Toward Resilient Network Slicing for Satellite–Terrestrial Edge Computing IoT

Haitham H. Esmat, *Member, IEEE*, Beatriz Lorenzo, *Senior Member, IEEE*, and Weisong Shi, *Fellow, IEEE*

***Abstract*—Satellite–terrestrial edge computing networks (STECNs) emerged as a global solution to support multiple Internet of Things (IoT) applications in 6G networks. The enabling technologies to slice STECNs, such as software-defined networking (SDN), satellite edge computing (EC), and network function virtualization (NFV) are key to realizing this vision. In this article, we survey and analyze network slicing (NS) solutions for STECNs. We discuss slice management and orchestration for different STECNs integration architectures, satellite EC, mmWave/THz, and artificial intelligence solutions to make NS adaptive. In addition, we identify challenges and open issues to slice STECNs. In particular, resilient NS is crucial for essential and critical services. Network failures are unavoidable in large networks and can cause significant disruptions in NS, compromising many services. To this end, we present a resilient NS design to cope with failures and guarantee service continuity which is agnostic to the integration architecture and inherently multidomain. Further, we present strategies to achieve resilient networking and slicing in STECNs, including planning and provisioning of redundant network resources, design rules for service level agreement decomposition, and cross-domain solutions to detect and mitigate failures. Finally, promising future research directions are highlighted. This article provides valuable guidelines for slicing STECNs and will benefit key sectors, such as smart healthcare, e-commerce, Industrial IoT, education, and among others.**

***Index Terms*—Dynamic orchestration, Internet of Things (IoT), mmWave/THz, mobile edge computing (MEC), network intelligence, network slicing (NS), satellite–terrestrial edge computing networks (STECNs), resilient networking, software-defined networking (SDN).**

I. Introduction

SMART devices have been massively deployed in many fields, such as Industrial Internet of Thing (IIoT), transportation networks, and environment monitoring networks [1]. These large Internet of Things (IoT) deployments require mobile edge computing (MEC) [2], [3] to perform intensive computations for IoT devices with limited battery, communication, and computing capabilities. In parallel, the development and cost effectiveness of satellite equipment have led to satellite–terrestrial edge computing networks (STECNs) with multitier edge computing (EC) capabilities that can extend the connectivity to airborne networks and satellites, solving problems of deployment, coverage, and capacity, commonly faced in terrestrial networks. Unmanned aerial vehicles (UAVs) can assist terrestrial and satellite networks in data collection, execute some computation tasks, and relay the data to other nodes for further processing or aggregation. Thus, STECNs can provide diverse services that require low computational power with reduced delay. These services include emergency communications (EComms) [4], monitoring and reconnaissance (MAR) [5], in-space backhauling [6], and integration of cyber–physical systems [7]. In terrestrial networks, the tradeoff between latency and computational power in MEC has been extensively studied [8]. However, further research is needed to analyze such tradeoffs in satellite server EC for different IoT applications [9]. Several satellite–terrestrial integration architectures, as shown in Fig. 1, are proposed to expand network coverage, cope with mobility, and improve spectrum efficiency. The analysis of the performance considering the multitier EC capabilities is needed to exploit the characteristics of the different integration architectures for service differentiation.

Satellite networks should be virtualized for seamless integration with current 5G terrestrial networks [10]. Network slicing (NS) is a promising virtualization technique where the infrastructure is shared by multiple tenants (operators) to serve multiple service classes simultaneously [11]. Logical slices are created through the terrestrial–air–satellite segments, including communication, computing, and storage resources. Each slice is an independent virtual network expected to guarantee certain service level agreement (SLA) and provide complete network functionalities. NS in STECNs is multidomain since different operators provide the service at the terrestrial, air, and satellite segments, which will be referred to as domains. Moreover, different stakeholders may coexist and request different services, so STECNs must support multitenancy and functional service isolation. A software-defined network (SDN) implementation for low Earth orbit (LEO) satellite networks is presented in [12] to achieve high reliability and low latency in control links. A virtualized satellite–terrestrial network architecture is designed in [13] to support multiple Quality of Service (QoS) requirements through spectrum allocation. However, slicing in STECNs is quite challenging due to the following.

1) Different providers need to combine their resources to create slices that guarantee coverage across large geographical areas and administrative domains.

Manuscript received 30 January 2023; revised 2 May 2023; accepted 15 May 2023. Date of publication 18 May 2023; date of current version 8 August 2023. This work was supported in part by the U.S. National Science Foundation under Grant CNS-2008309 and Grant CNS-2225427. *(Corresponding author: Beatriz Lorenzo.)*

Haitham H. Esmat and Beatriz Lorenzo are with the Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003 USA (e-mail: habdelhafez@umass.edu; blorenzo@umass.edu).

Weisong Shi is with the Department of Computer and Information Sciences, University of Delaware, Newark, DE 19716 USA (e-mail: weisong@udel.edu).

Digital Object Identifier 10.1109/JIOT.2023.3277466

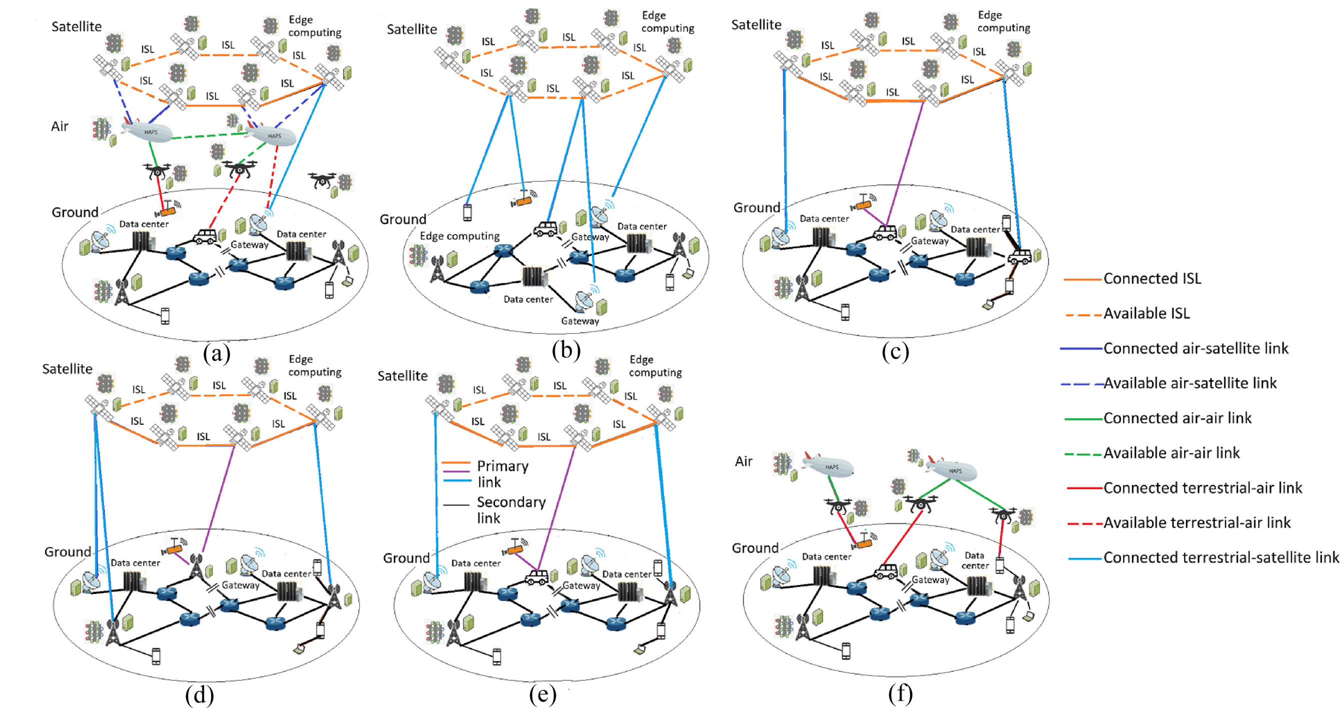

Fig. 1. STECN integration architectures. (a) Terrestrial–air–satellite network. (b) ITSN. (c) STRN. (d) STBN. (e) CogSTN. (f) HACN.

2) The network dynamics impact resource availability and thus, the SLA needs to be decomposed dynamically in different domains as the STECNs network topology evolves.
3) Performance maintenance throughout the slice life cycle is more demanding.
4) Network failures are difficult to predict and can compromise many services.

Existing works have overlooked the multidomain aspect of slicing STECNs and failures.

Failures due to natural or human intervention are unavoidable in large networks and can disrupt critical services having catastrophic effects [14]. Network resilience refers to the ability of the network to maintain acceptable performance in the presence of failures or other threats to its normal operation. Achieving resilience in STECNs is especially challenging since the design and implementation of resilient communication protocols and algorithms are further complicated by mobility and network dynamics. Important questions are how to maintain connectivity in STECNs and guarantee performance, which network integration architecture is more resilient to failures, and how to design resilient resource management and orchestration (MANO) schemes for heterogeneous services, and among others. Existing works on resilient mobile networking [15], [16] mainly focus on technologies that enhance network survivability and fault tolerance. However, in addition to maintaining connectivity, new applications also require faster, more efficient, and energy saving-communications. In fact, next-generation networks must meet stringent service requirements, such as low latency, high reliability, and throughput, all of which require optimal resource utilization even in the event of a failure [17]. Recent techniques, such as millimeter-wave communications, artificial intelligence (AI), traffic offloading, and service migration schemes contribute to the support of resilient network services. In this article, we leverage these efforts and provide further solutions at the ground, airborne, and satellites to palliate network failures and build resilient NS mechanisms. Our aim is to design schemes that provide adaptability to traffic dynamics, autonomous management, and reconfiguration against multiple failures to guarantee service continuity. We present a holistic approach to improve the resilience within each domain and across domains to avoid improper network integration.

A. Paper Motivation

This article investigates and analyzes NS in satellite–terrestrial networks, rise awareness of the importance of resilient slicing in STECNs, and presents a resilient NS design. A comprehensive survey and analysis of slicing STECNs are missing. Existing surveys on NS focus on terrestrial networks, as summarized in Table I. In particular, the topics covered include NS architecture, enabling technologies, standardization efforts, security, and use cases. Khan et al. [18] investigated NS in several IoT smart applications, taxonomy, design principles, service function chaining schemes, physical infrastructures, and security. Addad et al. [19] investigated interslice mobility management in 5G, drone traffic control, and autonomous vehicles. Taleb et al. [20] studied multidomain NS in terrestrial networks. Afolabi et al. [21] studied end-to-end (E2E) performance in NS, enabling technologies, and standardization efforts. Wijethilaka and Liyanage [22] analyzed how NS contributed to IoT realization and the application of blockchain and AI in NS. Wu et al. [23] surveyed NS for IIoT services with a special focus on smart transportation, smart energy, and smart factory. Debbabi et al. [24] reviewed

TABLE I
LITERATURE REVIEW (PARTIALLY COVERED: ∂ AND COVERED:✓)

Reference	Network slicing	Multi-domain	Non-terrestrial networks	SLA	Resiliency	Enabling technologies	MANO	mmWave/THz	Security/privacy	Standards	AI	Applications	Remarks
[18]	✓					✓	✓		✓		✓	Smart devices	Taxonomy and requirements of network slicing in IoT smart applications, key design principles, service function chaining schemes, physical infrastructures, and security.
[19]	✓	✓				✓	✓		✓			Autonomous vehicles	Investigates inter-slice mobility management in 5G when using drones for traffic control, and autonomous vehicles for video streaming.
[20]	✓	✓				✓	✓		✓	✓		5G	Multi-domain network slicing in terrestrial networks.
[21]	✓	✓				✓	✓		✓	✓		5G	End-to-end network slicing, enabling technologies, latest solutions, and standardization efforts.
[22]	✓	✓				✓	✓		✓		✓	IoT	Analyzes how network slicing contributed to IoT realization together with emerging technologies such as blockchain and AI and IoT integration.
[23]	✓	✓		✓		✓	✓		✓	✓	✓	IIoT	Network slicing for Industrial Internet of Things services with a special emphasis on smart transportation, smart energy, and smart factory.
[24]	✓	✓		∂		✓	✓		✓			5G	Network slicing with a focus on management and orchestration across multiple domains and the algorithmic aspects.
[25]	∂	✓	✓			∂				✓		IoT	LEO satellites for IoT and coverage and latency in rural and underserved areas. Open challenges and areas of further research are discussed.
[26]	✓		✓			✓		✓	∂	✓	✓	IoT	Evolution of non-terrestrial networks, important features of their integration, relevance to 5G and 6G, architectures, and higher layer issues.
[27]	✓		✓			✓	✓			✓	✓	5G	Satellite communications, applications needs for wider coverage, integration between wired and wireless networks for media streaming.
[28]			✓			✓		∂	✓	∂		IoT	Cross-layer design, and resource management and allocation for satellite, aerial, and terrestrial communications.
[29]			✓			✓		∂	✓		✓	IoT, 6G	Satellite-terrestrial networks towards 6G objectives of seamless connectivity for everyone and everything, and integration architectures.
[30]			✓			∂			✓		✓	IoT	AI-based solutions to satellite communication problems such as resource management, control, energy consumption, and security.
[31]			✓							✓		5G	Challenges in the application of the 5G standard to Non-Terrestrial Networks, and design considerations for applicability of 5G in Satellite RANs.
Our work	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	IoT, 6G	Resilient network slicing design, enabling technologies, resilient solutions for slicing communications and computing in STECNs.

MANO architectures for NS across multiple domains and their algorithmic aspects.

Some recent surveys on satellite networks identify NS as an enabling technology for the integration of satellite and terrestrial networks [25], [26], [27]. However, the main focus of these surveys is on the evolution of satellite–terrestrial networks, the important features of their integration, and their relevance beyond 5G and 6G ecosystems. Centenaro et al. [25] presented various architectures, higher-layer aspects pertinent to this integration, and use cases. An in-depth literature review of the driving factors spurring renewed interest in satellite communications is presented in [27] together with application needs for wider coverage and integration between existing wired and wireless networks. Liu et al. [28] summarized existing works on resource management and cross-layer optimization in satellite, airborne, and terrestrial networks. Zhu and Jiang [29] surveyed various integration architectures, their applications, and challenges for seamless integration. Fourati and Alouini [30] presented AI-based solutions for resource management, control, security, and energy usage in satellite networks.

B. Our Contribution

We are the first to present a survey on slicing satellite–terrestrial networks. The obtained knowledge would be crucial to accommodate different use cases in these networks and design resilient NS schemes. The main contributions of this article are summarized as follows.

