European wireless conference; 2016 May 18-20; Oulu, Finland*. Piscataway: IEEE Press; 2016. p. 1–6.
84. Peng H, Ye Q, Shen X. Spectrum management for multi-access edge computing in autonomous vehicular networks. *IEEE Trans Intell Transp Syst* 2020;**21**(7):3001–12.
85. Deng ZL, Du Q, Li N, et al. RL-based radio resource slicing strategy for software-defined satellite networks. *2019 IEEE 19th international conference on communication technology (ICCT); 2019 Oct 16-19; Xi'an, China*. Piscataway: IEEE Press; 2019. p. 897–901.
86. Zhang N, Zhang S, Yang P, et al. Software defined space-air-ground integrated vehicular networks: Challenges and solutions. *IEEE Commun Mag* 2017;**55**(7):101–9.
87. Albonda HDR, Perez-Romero J. An efficient RAN slicing strategy for a heterogeneous network with eMBB and V2X services. *IEEE Access* 2019;**7**:44771–82.
88. Sun Y, Feng G, Zhang L, et al. Distributed learning based handoff mechanism for radio access network slicing with data sharing. *ICC 2019 - 2019 IEEE International Conference on Communications (ICC); 2019 May 20-24; Shanghai, China*. Piscataway: IEEE Press; 2019. p. 1–6.
89. Wang G, Zhou S, Zhang S, et al. SFC-based service provisioning for reconfigurable space-air-ground integrated networks. *IEEE J Sel Areas Commun* 2020;**38**(7):1478–89.
90. Chen K, Zhao S, Lv N. Network monitoring information collection in the SDN-enabled airborne tactical network. *Int J Aerosp Eng* 2018;**2018**:1–20.
91. Hawilo H, Shami A, Mirahmadi M, et al. NFV: State of the art, challenges, and implementation in next generation mobile networks (vEPC). *IEEE Netw* 2014;**28**(6):18–26.
92. Cunha VA, da Silva E, de Carvalho MB, et al. Network slicing security: challenges and directions. *Internet Technol Lett* 2019;**2**(5):e125.
93. Li X, Samaka M, Chan HA, et al. Network slicing for 5G: challenges and opportunities. *IEEE Internet Comput* 2017;**21**(5):20–7.
94. Haider S, Akhunzada A, Mustafa I, et al. A deep CNN ensemble framework for efficient DDoS attack detection in software defined networks. *IEEE Access* 2020;**8**:53972–83.
95. Prados-Garzon J, Ramos-Munoz JJ, Ameigeiras P, et al. Modeling and dimensioning of a virtualized MME for 5G mobile networks. *IEEE Trans Veh Technol* 2017;**66**(5):4383–95.
96. Chiang M, Zhang T. Fog and IoT: An overview of research opportunities. *IEEE Internet Things J* 2016;**3**(6):854–64.
97. Kai K, Cong W, Tao L. Fog computing for vehicular Ad-hoc networks: paradigms, scenarios, and issues. *J China Univ Posts Telecommun* 2016;**23**(2):56–96.
98. Abbas N, Zhang Y, Taherkordi A, et al. Mobile edge computing: a survey. *IEEE Internet Things J* 2018;**5**(1):450–65.
99. Wang S, Zhang X, Zhang Y, et al. A survey on mobile edge networks: Convergence of computing, caching and communications. *IEEE Access* 2017;**5**:6757–79.
100. Mao Y, You C, Zhang J, et al. A survey on mobile edge computing: The communication perspective. *IEEE Commun Surv Tutorials* 2017;**19**(4):2322–58.
101. Xu J, Chen LX, Zhou P. Joint service caching and task offloading for mobile edge computing in dense networks. *IEEE INFOCOM 2018 - IEEE conference on computer communications; 2018 Apr 16-19; Honolulu, USA*. Piscataway: IEEE Press; 2018. p. 207–15.
102. Cheng X, Lyu F, Quan W, et al. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach. *IEEE J Sel Areas Commun* 2019;**37**(5):1117–29.
103. Zhang Y, Di B, Wang P, et al. HetMEC: Heterogeneous multi-layer mobile edge computing in the 6G era. *IEEE Trans Veh Technol* 2020;**69**(4):4388–400.
104. Wang X, Han Y, Wang C, et al. In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Netw* 2019;**33**(5):156–65.
105. Wang SQ, Urgaonkar R, Zafer M, et al. Dynamic service migration in mobile edge-clouds. *2015 IFIP networking conference (IFIP Networking); 2015 May 20-22; Toulouse, France*. Piscataway: IEEE Press; 2015. p. 1–9.
106. Katsalis K, Papaioannou TG, Nikaein N, et al. SLA-driven VM scheduling in mobile edge computing. *2016 IEEE 9th international conference on cloud computing (CLOUD); 2016 Jun 27 - Jul 2; San Francisco, USA*. Piscataway: IEEE Press; 2016. p. 750–7.
107. Wen Z, Cala J, Watson P, et al. Cost effective, reliable and secure workflow deployment over federated clouds. *IEEE Trans Serv Comput* 2017;**10**(6):929–41.