1) We discuss the challenges of slicing STECNs and analyze NS in different satellite–terrestrial integration architectures, enabling technologies to slice STECNs, multidomain MANO, AI-based slicing schemes, and open issues.
2) We identify possible types of failures in STECNs, present solutions to enhance the resilience of communications and computing, and discuss resilience metrics and the challenges of designing resilient NS schemes in STECNs.
3) A resilient NS design is presented to cope with multiple failures and guarantee service continuity. Our design is agnostic to the integration architecture and includes planning and provisioning of redundant network resources, design rules for SLA decomposition, cross-domain solutions to perform failure control, and collaboration strategies for resource sharing.
4) Future research directions and guidelines to build resilient NS solutions are provided.

The remainder of this article is organized as follows. Section II introduces STECN architectures and the enabling technologies to slice STECNs. NS in STECNs is discussed in

Section III together with the latest projects, and standardization efforts. Resilient networking and slicing strategies are given in Section IV. Our proposed resilient NS scheme is presented in Section V. Finally, Section VI discusses open issues and future research directions.

II. Satellite–Terrestrial Edge Computing Networks and Enabling Technologies

This section introduces STECN architectures, and the enabling technologies to slice STECNs.

A. Satellite–Terrestrial Integration Architectures

A generic STECN is shown in Fig. 1(a), which can support different applications by using resources at the terrestrial–air, terrestrial–satellite, or terrestrial–air–satellite domains. The elements of the generic STECN are AS follows.

1) *Terrestrial IoT:* Consisting of IoT devices, cellular terminals, machine-type communication terminals (dual-mode C-band with terrestrial networks or Ka-band with LEO), and terrestrial EC facilities (e.g., MEC).
2) *Low-Altitude Platforms (LAPs):* Consisting of UAVs that monitor the environment, process their own data and the data collected from IoT devices, and offload it to the high-altitude platforms (HAPs) or satellites for further processing.
3) *HAPs:* Consisting of EC-enabled HAPs with broadband wireless connections deployed in the stratosphere. They provide wider coverage and have more powerful computational capabilities than LAPs. They are suitable for coordinating terrestrial/satellite resources, performing computation offloading, and caching fundamental data associated with the applications to reduce relay and energy consumption for IoT devices.
4) *Satellites:* Consisting of LEO satellites that collect their own data from the environment, execute their own data, and the data collected from UAVs locally when the backhaul links from satellites to ground stations (GSs) are not available. Besides, LEO satellites can relay the data using intersatellite links (ISLs) to other satellites or to a GS for more complex processing.
5) *Gateway:* Consisting of a GS that can analyze raw data received from LEO satellites or send it to a remote data center. The generic architecture can support heterogeneous requirements by using different topological configurations, as shown in Fig. 1. For example, delay-tolerant data from a mobile terminal can be relayed by the satellite network to the gateway, and the computing task can be served at the gateway cloud. Likewise, intermediate drones and satellites can process delay-demanding tasks to reduce latency. Configuration decisions involve task allocation to EC servers, routing, and resource allocation and require an advanced ML algorithm for autonomous management and adaptation of the configuration to the wireless environment.

A detailed survey on satellite–terrestrial integration architectures is presented in [29]. By combining the advantages of the different network segments, more efficient resource allocation, mobility management, and service continuity can be achieved. In particular, the following integration architectures have been proposed.

1) Integrated Terrestrial–Satellite Network [32]: An integrated terrestrial–satellite network (ITSN) consisting of LEO satellites augmented with a terrestrial network with a number of GSs that can access data centers is shown in Fig. 1(b). The user's devices communicate with the satellites through terrestrial-to-satellite links. Satellites can relay information to neighboring satellites in the same or adjacent orbits using ISLs.

2) Satellite–Terrestrial Relay Network [33], [34]: A hybrid satellite–terrestrial relay network (STRN) is shown in Fig. 1(c) in which a relay is used to transmit to the satellite when there is no direct link available between the user and the satellites. The unavailability of the direct link may be due to rain/fog attenuation, obstacles, etc., which lead to the masking effect. Sreng et al. [34] extended this architecture to a cooperative architecture that achieves higher diversity by combining both the masked direct link and the relay link.

3) Satellite–Terrestrial Backhaul Network [35]: In the satellite–terrestrial Backhaul network (STBN), satellite backhaul links connect terrestrial access points and base stations (BSs) to the core network, as shown in Fig. 1(d). In emergency situations, satellite backhaul links can maintain the network connection as a backup to enable reliable communications even if terrestrial links fail.

4) Cognitive Satellite–Terrestrial Network [36]: In a cognitive satellite–terrestrial network (CogSTN) spectrum resources can be dynamically shared among satellite and terrestrial networks. An example of CogSTN is shown in Fig. 1(e) where the satellite and the terrestrial networks act as the primary and secondary networks, respectively. CogSTN can also incorporate terrestrial relays to assist the primary network transmission to the end users resulting in a hybrid STRN known as CogSTRN [37].

5) Cooperative Satellite–Terrestrial Network [34]: In a cooperative satellite–terrestrial network (CooSTN), the satellite and terrestrial networks cooperate to serve users who need to be equipped with dual-mode terminals to access both networks. By exploiting the resources available on both networks, overall performance and user satisfaction can be improved. For instance, users located in urban areas can be served by the terrestrial network if it is not congested. Otherwise, users will be served by the satellite network. Similarly, users located in remote areas (rural, sea, or airspace), outside the coverage of the terrestrial network, can be served by the satellite network.

6) Hierarchical Aerial Computing Network [38]: In hierarchical aerial computing network (HACN), the high-altitude and LAPs in the aerial access network, such as HAPs and UAVs, respectively, provide effective MEC services for IoT applications. In particular, a great amount of research has focused on the partial integration of UAVs to collect and process IoT data given their advantages in terms of fast deployment, low latency, and processing capacity [39]. HACNs are especially needed in remote areas or emergency situations with limited access to terrestrial cellular networks. An example of HACN is shown in Fig. 1(f).

Further research on EC for STECNs is needed to exploit its advantages in these architectures.

B. Satellite Edge Computing

MEC in STECNs [40], [41], [42] can take place in the terrestrial domain (i.e., terrestrial offloading), air domain (i.e., airborne offloading), and satellite domain (i.e., satellite-borne offloading). MEC brings the following advantages in STECNs.

1) Content Caching: Decreases the traffic in STECNs by precaching popular content and thus, reduces the retransmission of the same content. The application data requested can be quickly accessed which reduces delay and congestion in the network, and improves resiliency. This is especially relevant when transmitting multimedia data.

2) Computation Offloading: Offloading computation tasks closer to end users improve resource efficiency and service performance and also reduces the energy consumption caused by long-distance data transmissions. As the computational power of EC servers is limited it is important to define offloading policies and scheduling schemes to fulfill diverse network requirements. Delay-sensitive and compute-intensive applications benefit the most from computation offloading.

3) Network Services: MEC provides the capability to implement new network services in STECNs, such as edge intelligence, deployment of virtual network functions (VNFs), and among others. Since the topology of STECNs changes dynamically, network services to manage dynamic resource availability can effectively improve resource utilization. For instance, the terrestrial network can receive information regarding resource availability in the satellite network and make timely offloading decisions.

4) Efficient Use of Satellite Bandwidth Resources: Satellites can process locally their captured remote sensing information or offload it to airborne platforms, reducing the amount of data transmitted back to the ground.

5) Reliability of Communication: Satellite networks can provide redundant/backup transmissions for the terrestrial network, increasing the reliability of the communication.

6) Security and Privacy: MEC improves security and privacy by reducing the exposure of the users' information (reducing the amount of data transmitted throughout the network, which reduces the risk of data leaks).

The design aspects of incorporating MEC in STECN are discussed in [41]. In particular, they address QoS requirements, fault/failure recovery, multinode task scheduling, cooperative computation offloading, and mobility management. A cooperative computation offloading model is presented by Zhang et al. [43] to achieve parallel computation in STECNs. They describe several task scheduling models in detail. In [42] a task allocation algorithm for LEO satellite EC is presented based on a greedy strategy for the Walker Delta satellite constellation. They analyze the performance of their proposed algorithm in terms of computational cost. In [44] a novel MEC framework for terrestrial–satellite IoT is proposed together with an energy-efficient computation offloading algorithm. The computation offloading problem from IoT devices to LEO satellites is solved by sequential fractional programming to minimize the latency and then by Lagrangian dual decomposition method.

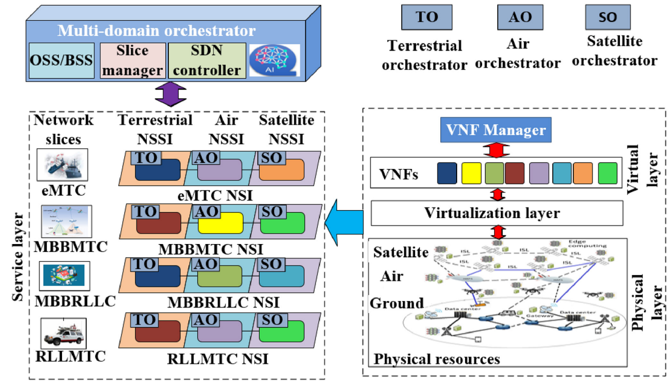

Fig. 2. Virtualization architecture for STECNs for MBBRLLC, MBBMTC, eMTC, and RLLMTC services.

C. Network Function Virtualization

Network function virtualization (NVF) decouples software and hardware equipment and enables the virtualization of traditional physical network equipment, such as switches, routers, and firewalls, to run in virtual machines (VMs). By introducing NFV into satellite–terrestrial networks, the available resources (e.g., bandwidth, computing, storage, etc.) can be abstracted into a resource pool and allocated on-demand for IoT services. The NFV architecture has three components: 1) VNFs which are the software form of the network functions deployed on virtual environments; 2) NFV Infrastructure that contains the resources upon which the VNFs will be deployed; and 3) MANO that is responsible for the lifecycle MANO of VNFs, NFVI, and deployed services. Fig. 2 shows the virtualization of STECNs which consist of a physical layer, a virtual layer, and a service layer. The physical layer represents the physical infrastructure of the STECN which includes all the physical resources required to host and run the virtualized environment (e.g., servers, storage devices, network switches, and other hardware components). The virtual layer sits on top of the physical layer and is responsible for creating and managing VMs and VNFs. It abstracts the physical hardware for the higher-level service layer as a pool of virtual resources that can be dynamically allocated. The service layer provides the actual services that users and applications interact with. It includes the operating systems, middleware, applications, and other software components that run on the virtualized infrastructure. The service layer is responsible for managing the resources provided by the virtual layer to ensure the required levels of performance, availability, and security. By NFV, the required services of a slice can be represented by a service function chain (SFC) containing a sequence of VNFs. These VNFs are deployed on different network nodes in STECNs, e.g., IoT devices, vehicles, UAVs, satellites, etc. For instance, a monitor/surveillance service involves the following network functions [45]: recording the video, subtracting background, compressing and decoding the data, and detecting and classifying objects. Different network nodes will implement these functions based on their capabilities and characteristics.

The mapping of VNF in an SFC onto physical nodes in STECNs is studied in [45] and [46]. In [46], the VNF mapping process is studied in static scenarios. Their objective is to maximize the service requests served while minimizing the cost of running the VNFs on different network nodes. They formulate the problem as an integer nonlinear programming problem. Li et al. [45] investigated the dynamic VNF mapping and scheduling for Internet of Vehicles (IoV) services. The VNF mapping strategy is readjusted due to varying resources or network requirements in the slice. They model the delay and migration cost experienced by VNF live migration and reinstantiation. The problem is formulated as a mixed-integer linear programming (MILP) and solved by Tabu Search-based algorithms.

Several works study how to deploy VNFs in satellite edge servers via centralized [47] and distributed approaches [48]. In [48] the VNF placement problem is solved to minimize the deployment cost by using a decentralized resource allocation algorithm based on a potential game and finding a Nash equilibrium. Gao et al. [49] assumed that a priori information about service types, resource requirements, and running periods for different services can be obtained by satellite mission planning. They aim to minimize resource utilization and the number of servers used for service chaining placement.

D. SDN for Satellite Networks

Future STECNs will include a large number of satellites and thus, control of the network becomes more complex. The successful application of SDN in terrestrial networks to manage large and complex networks anticipates significant benefits in STECNs in terms of more flexible configurations and management. Therefore, SDN is being introduced in STECNs [50] to optimize the network and coordinate network resources by separating data forwarding from the SDN control plane and enabling open programmable control. Moreover, SDN provides a time-sensitive response to serve critical services and optimizes local resources. The SDN architecture consists of four parts: infrastructure layer (i.e., data plane), control layer, application layer, and two open programming interfaces, northbound interface (NBI) and control–data–plan interface (CDPI), also known as southbound interface (SBI) [51]. The control layer communicates with the application layer using NBI, whereas SBI is used for communication with the data plane. The CDPI provides a standard interface between the network control and data planes, allowing them to exchange information and coordinate their actions. The NBI allows external applications to program the network and retrieve real-time information about network events and status. This information can be used to make dynamic decisions about network traffic management, network optimization, and security. The application layer contains virtualization, resource allocation, routing, load balancing and security policies, and slice and service primitives. The virtualization policy is responsible for creating and managing virtualized resources. The resource allocation policy

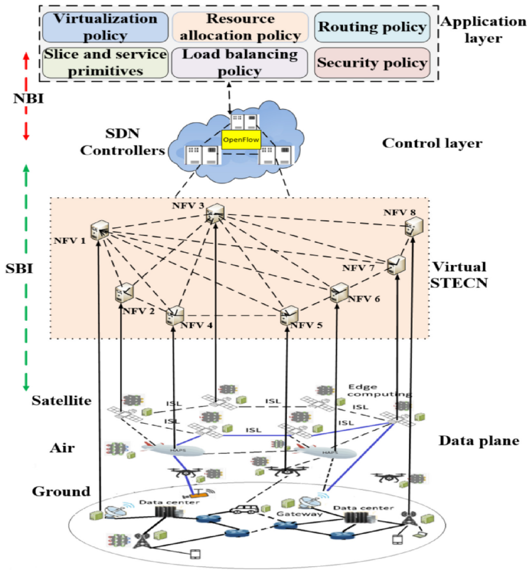

Fig. 3. Architecture of SDN/NFV for STECNs.

determines how physical resources are allocated to virtualized environments based on predefined rules and policies. The routing policy determines how network traffic is directed between virtualized environments, ensuring efficient and secure communication between different system parts. The load balancing policy ensures that workload is evenly distributed across available resources, preventing any single resource from becoming overloaded and causing performance issues. A security policy defines the security rules and measures applied to virtualized environments, ensuring that data and resources are protected from unauthorized access. Slice and service primitives are programming interfaces that allow users or applications to access and interact with virtualized resources and services. An SDN/NFV architecture for STECNs is illustrated in Fig. 3 where the data and control planes are decoupled. SDN controllers orchestrate the VNFs to provide 6G services. To perform VNF scheduling and mapping, the SDN controllers store information, such as resource availability, QoS requirements, network conditions, and the physical locations of the NFV nodes.

The management architecture of STECNs should have a hierarchical and domain-based SDN control plane. Flat management schemes have limited scalability which translates into low efficiency of data synchronization among controllers. However, multicontroller deployments in STECNs face the following challenges: 1) dynamic STECN topology; 2) large-scale network; 3) increase or decrease of network nodes (e.g., vehicles, UAVs); and 4) unbalanced distribution of terrestrial users.

Recent works studied intelligent MANO schemes to leverage SDN in STECNs. Ferrús et al. [52] discussed the benefits associated with SDN and NFV in satellite networks and possible applications. Chen et al. [53] designed a hierarchical domain-based SDN architecture for STECNs in which the SDN control plane is divided into two layers. A primary control layer is implemented on the terrestrial network, while a secondary one is deployed on the satellite network. The SDN data plane encompasses terrestrial, air, and satellite networks. They determine the number of controllers and their relative positions by using the average network delay and the controllers' load. The SERvICE framework [54] consists of a novel integrated satellite–terrestrial network based on SDN and NFV that includes LEO, medium Earth orbit (MEO), and geostationary equatorial orbit (GEO) satellites. They introduce a management plane, control plane, and forwarding plane with the aim to achieve flexible satellite network traffic engineering (TE) and fine-grained QoS guarantee. They describe their prototype implementation with the help of delay-tolerant networks and OpenFlow. Mendoza et al. [55] described an SDN-based framework for the integration and management of the satellite capacity in STBN. They develop a TE application to manage dynamically a steerable satellite capacity for resilience or emergency purposes. The proposed scheme achieves optimal allocation of the available satellite and terrestrial capacity under both failure and nonfailure conditions. Oubbati et al. [56] surveyed existing works on UAV-assisted networks enabled by SDN and NFV. They discuss several use cases and open problems like the role of UAVs in cellular communications, routing, and monitoring. Several works [57], [58] study the incorporation of SDN into UAVs networks to ensure robust connectivity. For instance, Secinti et al. [57] proposed a multipath disjoint routing protocol to avoid link failures among UAVs transmissions. Their SDN architecture is aware of the limited computation resources of UAVs. Iqbal et al. [58] designed an SDN-based architecture for UAV networks to maximize network availability by predicting connectivity disruptions.

The controller placement problem (CPP) in SDN-based STECNs is quite complex due to the dynamic network topology and traffic variations. Distributing the load of the SDN controllers and finding their best placement is particularly critical to achieve efficient management and control of the network resources. Two approaches have been adopted: 1) static controller placement (SCP) and 2) dynamic controller placement (DCP). SCP has been applied to terrestrial networks [59] but cannot be directly adopted in satellite scenarios whereas DCP for satellite networks [60] need a controller migration strategy to adapt to topological dynamics and traffic variations. The drawback of the latter approach is the cost of migrating controllers frequently from one satellite to another which requires additional bandwidth resources. In [61], an SCP with dynamic assignment (SPDA) method is presented to solve the CPP. They incorporate SDN controllers into a fixed set of satellites, and then dynamically assign switches to existing controllers based on their traffic load and current latency. They show that SPDA consumes lower bandwidth than methods involving controller migrations.

The application of SDN to TE in broadband LEO satellite networks is studied in [62]. They investigate unicast and multicast TE using a centralized SDN management scheme. Unicast TE is used to support ubiquitous network access to LEO satellites and basic network services. They adopt a k-segment routing strategy based on segment routing techniques [63], and achieve low cost and near-optimal maximal link utilization. Multicast TE is used for content distribution and video streaming. They employ rectilinear Steiner trees (RSTs) [64] to maximize bandwidth savings and obstacle-avoiding RSTs (OARSTs) [65] to handle the contention of multiple multicast groups. Due to the movement of satellites, SDN control information is used in [66] to improve handover performance. Tao et al. [67] presented a load-balancing scheme to improve the performance of the handover process and distribute the traffic throughout the network to minimize congestion. By using a resilient congestion estimation scheme, they detour the traffic around nearby satellites to reduce the traffic load based on the network condition. IoT devices may have dual-mode terminals to access the terrestrial and satellite networks which complicates access control. Yan et al. [68] designed an SDN-based dynamic channel allocation algorithm. Centralized access control for IoT terminals is performed by an SDN integrated control center, and then channels are allocated dynamically by a satellite or terrestrial controller according to the priority of the terminals.

E. Summary and Discussion

This section presents different STECN integration architectures and enabling technologies to slice STECNs. Several works study task offloading and collaboration schemes to share computing resources in STECNs [41], [42], [43]. However, further work is needed to analyze the performance of MEC capabilities in different integration architectures and its improvements in terms of spectrum efficiency, delay, congestion, and security. Moreover, the characteristics of an SDN/NFV architecture for STECNs are detailed together with their advantages in control and management. More studies should be conducted on how to leverage the SDN controllers' information to cope with intermittent resource availability (per domain and across domains) and design adaptive routing, scheduling, and handover schemes.

III. Network Slicing in STECNs

This section presents multidomain NS, summarizes existing works on slicing STECNs, and discusses the challenges to slice STECNs. In addition, AI-based slicing schemes are presented together with the latest projects and standardization efforts.

A. Dynamic and Multidomain Network Slicing

NS is introduced into satellite systems to fulfill various service demands and user requirements, and enhance resource efficiency and flexibility [10]. By NS, the shared physical infrastructure is divided into multiple E2E logical networks, i.e., network slice instances (NSIs), where each E2E NSI is an independent virtual and programmable network architecture tailored to meet E2E SLAs. NFV and SDN allow NS to be feasible, where the former decouples the network functions from dedicated hardware and runs them as software in VMs, while the latter decouples the data plane from the control plane [69].

NS in STECNs is inherently dynamic and multidomain. NS under traffic uncertainty has been studied in [70], [71], [72], and [73] for terrestrial networks. Feng et al. [70] examined the tradeoff between revenue and delay in slice admissions using the Lyapunov optimization for a single operator and traffic class. Chien et al. [71] proposed an algorithm to prevent the over-provisioning of computing and network resources based on measured statistics of typical 5G services. Dynamic network slice reconfiguration and recovery schemes are proposed in [74].

In multidomain NS, a slice contains resources from domains that belong to different network operators (i.e., terrestrial, air, and satellite). In Fig. 2, an architecture of intelligent NS management in STECNs is shown to serve mobile broadband reliable low latency communication (MBBRLLC), mobile broadband machine type communication (MBBMTC), best-effort enhanced Machine Type Communication (eMTC) and Best-effort reliable low latency machine type communication (RLLMTC) traffic. This architecture is composed of a multidomain orchestrator, network slices, physical resources, a virtualization layer, VNFs, a VNF manager, domain orchestrators, and intelligent NS services. A multidomain network orchestration architecture incorporates an operation support system and a business support system (OSS/BSS), a slice manager, an SDN controller, and intelligent NS management. The latter plays a key role to connect the underlying multidomain network infrastructure and create heterogeneous network slices for 6G services.

Several works study multidomain NS in terrestrial networks [20], [75]. Kukliński et al. [75] proposed a modular architecture to create multidomain slices as a combination of single domain ones and independently of the underlying technology. They assume that slices have their own MANO support embedded on the slice. In [20] a multidomain MANO architecture is presented for federated resource control. The architecture consists of four major strata: 1) multidomain service conductor stratum; 2) domain-specific fully fledged orchestration stratum; 3) subdomain MANO and connectivity stratum; and 4) logical multidomain slice instance stratum.

B. Network Slicing in STECNs

A slice scheduling algorithm is proposed in [40] to access satellite servers for various IoT application demands, such as mission-critical IoT, massive IoT, and 6G backhaul. They formulate the offloading problem as a multiobjective optimization problem with respect to transmission and computational power, and latency constraints. The problem is solved using a heuristic algorithm with time-varying satellite constellations topologies. Kak and Akyildiz [7] presented an autonomous NS framework for ultradense CubeSat systems. Their objective is to minimize the violation of the SLA associated with the slice. The slices are created by finding the optimal gateways and CubeSats needed to route the traffic and the required resources to serve each traffic class.

Drif et al. [69] integrated satellite systems with 5G as a transport network between core networks and the 5G terrestrial radio access. They focus on network design aspects in addition to network MANO solutions for this integration. They introduce a slice classifier and a satellite management network, and present primitives for slicing the space segment. For example, the LEO satellite is used by MIoT slice and GEO is used by 5G enhanced Mobile Broadband (eMBB) slice. A space-based EC system architecture customizable to support NS is presented in [76]. A resource management mechanism is developed to allocate resources based on the QoS requirements and specific application scenarios. In [77], E2E slicing of the satellite backhaul is discussed based on differentiated traffic needs and using LEO and GEO paths. Bisio et al. [78] presented a network-slicing solution focused on eMBB services within space–ground integrated networks. In [79], the concept of QoS was expanded to quality of experience, and the authors provided numerical evidence that NS can meet the varying demands of both users and different industries. Ahmed et al. [80] optimized resource usage in integrated 5G satellite networks through on-demand NS. Drif et al. [69] introduced a framework for slicing satellite networks, aimed at promoting the integration of terrestrial networks and satellite services. In addition, satellite networks can bring new use cases for NS as discussed in [26], [81], and [82].

NS in STECNs applied to vehicular networks is investigated in [83] and [84]. A space–air–ground integrated vehicular network (SAGVN) is presented to serve automotive services and meet ultrareliability low-latency communication (URLLC) requirements when transmitting bandwidth-intensive content [85]. Lyu et al. [83] investigated SAGVN and present an online control architecture for dynamically slicing SAG spectrum to deliver isolated vehicular services. Online decisions are made regarding slice admission and scheduling, resource allocation, and UAV dispatching for various services. They develop a network revenue function that integrates time-averaged system throughput and UAV dispatching costs, as well as all service queue backlogs. Using Lyapunov optimization theory, the goal is to maximize network revenue while achieving stability in the time-averaged queue. Wu et al. [84] investigated the problem of resource slicing and scheduling (RSS) to meet delay-sensitive services (DSSs) and delay-tolerant services (DTSs) in space–terrestrial integrated vehicular networks. RSS is used to allocate spectrum resources to distinct slices and decide user association and bandwidth allocation for particular vehicles. They first develop a joint RSS (JRSS) problem to minimize the long-term network cost, which includes the cost of slice reconfiguration, decision trees (DT) delay, and DSS requirement violation under dynamic network conditions. Since RSS decisions are interdependent with distinct timelines, they partition the JRSS problem into two subproblems with different time scales, and present an RL-based JRSS scheme to solve them.

C. AI-Based Network Slicing Design for STECN

The multitier nature of STECNs and the heterogeneous requirements that they are expected to accommodate can increase significantly the complexity of optimizing the network, and managing and orchestrating network resources. In this regard, AI becomes an enabler to slice STECNs. AI-assisted solutions applied to MANO are intended to address a variety of optimization challenges, including intraslice resource management, VNF placement, VN embedding (VNE), and resource allocation between slices, as

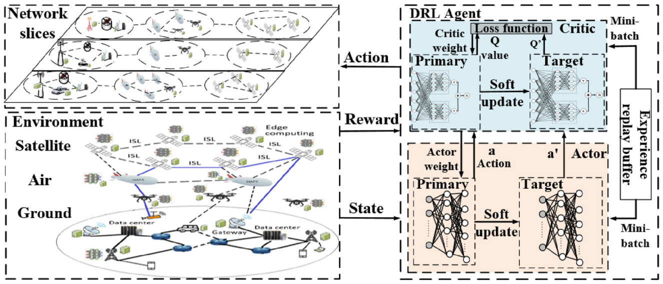

Fig. 4. Illustration of the DRL-based resource management for NS in STECNs.

illustrated in Fig. 4, where the deep reinforcement learning (DRL) learning agent observes the environment (state) and takes an action accordingly by the actor-network. Then, the agent receives the feedback (reward) and evaluates the taken action by using the critic network. The optimization problem for NS management in STECN can be solved by using centralized algorithms via one centralized learning agent or distributed algorithms through a collaboration among domain controllers in which each domain controller (agent) takes an action that maximizes the global reward. Lee et al. [86] jointly optimized the location of the UAV and the association between users, UAVs, and satellites via DRL. Rodrigues and Kato [87] presented a framework for NS to assign resources from both terrestrial and satellite networks using a machine-learning algorithm and ant colony optimization (ACO). The algorithm learns the patterns of user requests and the most frequently requested paths. By assigning a higher cost to these network paths, they will be allocated as a last resort. Their approach shows efficient resource management and significant improvements when compared to simpler NS strategies. However, the following aspects should be considered in designing AI-based techniques in NS for STECNs.

1) Frequent Topology Variations and Handovers: AI algorithms face difficulties in responding to dynamic and complex wireless contexts [81] and lack resiliency to adapt to dynamic environments [81]. We identify two types of variations.

1) *Predictable:* Good performance can be achieved in a dynamic environment when the input data is structured or has a particular probability distribution (e.g., the Poisson distribution for traffic requests or the Rician distribution for satellite-channel fading [27]). In [88], an NS orchestration and configuration scheme LEONS is presented based on restless multiarmed bandits (RMAB). The scheme relies on a belief in resource availability and SLA achievement to select the best slice configuration given the network condition and traffic requirements. The algorithm achieves fast configuration selection with low complexity. Kak and Akyildiz [7] introduced a novel topology construction mechanism based on Voronoi tessellation to map CubeSat constellations to a network topology. The resulting topology is used as the substrate to deploy different slices. Slice status information must be detected and reported in real time. However, the comparatively high propagation delay between terrestrial and satellite nodes may hinder real-time interaction between the control and data planes. Given that satellites move periodically, this kind of fluctuation is time varying but predictable [27]. Time-expanded graphs can be adopted to easily deploy existing AI methods directly.
2) *Unpredictable:* Traditional AI models [e.g., DRL and deep neural networks (DNNs)] work under ideal circumstances but unforeseeable NS management fluctuations may challenge their correct operation. For instance, unforeseen variations in network topology, unexpected user departure/arrival, bursty traffic needs, and drastic fluctuation in channel conditions are common in STECNs. When more tiers of satellites are considered (e.g., GEO and MEO), the problem becomes even harder. The reason for this is that a standard AI algorithm is highly dependent on observed data sets used for the training, and has limited generalization capability [89]. This leads to the problem of data drift in ML [90] in which the new inputs may no longer be relevant, resulting in unpredictable input–output relationships. As a result, the learning performance (e.g., convergence, accuracy, and loss) decreases in the new environment. In general, there are two approaches to dealing with the effects of data drift: a) passive and b) active [90]. The former updates the AI model on a regular basis by retraining it on the most recently observed data sets, regardless of whether the update is required. The latter approach evaluates first the need for the update and then retrains the model using a concept-drift trigger mechanism. Data drift is one of the most significant barriers to implementing AI in practical B5G systems [90]. Besides, they suffer from catastrophic forgetting, and thus the model needs to be retrained for every new task, which is time-consuming and inefficient [10]. Some progress has been made to improve the response time of learning models to dynamic environments by transfer learning [89], meta-learning, and lifelong machine learning (LL) [90], [91], [92].