108. Mobile Edge Computing (MEC): Framework and reference architecture. 2016. Standard No.: ETSI GS MEC 003.
109. Brik B, Frangoudis PA, Ksentini A. Service-oriented MEC applications placement in a federated edge cloud architecture. *ICC 2020 - 2020 IEEE International Conference on Communications (ICC); 2020 Jun 7-11; Dublin, Ireland*. Piscataway: IEEE Press; 2020. p. 1–6.
110. Lim WYB, Luong NC, Hoang DT, et al. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Commun Surv Tutorials* 2020;**22**(3):2031–63.
111. Lu X, Liao Y, Lio P, et al. Privacy-preserving asynchronous federated learning mechanism for edge network computing. *IEEE Access* 2020;**8**:48970–81.
112. Shao M, Liu J, Yang Q, et al. A learning based framework for MEC server planning with uncertain BSs demands. *IEEE Access* 2020;**8**:198832–44.
113. Chen M, Mozaffari M, Saad W, et al. Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience. *IEEE J Sel Areas Commun* 2017;**35**(5):1046–61.
114. Zhou CH, Wu W, He HL, et al. Delay-aware IoT task scheduling in space-air-ground integrated network. *2019 IEEE global communications conference (GLOBECOM); 2019 Dec 9-13; Waikoloa, USA*. Piscataway: IEEE Press; 2019. p. 1–6.
115. Mao S, He S, Wu J. Joint UAV position optimization and resource scheduling in space-air-ground integrated networks with mixed cloud-edge computing. *IEEE Syst J* 2021;**15**(3):3992–4002.
116. Wang D, Tian X, Cui H, et al. Reinforcement learning-based joint task offloading and migration schemes optimization in mobility-aware MEC network. *China Commun* 2020;**17**(8):31–44.
117. Zhou S, Wang G, Zhang S, et al. Bidirectional mission offloading for agile space-air-ground integrated networks. *IEEE Wirel Commun* 2019;**26**(2):38–45.
118. Zhou C, Wu W, He H, et al. Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN. *IEEE Trans Wireless Commun* 2021;**20**(2):911–25.
119. Gu SS, Sun XY, Yang ZH, et al. Energy-aware coded caching strategy design with resource optimization for satellite-UAV-vehicle integrated networks. *IEEE Internet Things J* 2021. https://doi.org/10.1109/JIOT.2021.3065664.
120. Shen Y, Shi Y, Zhang J, et al. Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis. *IEEE J Sel Areas Commun* 2021;**39**(1):101–15.
121. ONAP Consumption models[Internet]. Available: https://www.onap.org/wp-content/uploads/sites/20/2020/06/ONAP_EUAG_Whitepaper_061720.pdf.
122. Ontology (information science)[Internet]. Available: https://en.wikipedia.org/wiki/Ontology(information_science).

123. Lopez de Vergaro JE, Villagra VA, Asensio JI, et al. Ontologies: Giving semantics to network management models. *IEEE Netw* 2003;**17**(3):15–21.
124. de Vergara JEL, Villagra VA, Berrocal J. Applying the Web ontology language to management information definitions. *IEEE Commun Mag* 2004;**42**(7):68–74.
125. Sahlmann K, Scheffler T, Schnor B. Ontology-driven device descriptions for IoT network management. *2018 Global Internet of Things Summit (GIoTS); 2018 Jun 4-7; Bilbao, Spain*. Piscataway: IEEE Press; 2018. p. 1–6.
126. Ding G, Wang J, Wu Q, et al. Robust spectrum sensing with crowd sensors. *IEEE Trans Commun* 2014;**62**(9):3129–43.
127. Wild T, Braun V, Viswanathan H. Joint design of communication and sensing for beyond 5G and 6G systems. *IEEE Access* 2021;**9**:30845–57.
128. Chen X, Feng Z, Wei Z, et al. Performance of joint sensing-communication cooperative sensing UAV network. *IEEE Trans Veh Technol* 2020;**69**(12):15545–56.
129. Sezer OB, Dogdu E, Ozbayoglu AM. Context-aware computing, learning, and big data in Internet of Things: A survey. *IEEE Internet Things J* 2018;**5**(1):1–27.
130. GLUE's benchmark leaderboard[Internet]. Available from: https://gluebenchmark.com/.
131. Zeydan E, Turk Y. Recent advances in intent-based networking: A survey. *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring); 2020 May 25-28; Antwerp, Belgium*. Piscataway: IEEE Press; 2020. p. 1–5.
132. Bankey V, Upadhyay PK. Physical layer security of multiuser multirelay hybrid satellite-terrestrial relay networks. *IEEE Trans Veh Technol* 2019;**68**(3):2488–501.
133. Li BH, Liang RC, Zhou W, et al. LBS meets blockchain: an efficient method with security preserving trust in SAGIN. *IEEE Internet Things J* 2021. https://doi.org/10.1109/JIOT.2021.3064357.
134. Sun W, Wang L, Wang P, et al. Collaborative blockchain for space-air-ground integrated networks. *IEEE Wirel Commun* 2020;**27**(6):82–9.
135. Tselios C, Politis I, Kotsopoulos S. Enhancing SDN security for IoT-related deployments through blockchain. *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN); 2017 Nov 6-8; Berlin, Germany*. Piscataway: IEEE Press; 2017. p. 303–8.