2) Large-Scale System: It is hard for AI techniques to attain sufficient performance in real operations when network scale and service-oriented needs expand. Many open difficulties arise, including how to collect and analyze vast volumes of labeled data, how to extract features from raw data, and which particular features should be extracted.

3) Limitations in Network Resources: To select the right AI model to solve a specific problem, a tradeoff analysis between cost and AI performance is needed. Multidimensional resources (e.g., communication, computation capabilities, and flexible resource configuration) and their cost in NS must be considered. For instance, using cloud computing resources to run AI algorithms will result in higher cost and security concerns, and may be still profitable depending on the application. Likewise, this tradeoff analysis is needed when using AI algorithms in Earth-observing satellites since energy cost is critical to preserve their autonomy and acquire more images [93].

4) High Performance Demands: ML has been adopted in satellite communications for energy management [94], anti-jamming [95], beam hopping [96], and among others. Due to the relatively high propagation delay in satellite data transmission, the services offered by different slices may be better suited for latency-tolerant communications, such as eMBB or massive MTC services. Currently, AI's intrinsic limitations make it difficult to extend their application to safety-sensitive communications, such as mission-critical communications, remote healthcare, and autonomous vehicles. For further exploration the previous settings, excellent prediction accuracy, adequate problem representation, and error-tolerant frameworks are required. To perform online network slice management, an AI algorithm with capability of fast response and resilience is critical. In reality, an AI algorithm may have a long convergence time [97], resulting in long-term performance degradation. This is because dynamic circumstances can lead to rapid parameter changes over time (e.g., link failure probability, capacity, and delay). In fact, the selected AI model needs to be regularly refined and updated in the new environment. Otherwise, it may become sensitive to forthcoming unknown facts.

D. Technical Challenges of Network Slicing in STECNs

Slicing in STECNs presents the following challenges.

1) Multiple Service Providers and/or Administrative Domains: Each slice encompasses resources from multiple service providers (SPs) and/or administrative domains, as shown in Fig. 6. Accommodating the requirements of different services with the infrastructure and resources owned by multiple stakeholders requires efficient MANO schemes. Multidomain NS requires collaboration between different entities for resource sharing and fulfilling requirements across multidomains. Besides, a multidomain control plane architecture is needed to assure E2E performance throughout the slice life cycle. Some of the challenges in implementing fully autonomous resource MANO solutions stem from the lack of standardized interfaces across component vendors. To palliate that, satellite operators are using software-defined satellites that enable control and management on-the-fly as already used for the terrestrial and aerial components [98], [99]. Besides, SDN enables radio updates, such as the ability to change transmission power, and spectrum bands and facilitates multitenant support. By using network virtualization and slicing, SPs can share their network resources to achieve E2E connectivity. EC providers in STECNs provide local computing and storage resources for different applications and the ability to exchange real-time information on local network conditions for control and management purposes, as detailed in Section II.

2) Service Level Agreement Decomposition: The SLA needs to be decomposed dynamically into partial SLAs that each of the domains can support based on their network condition. However, it is challenging to guarantee the SLA since encompasses different domains with heterogeneous resources, different traffic dynamics, and wireless characteristics. In addition, the orchestrator may not have information on the state of available resources per domain. Existing works on SLA decomposition focus on the terrestrial network. De Vleeschauwer et al. [100] built a risk model based on the responses of each domain controller to previous SLA decomposition requests. They decompose the E2E SLA based on the estimate obtained by the risk models of all involved domains. Kovacevic et al. [101] presented a network latency equalization scheme that redefines the delay requirements in each network domain depending on the conditions in the other domains. A delay larger than a threshold (debt) is allowed in certain domains if other domains can compensate it by transmitting the messages with less latency (credit).

3) Vulnerabilities to Failures: Performance maintenance throughout the slice life cycle is more demanding given unpredictable network dynamics or failures that can compromise many slices which in turn will impact many services. A few works evaluate the vulnerabilities of NS in terrestrial networks [102], [103], [104] and propose solutions to improve its resiliency [105], [106], [107]. In [105] a latency-aware NS scheme is presented to provide resiliency by using backups for various network components. They propose a genetic algorithm to find a backup strategy. In [106], a service shifting mechanism is proposed to upgrade or downgrade a VNF graph consisting of a set of VNFs to react to infrastructure problems. In [107], a dynamic redundant slice allocation strategy based on channel fading characteristics is presented for error resiliency of video transmission over wireless channels. These works focus on providing resilience to a specific type of failure or traffic class.

4) Privacy and Security: Classical machine learning algorithms, such as DT, random forest (RF), K-Means, and Bayesian network have been used to detect network attacks [108], and recently to detect NS attacks [102], [103]. In [102] the Denial-of-Service (DoS) attack is analyzed in NS. The resource orchestrator uses a learning-assisted algorithm to learn the slice performance utility. The DoS attack is detected when the efficiency of the resource utilization changes dramatically, and the impact of the attack is mitigated by reducing the resource allocation to the slice. A cooperative anomaly detection algorithm based on transfer learning is presented in [103] to detect the anomalies of nodes in a slice. To determine its cause, the anomaly is classified as normal, degradation, deterioration, and outage. In [104] a privacy-preserving authentication protocol is developed to mitigate network slice topology learning attacks by protecting users' service access behavior from third-party providers. In [109], the vulnerabilities of ML models to adversarial attacks are highlighted together with the importance to assess the robustness of the models before integrating them into deployed systems.

E. Satellite Projects

Many U.S. agencies (e.g., DARPA, NSF, and NASA) have programs to exploit space communication. DARPA has directed its effort toward adaptable optical communications to interconnect different constellations into a resilient "space layer" for the Internet of small satellites [110]. NSF has the Spectrum and Wireless Innovation enabled by Future Technologies (SWIFT) program that considers effective spectrum utilization in satellite communication systems [111]. NASA establishes the Near Space Network that provides data and communications, telemetry, commanding, and ground-based tracking services to a wide range of customers with satellites (LEO, GEO, and MEO) [112].

Integrated satellite–terrestrial systems for ubiquitous beyond 5G communications has been examined in [113] to create an essential shift in the current 5G wireless networks toward secured, self-organized, intelligent, and ubiquitous satellite–terrestrial integrated network exploiting ground-breaking satellite communications technologies. A proof-of-concept testbed for satellite–terrestrial integration in the 5G Context has been investigated by SATis5 Team [114]. In [115] funded by European Commission Horizon 2020 program, satellite integration into 5G has been studied with a focus on satellite backhauling solutions. Building upon synergies with SATis5, the 5G-VINNI experimentation facility site hosts satellite 5G testbed nodes equipped with SDN/NFV/MEC capabilities and enables eMBB and mMTC traffic over satellites. In [116], an autonomous NS is investigated for integrated satellite–terrestrial transport networks. EU research projects, such as 5G!Pagoda [117], and SliceNet [118], as part of the European Commission Horizon 2020 program has proposed a dynamic creation, orchestration, and management of network slices for different use cases on multidomain 5G networks.

F. Standardization

The 3rd Generation Partnership Project (3GPP) standardization efforts for the integration of terrestrial and nonterrestrial networks (NTNs) are discussed in detail [119]. Ongoing efforts focus on defining NTNs with interoperable interfaces in order to achieve seamless connectivity [120]. Regarding the standardization process for multidomain NS, ETSI NFV MANO framework is working on transitioning from core NS toward peer-to-peer cross-domain orchestration and management [121]. The report provides recommendations and discusses different slicing architectures for multidomain MANO. The 3GPP standardization efforts for IoT traffic support are summarized in [25].

G. Summary and Discussion

This section comprehensively reviews existing works on slicing STECNs, including AI-based solutions, and the latest research projects and standardization efforts in satellite–terrestrial systems are summarized. Further investigation is needed to address the challenges of slicing STECNs due to their large scale, multidomain and dynamic nature, and vulnerability to failures. Although multidomain slicing under traffic uncertainty has been studied in terrestrial networks [20], [70], [75], new slicing schemes and SLA definitions that account for mobility, heterogeneous resources, failure recovery, and multiple SPs are needed in STECN.

IV. Resilient Networking and Slicing in STECNs

In this section, we present strategies to achieve network resiliency and discuss techniques for resilient communication and computing, and resilience metrics. Some of these techniques will be adopted in Section V to design resilient NS.

A. Type of Failures and Strategies for Resilient STECNs

STECNs face challenges imposed by the dynamics of the network topology, the intermittent visibility of ISLs, and UAVs operation, which makes them susceptible to external disruptions. In addition, failures due to natural or human intervention are unavoidable in large networks. A failure at any point in the terrestrial–air–satellite segments propagates throughout the network and it will affect multiple slice domains, disrupting the service. Hence, resilience is crucial in STECNs to avoid disrupting the service for many applications, such as data collection and offloading [122], rescue operations [123], security and surveillance [124], etc. To design strategies that guarantee connectivity in STECNs, the following aspects must be taken into account.

1) A large number of nodes (IoT devices, UAVs, and satellites) can be compromised, i.e., under massive destruction.
2) Re-establish the communication may need a lot of resources which is costly and even so may not be suitable for real-time executions.
3) Computing task execution and storage may be compromised too.
4) Failures may be predictable if they follow a failure pattern, in which case they can be mitigated or even avoided. On the other hand, unpredictable failures could compromise network performance if not handled carefully.

We can also classify the failures based on their duration as permanent, intermittent, and transient, as shown in Fig. 5, with the following characteristics.

1) *Permanent:* A permanent failure of a terrestrial, UAV, or satellite node occurs due to natural or human-induced disruptions, and it results in a permanently unavailable node and unavailable links toward that node. We assume that the failure is permanent when it lasts longer than a predetermined threshold.
2) *Intermittent:* Intermittent failures are associated with loss of communication links due to mobility and packet drops, unavailability of UAVs due to interruption in power supply, etc.
3) *Transient:* Transient failures are identified using detection mechanisms frequently based on information redundancy, such as error correction coding (ECC) and their mitigation is often based on retransmission (timing redundancy). Many of these solutions are cross-layer as detection and mitigation typically happen in different layers with communication between them. After a period of time, the failure is mitigated and normal communication is resumed.

The strategies to achieve network resiliency are broadly categorized as follows.

1) Proactive Protection Strategies: Prepare the networks before a disaster occurs. Over-provisioning resources to prevent failure at VNs in the substrate network is discussed in [125]. To rebalance the load on the substrate network after the arrival or departure of VN requests, Gao and Rouskas [126] proposed dynamic re-embedding of both virtual nodes and links. Moreover, several works develop VNE

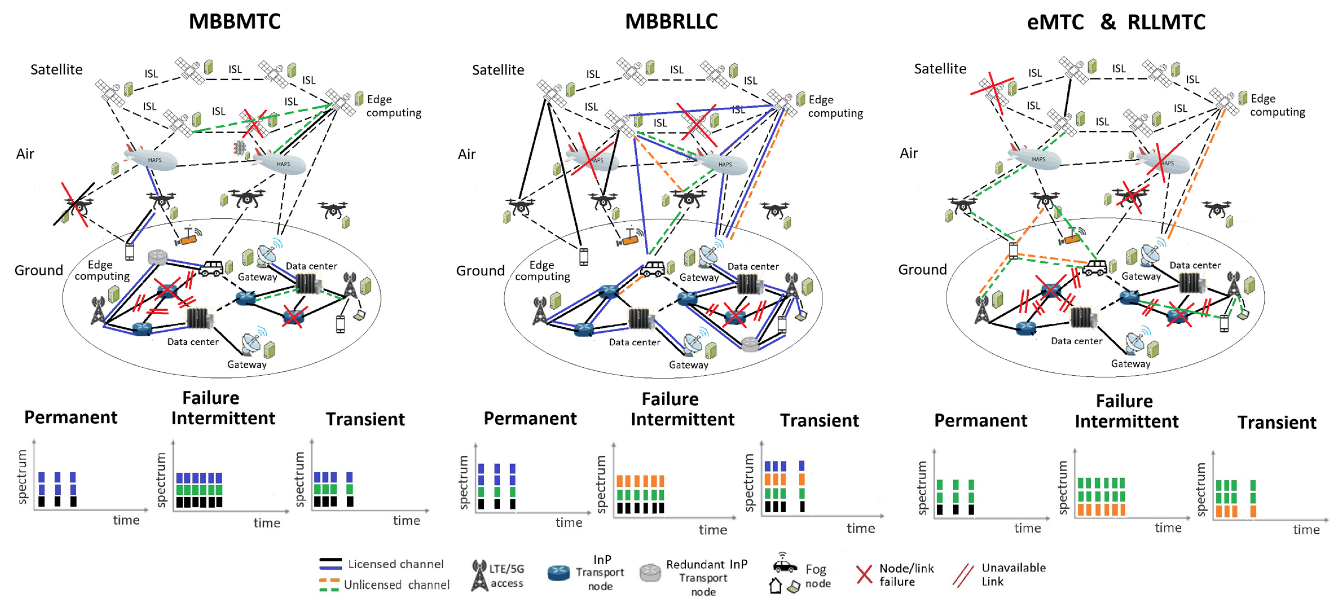

Fig. 5. Slice reconfiguration for MBBMTC, MBBRLLC, eMTC, and RLLMTC under permanent, intermittent, or transient failures.

schemes that can guarantee connectivity when multiple failures occur [127]. These mechanisms are effective for predictable failures but there is no guarantee after massive, and unpredictable failures. Shahriar et al. [128] focused on slicing 5G transport networks while providing dedicated protection by using backup resources. To reduce the idle backup resources, they adopt the following techniques: 1) bandwidth squeezing to reduce the protection bandwidth compared to the original request and 2) survivable multipath provisioning to leverage the capability of elastic optical networks to fine-tune spectrum allocation, adapt the modulation, and perform error correction for allocated spectrum resources.

2) Reactive Post-Failure Strategies: A recovery process is needed after failures occur with recovery actions designed to optimize the performance. Zhang et al. [129] studied progressive slice recovery by optimizing a sequence of reparation steps to recover the slice as a result of node or link failures. The recovery strategies include changing slice embedding, and/or enforcing different options of slice connectivity. The best recovery sequence is obtained by solving an integer linear programming (ILP) problem. To achieve scalability, they devise a two-phase progressive slice recovery meta-heuristic algorithm with a small optimality gap. Thiruvasagam et al. [105] leveraged multiconnectivity in resource-limited MEC cloud facilities to improve the resiliency of the system and orchestrate the deployment of network slices. They aim to achieve ultralow latency requirements and guarantee service continuity under: 1) failure of communication links; 2) failure of VNFs; 3) failure of local servers within MEC; and 4) failure of an entire MEC cloud facility at the regional level.

3) Adaptive Strategies to Prevent Future Failures: A scalable DRL-based routing scheme for SDN (ScaleDeep) is proposed in [130], which is resilient to network topology changes. The scheme automatically adjusts the routing policy to improve network performance. A set of critical nodes are selected as driver nodes to improve the algorithm's convergence. Yuan et al. [131] studied how to schedule resources efficiently in an over-loaded LEO-terrestrial network to serve a large number of users and provide resilient communications given the dynamics of the wireless environment. They propose an enhanced meta-critic learning algorithm (EMCL) to adapt to nonideal dynamic environments and show their effectiveness and fast-response capabilities. A user association mechanism is presented in [132] to balance the accessible capacity of satellite–terrestrial networks under intermittent connectivity and dynamic backhaul capacity. They adjust the network load by jointly considering BSs traffic load, task classification, and backhaul capacity of LEO-based BSs. A dynamic cell range extension algorithm is developed to achieve resilient backhaul capacity and meet the users' task processing requirements. Yang et al. [133] developed a resilient networking architecture for LEO remote sensing. They propose path-quality aided and lifetime-aware dynamic routing for enhancing the robustness of routing against topology changes. To achieve resilience against ISL intermittence and mobility, a hop-by-hop data transmission algorithm is proposed together with data caching for resilient satellite access switching.

B. Solutions to Enhance Resilience of Communication and Computing in STECNs

We discuss solutions to improve resiliency in communications and computing that can be adopted by any STECN integration architecture.

1) *Context-Aware Protocols:* Given the different distribution of terminals, propagation characteristics, and working conditions, efficient communication protocols must be context-aware. The context-awareness includes information related to mobility, traffic, and available resources which vary at each segment in STECNs. To integrate STECNs and provide resilient communication and computing, the following protocols are needed.

1) *Routing:* Routing algorithms will use context-aware information to route the tasks to EC servers based on their load, the anticipated duration of the server availability (especially at UAVs and satellites), and the available spectrum. Since controllers are only aware of the availability of their local resources, routing decisions must be distributed and avoid frequent reconfigurations. Relay selection schemes are investigated in [134] for HSTRN to improve the outage and network capacity. With the incorporation of terrestrial relays, spatial diversity can be exploited to improve the system's performance.
2) *Mobility Management:* Since domains have different mobility patterns, mobility management must be integrated to avoid intermittent transmissions. Liu et al. [28] reviewed mobility management, the challenges in the integration of satellite and terrestrial networks, and their physical-layer characteristics.
3) *Resource Management:* Spectrum is a scarce resource in both satellite and terrestrial networks. The cognitive radio (CR) technique is adopted in CogSTN to enable dynamic spectrum sharing among both networks [135]. The outage probability in the CogSTRN is analyzed in [37] and [136]. In [37] power control schemes are presented to control the interference caused to the primary terrestrial network and the nonorthogonal multiple access (NOMA) technique is applied in [136] for simultaneous transmission in the primary and secondary networks. In addition higher frequency bands, such as mmWave and terahertz, have attracted much interest to improve the spectrum efficiency by spectrum sharing [137], and thus, these works are explained next in a dedicated section.
4) *Load Balancing and Traffic Offloading:* They are crucial to avoid uneven service distribution and improve the wireless backhaul capacity of satellites which is shared by a large number of BS within a certain coverage. The performance evaluation needs to analyze the tradeoffs among latency, throughput, energy, and spectrum efficiency to address different traffic offloading decisions. Multimedia traffic offloading schemes are presented in [138]. They use the wide coverage of satellites to offload popular content for multicast/broadcast transmission, which reduces congestion and uses network resources efficiently. Latency is particularly important in the design of dynamic offloading schemes due to the long propagation delay in the satellite–terrestrial transmissions [139].

2) *Cooperation Between Both Networks:* As in CooSTN architecture, terrestrial and satellite networks can collaborate to serve users with dual-mode terminals based on their service requirements or signal intensity. Besides, simultaneous transmissions in both networks can be utilized to improve the capacity and reliability of the transmission by spatial diversity. Multipath communication enhanced with network coding is used in [140] to maximize the system throughput when users can transmit simultaneously to the satellite and the cellular BS. The airborne network can facilitate collaboration between satellite and terrestrial networks by providing connectivity to users out of their coverage range [141]. The backhaul transmission is enabled by either satellites or terrestrial BSs. However, the design and optimization of airborne networks (e.g., HACN and terrestrial–airborne–satellite) is more challenging than in other STECNs architectures due to their time-varying characteristics, heterogeneity, and self-organization capabilities.

3) *Service Migration:* It facilitates service continuity when the user moves from the coverage area of one MEC platform to another. When service migration occurs, the initial platform forwards the application data to the new platform. Alternatively, premigration can be performed which ensures that the migration is completed as the user enters the new service area and minimizes the impact on the computing task. To conduct efficient service migration, finding the right migration time is crucial. Next, we analyze service migration in each domain.

1) *Satellite Networks:* Satellites follow specific orbits and their movement is periodic. Thus, their positions at each moment and their route are predictable. Satellites can perform premigration and determine the route in real time based on the position of the user and nearby satellites. Satellites should complete the service migration before the user enters the coverage area of the next satellite.
2) *Air Networks:* Unpredictable external destructions (UEDs) may occur in air networks when performing service migration. To re-establish the communication, a self-healing method based on swarm networks is proposed in [142]. They present a graph convolutional neural network (GCN) to obtain the recovery topology online for one-off UEDs. Then, they develop a GCN-based trajectory planning algorithm that can be applied to general UEDs and makes UAVs re-establish communication during the self-healing process. The online executions of the GCN follow a meta-learning scheme.
3) *Terrestrial Networks:* Since the main terrestrial network infrastructure, such as access points and BSs are fixed, the service migration focuses on predicting the users' movement trajectory.

Service migration is also an effective way to recover from network failures. If the channel state suddenly changes or the available computing and/or the storage of the current node are insufficient to meet the requirements of the computation task, it should be migrated to another node.

4) *Computation Task Reinstantiation:* If the computation task cannot be repaired it will need to be reinstantiated. This is the case when a node or container running the task suddenly fails. The task will be reinstantiated upon retransmission of the computation request.

C. mmWave and THz

The LEO satellites currently operate in Ku-band, Ka-band, and Q-/V-bands. It is expected that to fulfill the increasing data rates demand, higher frequencies in W-band and Terahertz (THz) band [143] will be used. At THz frequencies, there is a huge bandwidth available and antenna gains are highly directional with razor-sharp beams able to achieve ultrahigh throughput in near-Earth and deep space orbits. Besides, they provide secure communication with reduced eavesdropping [144]. Nie and Akyildiz [145] analyzed the channel propagation effects and the capacity of ISL channels at THz band to demonstrate its feasibility. At these frequencies, outages may occur due loss in the communication caused by rain, cloud, and atmospheric gases [146]. The impact of molecular absorption loss and noise, thermal noise, and noise due to the receiver circuit and antenna temperature on the SNR is studied in [147]. Xing and Rappaport [148] conducted measurements at 140 GHz in a rooftop surrogate satellite/tower BS. They show that by using elevation angles ≤ 15 for terrestrial emitters, the interference in the same or adjacent bands can be neglected at frequencies above 100 GHz when the following transmissions coexist: 1) passive satellite sensors and terrestrial terminals or 2) mobile links and terrestrial backhaul links.

Several works study the advantages of incorporating mmWave [149] in different integration architectures as discussed in the following.

1) Integrated Terrestrial–Satellite Network: Peng et al. [32] investigated user scheduling, hybrid beamforming, and resource allocation optimization in ITSN to improve transmission rate and energy efficiency. They present a spectrum-sharing scheme between satellite transmissions in the Ka-band and terrestrial links in mmWave frequencies. A hybrid analog–digital beamforming transmission scheme is presented with low complexity for onboard processing. They solve the optimization problem by minimum mean square error (MMSE) criterion and logarithmic linearization.

2) Cognitive Satellite–Terrestrial Network: Zhao et al. [150] considered a CogSTN in which the primary satellite network and the secondary terrestrial network transmit in the same mmWave bands. In the secondary network, there are multiple legitimate receivers (LRs) and an Eavesdropper (Eve) with imperfect channel state information (CSI). They study the optimization of the power allocation to maximize the secure energy efficiency (SEE) for NOMA in mmWave (mmWave-NOMA). The problem is nonconvex and is solved iteratively using successive convex approximation, a semidefinite program by Dinkelbach, and S-procedure technologies. In [151] the coexistence between a terrestrial mmWave cellular network and a GEO satellite network is investigated. They maximize the Ergodic capacity of the terrestrial users subject to the probability of causing interference to the satellite users. The problem is solved by a new virtual uplink-based beamforming algorithm.

3) Satellite–Terrestrial Relay Network: The downlink transmission in STRN is studied in [152] with a mmWave decode-and-forward (DF) relay. The satellite-to-relay link and relay-to-user link adopt SR fading and Nakagami-m fading, respectively. They derive the signal-to-interference-to-noise ratio (SINR) for both links and use the meta-distribution to obtain coverage information, i.e., the probability that a percentage of satellite-to-relay links and relay-to-user links are able to reach a target SINR threshold.

4) Satellite–Terrestrial Backhaul Network: Wang et al. [153] investigated beamforming schemes for STBN when the satellite backhaul and terrestrial access links operate on the same mmwave band. They aim to maximize the transmission rate subject to power requirements of the satellite and the BS, terrestrial users' QoS requirements, and satellite backhaul capacity. They transform the problem into a standard difference of convex problem under perfect CSI. Then, an iterative penalty function algorithm is derived to obtain suboptimal beamforming vectors when only imperfect CSI is available.

5) Hierarchical Aerial Computing Network: UAVs can significantly enhance the coverage and QoS of the terrestrial mmWave cellular networks either by serving as aerial access points or relays [154]. By packing large antenna arrays in a small area on the UAV, 3-D beamforming can be performed. Xiao et al. [155] surveyed the latest works on mmWave-UAV communications and networking. They discuss its potential and technical challenges and present an overview of mmWave antenna structures and channel modeling.

D. Resilience Metrics

Resilience is the ability of the network to provide and maintain an acceptable level of service in the presence of failures and challenges to normal operation [156]. The metric Quality of Resilience (QoR) refers to the service resilience characteristic, such as service continuity, downtime, or availability which are defined in the long term [157]. Unlike the QoS metrics, the QoR is not perceived directly by the users since users cannot distinguish if an increase in delay is due to congestion or network failures. The International Telecommunication Union–Telecommunication Standardization Sector (ITU-T) proposed several resilience metrics [157], [158]. Below we summarize some of these metrics and present new metrics related to resilience in STECNs.

1) *Instantaneous (Un)Availability:* Probability for a network element (node/link) of being (un)available a given time.
2) *Mean Time Between Failures/Interruptions:* Mean time between two consecutive failures/interruptions of a repaired element.
3) *Mean Time to Failure/Recovery:* Mean value of time since the network element recovered/failed until the next failure/recovery.
4) *Retainability:* Probability of continuing to provide a service.
5) *Packet Loss Ratio (PLR):* Ratio of the total number of packets lost and the total number of transmitted packets.
6) *Service Availability:* A fraction of the total service time in which the PLR is above a predefined threshold. Similarly, the service availability can be defined in terms of a threshold latency.
7) *Risk of Nonsatisfying the Service:* Probability that the service does not satisfy the service requirements defined in the SLA.
8) *Effectiveness of Failure Prevention:* Probability that the failure prevention mechanism is effective. It is

given by one minus the risk of nonsatisfying the service.

9) *Effectiveness of the Redundancy Strategy:* Probability that a failure downgrades the performance while using redundancy.

10) *Efficient Resource Utilization:* Measures the amount of additional resources allocated to achieve a low probability of transmission failure even without redundancy to make efficient use of the network resources.

E. Summary and Discussion

In this section, we review design aspects for resilient communication and computing in STECNs. In particular, we classify the types of failures based on their duration and predictability, and design strategies to protect the network and prevent future failures. The availability of higher frequency bands and their advantages in different STECN integration architectures are described together with resilience metrics. This review serves a solid basis for further research on strategies for resilient NS. For instance, proactive protection strategies such as overprovisioning can be adopted for interdomain slicing to avoid reconfiguring E2E routes whereas reactive strategies can be used intradomain to adapt to local changes.

V. Resilient Multidomain Network Slicing Scheme for STECNs

In this section, we present a resilient NS design to mitigate failures and achieve the SLA. We present solutions to cope with different types of failures and analyze their impact on the performance of different traffic classes. In addition, we present cross-domain solutions that include resource planning and provisioning, adaptive MANO, and collaboration between different domains.

A. Background and Motivation

Resilient NS is crucial to maintain acceptable performance in the presence of failures. In fact, network failures can cause significant disruptions in NS since the infrastructure is shared by multiple slices. Thus, failures in a node/link can compromise many services. In addition, service provisioning in STECNs involves different administrative domains in which providers interact to facilitate network resources (communication, computing, and storage) and meet the slice requirements. Since STECNs are dynamic and failures may occur, the level of performance degradation tolerated by each service must be specified in the SLA. In particular, the SLA must include the following specifications in STECNs: 1) service requirements (e.g., required bandwidth and delay); 2) service aspects (e.g., multiband selection, energy efficiency, and MEC); 3) design aspects (e.g., controllability, failure recovery, security, and autonomy); and 4) legal aspects to specify which entity is responsible for service degradation and the corresponding compensation.

In Fig. 6, we illustrate a multidomain NS scheme where tenants request slices from any traffic class described in Table II that correspond to typical application scenarios in STECNs [159]. MBBMTC supports high broadband data rates along with massive connectivity [e.g., in-space cellular backhaul remote connectivity (ICB)]. MBBRLLC offers high broadband data rates along with reliable, and low-latency communication [e.g., remote industrial automation (RIA)]. eMTC supports massive connectivity, and latency-tolerant communication (e.g., MAR). Finally, RLLMTC is characterized by low throughput, low latency, and massive connectivity (e.g., EComm). Each traffic class has different requirements and priorities in terms of minimum throughput $T_{\min}$, and maximum delay $d_{\max}$ given by ϑ and ε, respectively, and reliability requirements ρ that the tenant specifies when making the slice request. The multidomain NS scheme shown in Fig. 6 consists of the following components.

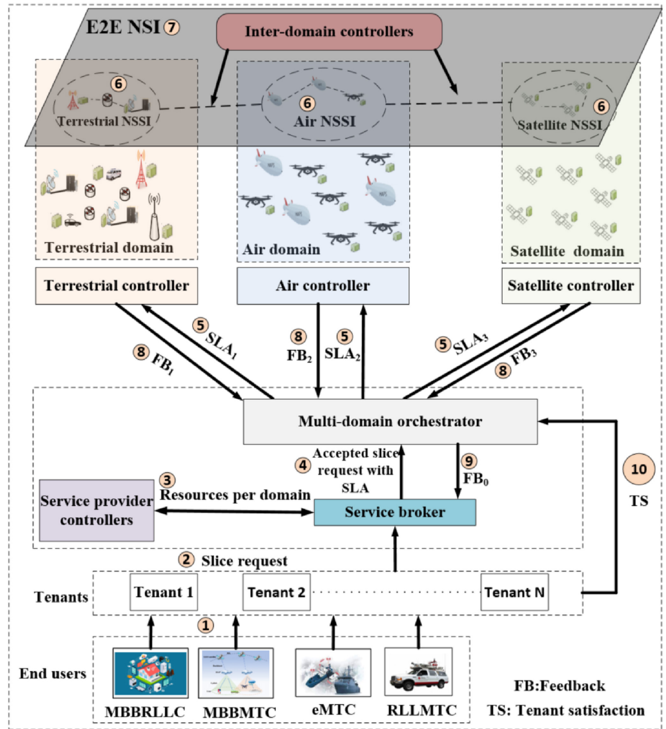

Fig. 6. Multidomain NS architecture.

TABLE II
SERVICE REQUIREMENTS FOR DIFFERENT CLASSES

Class	ϑ	ε	T_{min} (Mbps)	d_{max} (ms)	ρ (%)	packet size (kbit)	I_c	λ_c	μ_c
EComm	0.3	0.7	40	150	99.99	1	9	12	4
ICB	0.5	0.5	100	600	99.9	250	6	8	4
RIA	0.2	0.8	120	300	99	20	7	10	4
MAR	0.7	0.3	50	900	99.9	250	4	6	4

1) *SPs:* Own the network infrastructure and provide resources, including communication, computing, and storage.

2) *Tenants:* Request slices to serve their subscribers' service demands.

3) *Service Broker:* Conducts admission control and negotiates the price for the requested service with the SPs.

4) *Multidomain Orchestrator:* Analyzes the service requirements, decomposes the SLA among the domains, and

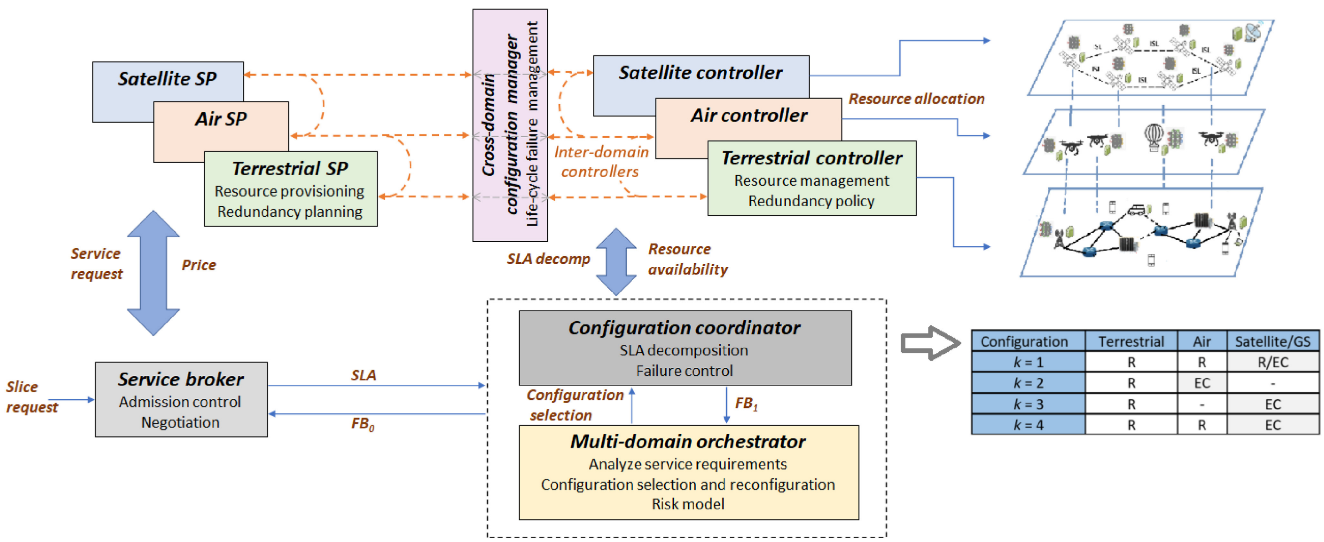

Fig. 7. Resilient multidomain NS model.

informs the service broker on the risk of nonachieving the SLA.

5) *Domain Controllers:* The intradomain controllers perform resource management per domain and interact with the interdomain controllers to allocate resources to meet the E2E SLA.

The process of serving the slice request with SLA guarantees consists of the next steps.

1) Tenants receive their user' traffic requests with different performance requirements.
2) Each tenant requests the slice and negotiates with the service broker the SLA and the price for the service.
3) The service broker interacts with the SPs to determine the suitable amount of radio, computing, and storage resources needed and the price. The SPs collaborate to reach the SLA and provide the necessary resources.
4) The service broker forwards the accepted slice request with its E2E SLA requirements to the multidomain orchestrator.
5) The multidomain orchestrator decomposes the E2E SLA into partial SLAs using a *decomposition rule* that indicates how the E2E SLA is decomposed per domain. For instance, an E2E delay requirement can be decomposed as the sum of the delays per domain, while the E2E throughput can be obtained as the minimum throughput in all domains. Similarly, the SLA parameters related to design aspects, such as failure recovery time, can be decomposed through different technologies and operators to guarantee link availability with a high probability. The service requirements depend on the traffic class, and should also be specified in the SLA.
6) The domain controllers allocate the resources to meet the partial SLA requirement and create terrestrial, airborne, and satellite network slice subnet instances (NSSIs).
7) The domain controllers and the interdomain controllers interact to create the E2E NSI by combining the NSSIs to serve the slice request.

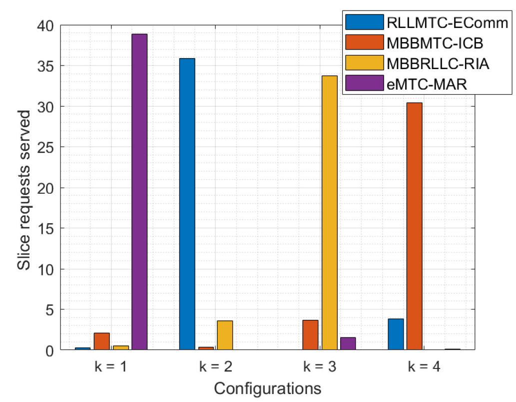

Fig. 8. Slice served per conf. and class for priority $\mathbf{O}$ = [0.1 0.5 0.1 0.2; 0.2 0.1 0.2 0.5; 0.2 0.3 0.5 0.1; 0.5 0.1 0.3 0.2].

8) after the slice is created, the domain controllers send feedback to the multidomain orchestrator regarding their achieved partial SLAs. The multidomain orchestrator uses this feedback to revise the decomposition rule if similar requests are received again.
9) The SLA is monitored during the slice lifecycle. In this process, the collaboration between SPs (step 3) will be revised according to the feedback received by the service broker. Likewise, the resource provisioning (step 3), and the decomposition (step 5) will be revised based on the feedback received in step 8.
10) Finally, the tenant assesses if the slice has met the service requirements and sends the tenant satisfaction (TS) feedback to the multidomain orchestrator. The multidomain orchestrator also considers this feedback to adapt the decomposition rule.

B. Resilient Network Slicing Architecture Design

In this section, we incorporate resilient considerations in our multidomain NS scheme, as shown in Fig. 7. To support heterogeneous service classes under different STECN integration architectures, our resilient multidomain design is agnostic to the STECN integration architecture. In fact, the integration

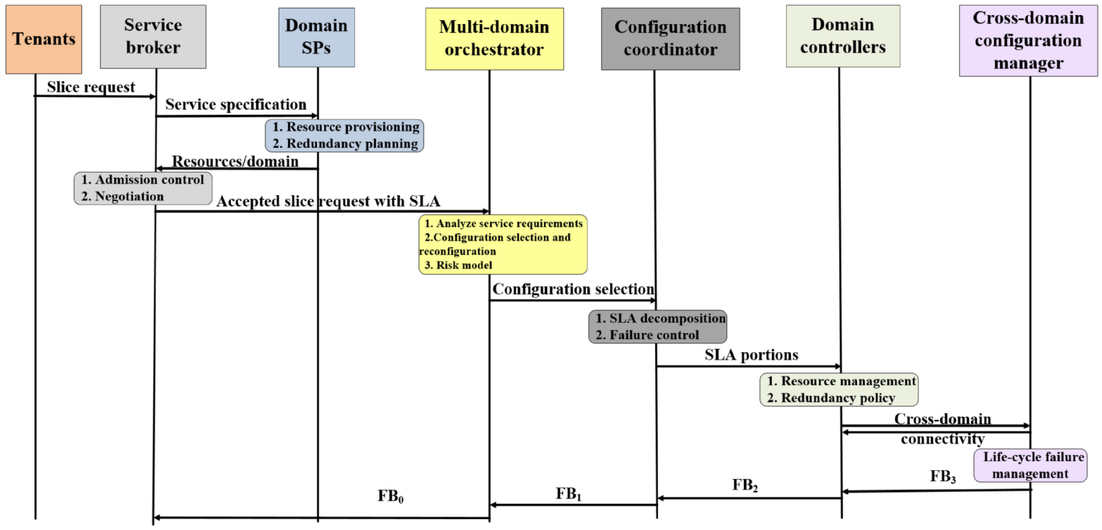

Fig. 9. Sequence diagram for creating a resilient slice in multidomain network.

architecture is abstracted in a slice configuration that defines a slicing topology with resources from different domains, i.e., terrestrial–air, terrestrial–satellite, terrestrial–air–satellite, and terrestrial–air–satellite–GS, and each domain may relay data (R) or perform EC. The configurations are described on the right table in Fig. 7. The multidomain orchestrator selects the configuration to serve a slice request depending on the traffic class and network condition, as shown in Fig. 8. For example, delay-tolerant data from an IoT device in ICB or MAR can be relayed by the satellite network to the gateway, and the computing task can be served at the gateway cloud (configuration 1). Likewise, intermediate UAVs (configuration 2) and satellites (configurations 3 and 4) can process delay-demanding tasks to reduce latency, such as in EComm or RIA. As previously discussed, each traffic class may have different tolerance to service degradation. Therefore, we incorporate redundancy planning and control strategies to cope with different failures and maintain acceptable performance. More details on these strategies are provided in Sections V-C and V-D. The resilient multidomain slicing scheme integrates the following new components and their functions, as described in the following.

1) *Configuration Coordinator:* Maps the service requirements by decomposing the E2E SLA into the domains involved in the selected configuration, monitors the performance of each configuration, and adapts the SLA decomposition accordingly to mitigate failures and achieve the E2E performance.
2) *Cross-Domain Configuration Manager:* Manages the resources per domain involved in that configuration, and interacts with SPs and configuration coordinator to perform life-cycle failure management.

The SPs, the multidomain orchestrator, and the domain controllers have additional responsibilities to the ones previously described. In particular, once the slice request is received, the SPs plan and provision resources and their backups (redundancy) to achieve the SLA. The multidomain orchestrator selects the slice configuration that minimizes the risk of nonachieving the SLA. Then, the configuration decision is forwarded to the configuration coordinator, which interacts with the domain controllers to decompose the SLA–assigns a fraction of the E2E SLA to each domain—and interacts with the cross-domain configuration manager that manages the resources in that configuration. Finally, the domain controllers allocate the resources and decide their redundancy policy to achieve the assigned SLA. The sequence of steps to create resilient network slices is shown in Fig. 9.

C. Network Planning, Provisioning, and Scaling

The domain SPs plan the resources based on the risk of nonsatisfying the SLA. Terrestrial, air, and satellite domains are prone to node and link failures. If not mitigated, they will compromise the SLA. To combat failures, redundancies must be provisioned not just at the physical infrastructure level, but also for each slice, so that the tenants do not experience any interruption. Network redundancy has been used to improve reliability in cognitive networks [3], [160], [161], reduce the delay [162], and increase robustness to failures in mobile networks [163]. Recently, a few works have applied redundancy to network slices [105], [107]. These works focus on a specific type of failure, e.g., server failure [105] or errors in the transmission [107]. Resource scaling is proposed in [164] to adapt resource provisioning to the network condition. In general, the domain providers can adopt the following types of redundancies.

1) *Hardware (HW) redundancy* incorporates terrestrial (fixed or mobile infrastructure), air, and/or satellite nodes to reroute the traffic avoiding compromised nodes, increase the spatial diversity to explore different available channels, and provide redundant computing and storage resources.
2) *Channel redundancy* increases the resilience when a channel becomes compromised or unavailable due to licensed users' activity. It can be used to transmit to the same or different nodes simultaneously. The redundant channels, licensed or unlicensed, are chosen based on the performance requirement and cost.

3) *Time redundancy* is useful when there is a transient failure. It consists of repeating transmissions in different slots. It can be combined with channel and HW redundancy to increase the chances to find an available connection.
4) *Information redundancy* increases resiliency to transmission errors due to loss or corrupted packets. By adding redundant bits to the original data, an error can be detected and/or corrected.

Ultimately, the slice broker makes the decision to accept/reject the slice request and revises its network planning and provisioning of resources based on the risk and the feedback from the controllers on their redundancy policies. Network planning and provisioning of redundancies to the network resources must consider the traffic class, failure rate, and the slice configuration. Slice redundancy allocation must be performed distributively by the domain controllers based on the decomposed SLA, but still transparent to tenants. If the resources are not longer needed they can be reused by other slice by dynamic resource scaling [164].

D. Traffic Aware Redundancy Policy

Each type of service has its own requirements which can be achieved by using a different amount and combination of redundant resources (i.e., spectrum, computing, and storage). We assume that tenants serve the 6G traffic classes [159] described in Section II and Table II with the following strategies.

1) *Mobile Broadband Machine Type Communication* [165], i.e., ICB, supporting high broadband data rates along with massive connectivity. This application transmits a high data volume and it has a large packet size. Thus, achieving a high throughput is more relevant than reducing the latency in resource allocation, as indicated in Table II. The most preferred configurations to serve this traffic are configurations 4, 3, and 1, respectively. The following redundancy strategies can be adopted per failures type.

1) *Permanent Failure:* Low redundancy level with hardware redundancy consisting of terrestrial and UAV nodes (configurations 1 and 4) and satellite nodes with high capacity links using licensed channels, and edge and cloud computing resources (configuration 1); channel redundancy using licensed channels, and information redundancy to avoid missing relevant data. As illustrated in Fig. 5, when one node fails, redundant links are used to reroute the traffic using redundant licensed channels to reconnect the new nodes to the network and to access edge or cloud computing resources.
2) *Intermittent Failure:* The level of redundancy is adjusted based on how frequently the interruptions happen. In addition to the redundancy used in the case of permanent failures, we incorporate channel redundancy on unlicensed channels and timing redundancy. Therefore, the transmissions are repeated in subsequent time slots using different channels, as shown in Fig. 5. Additional information redundancy is used to avoid choosing a redundant link that is intermittently available.

2) *Mobile Broadband Reliable Low Latency Communication* [166], i.e., RIA services, offering high broadband data rates along with reliable, and low latency communication. Given the high frequency of automation commands, this application has a small packet size and high packet arrival rate. Achieving low latency is more important than high throughput. The most preferred configurations are configuration 3, 2, and 1, respectively.

1) *Permanent Failures:* High level of redundancy with hardware redundancy can be applied using terrestrial and UAV nodes (configurations 1 and 2) and satellite nodes (configuration 3), as illustrated in Fig. 5; channel redundancy using licensed and unlicensed channels, redundant computing nodes, information redundancy to avoid missing relevant data and additional information about the network condition for rapid reconfiguration.
2) *Intermittent Failures:* May adopt a high level of redundancy like hardware redundancy using fog nodes, channel redundancy using licensed and unlicensed channels, timing redundancy, and information redundancy as with permanent failures.

3) *Best-Effort Enhanced Machine Type Communication* [167], i.e., MAR services, supporting massive connectivity, and latency tolerant communication. These services operate on a much longer time scale and thus they have a higher latency threshold. The priority to achieve high throughput and low latency is the same. This traffic is mostly served by configuration 1.

1) *Permanent Failures:* The level of redundancy is adjusted to the available resources. Hardware redundancy using fog nodes at the terrestrial domain, UAVs, and satellites; channel redundancy using licensed or unlicensed channels as available, and information redundancy about the network condition. In Fig. 5, we show the case of a permanent failure of a fixed terrestrial node, a UAV, an HAP, and a satellite node. As the links become unavailable, the traffic is rerouted using available nodes from the same or adjacent domains, and redundant cognitive channels.
2) *Intermittent Failures:* As in the case of permanent failures, the level of redundancy is adjusted to the available resources. In this case, since the failure is temporal we use redundant unlicensed channels and timing redundancy, as shown in Fig. 5.

1) Best-Effort Reliable Low Latency Machine Type Communication: Reference [168], i.e., EComm services, which are characterized by low throughput, low latency, and massive connectivity. This traffic is mostly served by configuration 2 since it achieves the lowest latency of all configurations. To mitigate failures, we will add more redundancy than in the previous case to achieve low latency.

The strategies to mitigate transient failures are a combination of the strategies used for permanent failures and intermittent failures, and they are adjusted based on the duration of the failure. The overprovisioning of resources needed to provide redundancy brings an additional cost for having backup resources. Thus, the redundancy strategy must be obtained per traffic class and failure type to optimize the tradeoff between the risk of nonachieving the SLA and the redundancy cost.

E. Multidomain Management and Orchestration

The proposed NS scheme includes a slice MANO system with a multidomain orchestrator, a configuration coordinator, a cross-domain configuration manager per configuration, and domain controllers. They interact with each other using a feedback control scheme that checks the QoS provided and adapts the slice decisions to the network condition. The following slice decisions are performed iteratively.

1) Slice Configuration Selection and Reconfiguration: The multidomain orchestrator selects the most suitable configuration based on the analysis of the slice service requirements. The configuration coordinator sends feedback to the multidomain orchestrator periodically on the achieved E2E SLA in the selected configuration. If the SLA cannot be achieved with the current available resources and network condition, the multidomain orchestrator selects another configuration. The configuration coordinator revises the SLA decomposition and the domain controllers remap the resources accordingly to reconfigure the slice and achieve the new request. If the remapping of resources requires additional resources, the domain controllers interact with their providers to update the resource provisioning and reduce the risk of nonachieving the SLA. If unpredicted failures occur, slice reconfiguration must be prioritized based on service contracts.

2) SLA Decomposition: The configuration coordinator adapts the SLA decomposition to mitigate failures, considering the redundancy policy at the different domains. For instance, it allocates a higher shared of the E2E delay to the domain that needs more redundant transmissions.

3) Redundancy Policy: Defines the level of redundancy (i.e., number of channels and nodes), the type of redundancy (i.e., hardware, channel/timing, and information), the type of resources used (i.e., spectrum (licensed/unlicensed channels), computing, and storage), and who provides the resources (i.e., terrestrial, air, and satellite). The controllers determine the redundancy policy for each traffic class and failure rate based on the assigned SLA. For instance, a controller may decide to use hardware redundancy when the number of errors is high instead of channel redundancy since the additional bandwidth will be wasted in transmitting information with errors. Likewise, it may decide to increase the amount of *metadata* information transmitted, which contains information related to congestion, reputation, routing, and topology, to verify the current network situation. Controllers will take action to improve the allocation (e.g., consider past demands in future decisions, update detection, mitigation mechanisms, etc.).

4) Life-Cycle Failure Control: The cross-domain configuration manager monitors the resources per domain in that configuration, and readjusts the resources to meet the SLA. If the current resources cannot meet the SLA, it interacts with configuration coordinator which modifies the SLA decomposition to compensate for the failures. If the current available resources under that configuration are still not enough, the multidomain orchestrator will change the configuration selection. Ultimately, if the current available resources are not enough to achieve the SLA, the cross-domain configuration manager will request more resources to the SPs. The allocation of additional resources is supervised by the service broker.

F. Hierarchical SDN-Based Multidomain Slicing

The feedback control scheme for MANO presented above facilitates the design of distributed learning algorithms to automatize management decisions. To implement an adaptive slicing management solution we consider a hierarchical network management architecture with four layers of SDN controllers with embedded AI capabilities. There is one central SDN controller (multidomain orchestrator) located at the GS cloud that manages the network slice requests and selects the slice configuration, an SDN controller per configuration (configuration coordinator) that coordinates the SLA decomposition across domains located at BSs, an SDN controller that manages the resources across domains per configuration (cross-domain configuration manager) located at MEC, and local SDN controllers (domain controllers) that manage resources within each domain distributed across computing nodes in each domain. Therefore, the multidomain orchestrator will learn to select the best configuration to serve slice requests for each traffic class based on the feedback from tenants' regarding their satisfaction with the service and feedback from the configuration coordinator with regard to the achieved SLA. Similarly, the configuration coordinator will learn to decompose the SLA based on the feedback on resource availability from each domain controller and feedback from the cross-domain coordinator on the cooperation between domains. Finally, the domain controllers will learn their redundancy policy based on the requested SLA and network condition to mitigate failures.

The feedback cycle allows the system the opportunity to learn the optimum resource allocation and admission policies per traffic class under various failure modes and traffic conditions. Moreover, slices can be managed efficiently through their lifecycle to meet the QoS requirements. ML algorithms can reduce the decision-making time due to their ability to interact and learn from the environment, and their capability to extract useful information from data [169], [170].

G. Cross-Domain Collaboration for Resource Sharing

Following the resilient NS model in Figs. 7 and 9, SPs receive a reward when the allocated slice satisfies the SLA minus the cost of provisioning redundancy. The reward depends on the traffic class and the controllers' redundancy policy which indicates the redundancy that the controller assigns to the slice in that domain to meet the partial SLA. Therefore, there is a tradeoff between reward and redundancy. To reduce redundancy, SPs can collaborate and share their resources among domains. To facilitate the collaboration, we explore the advantages of distributing the communication and computing tasks through STECNs nodes, as shown in Table III.

1) Computing tasks demanding a large number of resources should be decomposed into smaller tasks and distributed to different terminals.

TABLE III
DISTRIBUTED COMPUTING CONFIGURATIONS IN STECNs

Configuration	Terrestrial	Air	Satellite/GS
k=5	EC	EC	EC/EC
k=6	R	EC	EC/EC
k=7	EC	R	EC/EC
k=8	EC	EC	R/EC
k=9	EC	EC	EC
k=10	R	R	EC/EC
k=11	EC	R	R/EC
k=12	R	EC	R/EC
k=13	EC	R	EC
k=14	R	EC	EC

2) Intermediate nodes with computing capabilities that relay data for the user can also perform computing tasks on the way to the destination.
3) Different configurations provide different resource-sharing opportunities.

If a node fails to share the resources or behaves untruthfully, it will be penalized for compromising the service. This penalty must compensate for the actions taken to re-establish the service including the cost of redundancy based on the strategy. If the untruthful behavior persists, the SP may deny access to the infrastructure. Therefore, it is crucial to investigate incentives mechanisms that foster good behavior. Since we have multiple traffic classes and user terminals with heterogeneous capabilities, the valuation of their remaining resources should be considered when designing the incentives.

H. Scalability

The large-scale nature of STECNs with increasing number of IoT devices and services, and the expected increased in UAVs and satellites requires scalable solutions to adapt to the number of devices. The proposed slicing scheme takes into account scalability in the design with respect to the number of domains, configurations, and slice requests. The multidomain orchestrator achieves scalability of configurations thanks to the configuration coordinator. Similarly, the configuration coordinator achieves scalability with respect to the number of domains by relying in the cross-domain configuration manager. Resource scaling can also be performed distributively by the domain controllers.

I. Lessons Learned

1) Multidomain Orchestration and Management: Collaboration among different providers is needed to create slices across administrative domains and with heterogeneous resources. Existing works have overlooked the multidomain aspect of slicing STECNs. Therefore, the optimization of the resource allocation per domain and across domains, and collaboration schemes for traffic offloading between domains deserve further investigation.

2) SLA Decomposition Rule: The network dynamics affect the resource availability and thus, the SLA needs to be decomposed dynamically in different domains as the STECN topology evolves. The decomposition rule must be adaptive to the network condition and available resources. Distributed solutions are required to optimize the slice configuration, routing, and SLA decomposition. Moreover, if outages occur, the decomposition rule should consider redundancy to improve the resiliency in communications and computing.

3) SLA Violation Detection: Performance maintenance throughout the slice life cycle is more demanding in multidomain scenarios and requires further research. In addition, service requirements may be dynamic, and thus, mechanisms to monitor and evaluate the SLA are crucial to identify and prevent future failures.

4) Cost-Efficient Resilient Slicing Solutions: Existing works on resilient mobile networking [15], [16] mainly focus on maintaining the network connected. However, many applications have stringent requirements in terms of throughput, delay, or reliability. Thus, the design of resilient mechanisms should jointly consider other aspects, such as energy efficiency, latency, and communication cost.

VI. OPEN CHALLENGES AND FUTURE RESEARCH

In this section, we summarize open challenges and future research directions toward resilient slicing in STECNs.

A. Satellite Edge Computing

The multitier structure of STECNs brings up new research paradigms in MEC that require fine grain and cross-domain offloading decisions that adapt to the traffic dynamics and network conditions. The latency and computational power tradeoff for EC have been broadly studied in terrestrial networks [8], [171]. In [171] and [172] they focus on optimizing UAV trajectory to reduce the communication latency and energy consumption of IoT devices and show remarkable performance gain compared to fixed servers. However, there has been little work on EC facilitated by STECNs and its tradeoff analysis. In [173], satellites and UAVs are considered for heterogeneous traffic offloading. They present a fixed offloading partitioning to offload ultrareliable low-latency communications traffic to the air and terrestrial domains while eMBB traffic is offloaded to the satellite domain. However, adaptive computing task offloading decisions will be needed as more differentiated services become available.

B. Multidomain Network Slicing

Slicing STECNs requires coordination of resources across different domains to meet requirements for diversified services. Current slicing implementations are mostly centralized, which increases latency and overhead traffic in the network, or focused on a single domain. However, the mapping of requirements to network resources and the optimization of the SLA decomposition throughout the different domains should be distributed. Several works discussed how to decompose the E2E SLA throughout multiple domains in terrestrial networks [100], [101] but in STECNs further research is needed to guarantee the E2E SLA under uncertain resource availability [7].

C. Resilient Networking Protocols

Resilience should be an integral aspect in designing NS schemes for STECNs given their characteristics. The slicing schemes need to adapt to different failures, network conditions, and dynamic SLAs. Existing works focus on a specific type of failure, e.g., server failure [105] or errors in the transmission [107] without an integral resilient design. Therefore, how to schedule and manage resource redundancy automatically to augment the slices and meet the tenants' requirements remains an open problem.

D. Adaptive Network Slicing

The mobility of satellites and UAVs results in dynamic topologies which require adaptive resource allocation and slice scaling. Besides, the increasing number of satellites also brings more computing and processing capabilities and thus, the challenge of how to allocate them to perform computing tasks. The learning algorithms must be distributed and consider onboard energy, storage capacity, and computational power constraints [50]. The design should explore the unique characteristics of each domain and the challenges related to running in multiple heterogeneous networks. Since only a limited amount of informative data is available at each network node (generated by applications, sensing devices, users, etc.) and exchanging raw data between multiple devices is constrained to their battery and availability of network resources, new frameworks are needed to define distributed and cross-domain learning algorithms. Recent works on AI have focused on how to overcome inadequacies and future B5G or 6G systems are projected to include reliable and advanced AI solutions [174], which will rely on efficient decision making, knowledge exploitation, and sophisticated learning.

E. Privacy and Security

Several works address security challenges in satellite–terrestrial networks [175]. Satellites as eavesdroppers are considered in [176] and they analyze the ergodic and outage secrecy capacities. Ahmad et al. [177] discussed security challenges among interdomain and intradomain transmissions. However, there is a lack of research on privacy and security in slicing STECNs. Real-world scenarios must assume a strong adversary model in which a tenant/subscriber can evaluate all observable attributes of a slice and modify inputs to compromise the slicing management solution for their benefit. However, the lack of information sharing (e.g., resource management strategies and traffic patterns) between the SPs and tenants [102] makes the latter unable to detect attacks since they ignore how the SPs allocate the resources and whether the traffic load variations are expected or not. Similarly, the SPs are only aware of the physical infrastructure and ignore the mapping from virtual to physical resources. Sharing this sensitive information will bring significant overhead and make NS even more vulnerable to attacks. To this end, several works [102], [103], [104] propose learning-assisted NS solutions in terrestrial networks and analyze the performance under DoS attack [102], detect anomalies of nodes in a slice [103], and mitigate topology learning attack [104]. Nevertheless, research on how to incorporate security considerations in the SLA in STECNs remains to be studied.

VII. Conclusion

This article provides a comprehensive survey and analysis of NS in STECNs. First, we present the background on the integration architectures for STECNs, and the enabling technologies for slicing STECNs, such as satellite EC, SDN for satellite networks, and NFV. Then, we discuss multidomain NS, schemes for slicing STECNs, and its challenges. Further, we describe solutions to perform resilient networking and slicing and present a design for resilient multidomain NS in STECNs. Potential issues in NS due to network failures are far from well-investigated in both academia and industry. The proliferation of services enabled by the Internet of Remote Things will increase the number and types of slices running in the system and thus resilient design will benefit multiple applications. Future works can apply these concepts to enhance the resiliency in slicing different satellite–terrestrial integration architectures.

References

[1] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Aug. 2019.

[2] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[3] B. Lorenzo, J. Garcia-Rois, X. Li, J. Gonzalez-Castano, and Y. Fang, "A robust dynamic edge network architecture for the Internet of Things," *IEEE Netw.*, vol. 32, no. 1, pp. 8–15, Oct. 2018.

[4] B. Wang, Y. Sun, Z. Sun, L. D. Nguyen, and T. Q. Duong, "UAV-assisted emergency communications in social IoT: A dynamic hypergraph Coloring approach," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7663–7677, Aug. 2020.

[5] A. Kak and I. F. Akyildiz, "Designing large-scale constellations for the Internet of Space Things with CubeSats," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1749–1768, Feb. 2021.

[6] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "Joint HAP access and LEO satellite backhaul in 6G: Matching game-based approaches," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1147–1159, Apr. 2021.

[7] A. Kak and I. F. Akyildiz, "Towards automatic network slicing for the Internet of Space Things," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 1, pp. 392–412, Mar. 2022.

[8] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1171–1181, Dec. 2016.

[9] Y. Wang, J. Yang, X. Guo, and Z. Qu, "A game-theoretic approach to computation offloading in satellite edge computing," *IEEE Access*, vol. 8, pp. 12510–12520, 2020.

[10] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80–87, May 2017.

[11] J. García-Rois, B. Lorenzo, F. J. González-Castaño, F. Gil-Castiñeira, and J. Wu, "Slice allocation and pricing framework for virtualized millimeter wave cellular networks," *IEEE Access*, vol. 7, pp. 86349–86366, 2019.

[12] W. Cho and J. P. Choi, "Cross layer optimization of wireless control links in the software-defined LEO satellite network," *IEEE Access*, vol. 7, pp. 113534–113547, 2019.

[13] K. Lin, D. Wang, L. Hu, M. S. Hossain, and G. Muhammad, "Virtualized QoS-driven spectrum allocation in space-terrestrial integrated networks," *IEEE Netw.*, vol. 33, no. 1, pp. 58–63, Jan./Feb. 2019.

[14] N. Bartolini, S. Ciavarella, T. F. La Porta, and S. Silvestri, "On critical service recovery after massive network failures," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2235–2249, Aug. 2017.

[15] A. Pašić et al., "eFRADIR: An enhanced framework for disaster resilience," *IEEE Access*, vol. 9, pp. 13125–13148, 2021.

[16] M. Matracia, N. Saeed, M. A. Kishk, and M.-S. Alouini, "Post-disaster communications: Enabling technologies, architectures, and open challenges," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1177–1205, 2022.

[17] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, "High-reliability and low-latency wireless communication for Internet of Things: Challenges, fundamentals, and enabling technologies," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7946–7970, Oct. 2019.

[18] L. U. Khan, I. Yaqoob, N. H. Tran, Z. Han, and C. S. Hong, "Network slicing: Recent advances, taxonomy, requirements, and open research challenges," *IEEE Access*, vol. 8, pp. 36009–36028, 2020.

[19] R. A. Addad, T. Taleb, H. Flinck, M. Bagaa, and D. Dutra, "Network slice mobility in next generation mobile systems: Challenges and potential solutions," *IEEE Netw.*, vol. 34, no. 1, pp. 84–93, Jan./Feb. 2020.

[20] T. Taleb, I. Afolabi, K. Samdanis, and F. Z. Yousaf, "On multi-domain network slicing orchestration architecture and federated resource control," *IEEE Netw.*, vol. 33, no. 5, pp. 242–252, Sep./Oct. 2019.

[21] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 3rd Quart., 2018.

[22] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of Things realization in 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 957–994, 2nd Quart., 2021.

[23] Y. Wu, H.-N. Dai, H. Wang, Z. Xiong, and S. Guo, "A survey of intelligent network slicing management for Industrial IoT: Integrated approaches for smart transportation, smart energy, and smart factory," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1175–1211, 2nd Quart., 2022.

[24] F. Debbabi, R. Jmal, L. C. Fourati, and A. Ksentini, "Algorithmics and modeling aspects of network slicing in 5G and beyonds network: Survey," *IEEE Access*, vol. 8, pp. 162748–162762, 2020.

[25] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi, and L. Vangelista, "A survey on technologies, standards and open challenges in satellite IoT," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1693–1720, 3rd Quart., 2021.

[26] M. M. Azari et al., "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2633–2672, 4th Quart., 2022.

[27] O. Kodheli et al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 1st Quart., 2021.

[28] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 1st Quart., 2018.

[29] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 437–461, Jan. 2022.

[30] F. Fourati and M.-S. Alouini, "Artificial intelligence for satellite communication: A review," *Intell. Converg. Netw.*, vol. 2, no. 3, pp. 213–243, 2021.

[31] A. Sattarzadeh et al., "Satellite-based non-terrestrial networks in 5G: Insights and challenges," *IEEE Access*, vol. 10, pp. 11274–11283, 2022.

[32] D. Peng, A. Bandi, Y. Li, S. Chatzinotas, and B. Ottersten, "Hybrid beamforming, user scheduling, and resource allocation for integrated terrestrial-satellite communication," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8868–8882, Sep. 2021.

[33] H. Wu et al., "Impact of hardware impairments on outage performance of hybrid satellite-terrestrial relay systems," *IEEE Access*, vol. 7, pp. 35103–35112, 2019.

[34] S. Sreng, B. Escrig, and M.-L. Boucheret, "Exact symbol error probability of hybrid/integrated satellite-terrestrial cooperative network," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1310–1319, Mar. 2013.

[35] X. Artiga et al., "Shared access satellite-terrestrial reconfigurable backhaul network enabled by smart antennas at mmWave band," *IEEE Netw.*, vol. 32, no. 5, pp. 46–53, Sep./Oct. 2018.

[36] O. Y. Kolawole, S. Vuppala, M. Sellathurai, and T. Ratnarajah, "On the performance of cognitive satellite-terrestrial networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 668–683, Dec. 2017.

[37] Z. Li, F. Xiao, S. Wang, T. Pei, and J. Li, "Achievable rate maximization for cognitive hybrid satellite-terrestrial networks with AF-relays," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 304–313, Feb. 2018.

[38] H. Ahmadinejad and A. Falahati, "Forming a two-tier heterogeneous air-network via combination of high and low altitude platforms," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1989–2001, Feb. 2022.

[39] Y. Li et al., "Data collection maximization in IoT-sensor networks via an energy-constrained UAV," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 159–174, Jan. 2023.

[40] T. Kim, J. Kwak, and J. P. Choi, "Satellite edge computing architecture and network slice scheduling for IoT support," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14938–14951, Jan. 2022.

[41] R. Xie, Q. Tang, Q. Wang, X. Liu, F. R. Yu, and T. Huang, "Satellite-terrestrial integrated edge computing networks: Architecture, challenges, and open issues," *IEEE Netw.*, vol. 34, no. 3, pp. 224–231, May/Jun. 2020.

[42] Y. Zhang, C. Chen, L. Liu, D. Lan, H. Jiang, and S. Wan, "Aerial edge computing on orbit: A task offloading and allocation scheme," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 1, pp. 275–285, Jan./Feb. 2023.

[43] Z. Zhang, W. Zhang, and F.-H. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques," *IEEE Netw.*, vol. 33, no. 1, pp. 70–76, Jan./Feb. 2019.

[44] Z. Song, Y. Hao, Y. Liu, and X. Sun, "Energy-efficient multiaccess edge computing for terrestrial-satellite Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 14202–14218, Sep. 2021.

[45] J. Li, W. Shi, H. Wu, S. Zhang, and X. Shen, "Cost-aware dynamic SFC mapping and scheduling in SDN/NFV-enabled space–air–ground-integrated networks for Internet of Vehicles," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 5824–5838, Apr. 2022.

[46] G. Wang, S. Zhou, S. Zhang, Z. Niu, and X. Shen, "SFC-based service provisioning for reconfigurable space-air-ground integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1478–1489, Jul. 2020.

[47] Y. Cai, Y. Wang, X. Zhong, W. Li, X. Qiu, and S. Guo, "An approach to deploy service function chains in satellite networks," in *Proc. IEEE/IFIP Netw. Oper. Manag. Symp. (NOMS)*, 2018, pp. 1–7.

[48] X. Gao, R. Liu, and A. Kaushik, "Virtual network function placement in satellite edge computing with a potential game approach," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 2, pp. 1243–1259, Jun. 2022.

[49] X. Gao, R. Liu, and A. Kaushik, "Service chaining placement based on satellite mission planning in ground station networks," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 3, pp. 3049–3063, Sep. 2021.

[50] P. Kumar, S. Bhushan, D. Halder, and A. M. Baswade, "FybrrLink: Efficient QoS-aware routing in SDN enabled future satellite networks," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 3, pp. 2107–2118, Sep. 2022.

[51] B. Almadani, A. Beg, and A. Mahmoud, "DSF: A distributed SDN control plane framework for the east/west interface," *IEEE Access*, vol. 9, pp. 26735–26754, 2021.

[52] R. Ferrús et al., "SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges," *Phys. Commun.*, vol. 18, pp. 95–112, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490715000543

[53] C. Chen, Z. Liao, Y. Ju, C. He, K. Yu, and S. Wan, "Hierarchical domain-based multi-controller deployment strategy in SDN-enabled space-air-ground integrated network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 6, pp. 4864–4879, Dec. 2022.

[54] T. Li, H. Zhou, H. Luo, and S. Yu, "SERvICE: A software defined framework for integrated space-terrestrial satellite communication," *IEEE Trans. Mobile Comput.*, vol. 17, no. 3, pp. 703–716, Mar. 2018.

[55] F. Mendoza, R. Ferrús, and O. Sallent, "SDN-based traffic engineering for improved resilience in integrated satellite-terrestrial backhaul networks," in *Proc. 4th Int. Conf. Inf. Commun. Technol. Disaster Manag. (ICT-DM)*, 2017, pp. 1–8.

[56] O. S. Oubbati, M. Atiquzzaman, T. A. Ahanger, and A. Ibrahim, "Softwarization of UAV networks: A survey of applications and future trends," *IEEE Access*, vol. 8, pp. 98073–98125, 2020.

[57] G. Secinti, P. B. Darian, B. Canberk, and K. R. Chowdhury, "SDNs in the sky: Robust end-to-end connectivity for aerial vehicular networks," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 16–21, Jan. 2018.

[58] H. Iqbal, J. Ma, K. Stranc, K. Palmer, and P. Benbenek, "A software-defined networking architecture for aerial network optimization," in *Proc. IEEE NetSoft Conf. Workshops (NetSoft)*, 2016, pp. 151–155.

[59] T. Das, V. Sridharan, and M. Gurusamy, "A survey on controller placement in SDN," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 472–503, 1st Quart., 2020.

[60] A. Papa, T. De Cola, P. Vizarreta, M. He, C. M. Machuca, and W. Kellerer, "Dynamic SDN controller placement in a LEO constellation satellite network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2018, pp. 206–212.

[61] J. Guo, L. Yang, D. Rincón, S. Sallent, Q. Chen, and X. Liu, "Static placement and dynamic assignment of SDN controllers in LEO satellite networks," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 4, pp. 4975–4988, Dec. 2022.

[62] M. Hu, M. Xiao, W. Xu, T. Deng, Y. Dong, and K. Peng, "Traffic engineering for software defined LEO constellations," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 4, pp. 5090–5103, Dec. 2022.

[63] H. Tang, G. Liu, X. Chen, and N. Xiong, "A survey on Steiner tree construction and global routing for VLSI design," *IEEE Access*, vol. 8, pp. 68593–68622, 2020.

[64] C. Chu and Y.-C. Wong, "FLUTE: Fast lookup table based rectilinear Steiner minimal tree algorithm for VLSI design," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 1, pp. 70–83, Jan. 2008.

[65] C.-W. Lin, S.-Y. Chen, C.-F. Li, Y.-W. Chang, and C.-L. Yang, "Obstacle-avoiding rectilinear Steiner tree construction based on spanning graphs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 4, pp. 643–653, Apr. 2008.

[66] B. Yang, Y. Wu, X. Chu, and G. Song, "Seamless handover in software-defined satellite networking," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1768–1771, Sep. 2016.

[67] X. Tao, K. Ota, M. Dong, H. Qi, and K. Li, "Congestion-aware scheduling for software-defined SAG networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 4, pp. 2861–2871, Oct.–Dec. 2021.

[68] L. Yan, X. Ding, and G. Zhang, "Dynamic channel allocation aided random access for SDN-enabled LEO satellite IoT," *J. Commun. Inf. Netw.*, vol. 6, no. 2, pp. 134–141, 2021.

[69] Y. Drif et al., "An extensible network slicing framework for satellite integration into 5G," *Int. J. Satellite Commun. Netw.*, vol. 39, no. 4, pp. 339–357, 2021.

[70] J. Feng, Q. Pei, F. R. Yu, X. Chu, J. Du, and L. Zhu, "Dynamic network slicing and resource allocation in mobile edge computing systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7863–7878, Jul. 2020.

[71] H.-T. Chien, Y.-D. Lin, C.-L. Lai, and C.-T. Wang, "End-to-end slicing with optimized communication and computing resource allocation in multi-tenant 5G systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2079–2091, Feb. 2020.

[72] H. H. Esmat and B. Lorenzo, "Deep reinforcement learning based dynamic edge/fog network slicing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[73] H. H. Esmat and B. Lorenzo, "Self-learning multi-mode slicing mechanism for dynamic network architectures," *IEEE/ACM Trans. Netw.*, submitted for publication.

[74] M. R. Raza, M. Fiorani, A. Rostami, P. Öhlen, L. Wosinska, and P. Monti, "Dynamic slicing approach for multi-tenant 5G transport networks [invited]," *J. Opt. Commun. Netw.*, vol. 10, no. 1, pp. A77–A90, 2018.

[75] S. Kukliński, L. Tomaszewski, R. Kołakowski, and P. Chemouil, "6G-LEGO: A framework for 6G network slices," *J. Commun. Netw.*, vol. 23, no. 6, pp. 442–453, 2021.

[76] C. Suzhi et al., "Space edge cloud enabling network slicing for 5G satellite network," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2019, pp. 787–792.

[77] S. Roy, "System and methods for providing integrated 5G and satellite service in backhaul and edge computing applications," in *Proc. 38th Int. Commun. Satellite Syst. Conf. (ICSSC)*, 2021, pp. 125–132.

[78] I. Bisio, F. Lavagetto, G. Verardo, and T. de Cola, "Network slicing optimization for integrated 5G-satellite networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[79] T. de Cola and I. Bisio, "QoS optimisation of eMBB services in converged 5G-satellite networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12098–12110, Oct. 2020.

[80] T. Ahmed, A. Alleg, R. Ferrus, and R. Riggio, "On-demand network slicing using SDN/NFV-enabled satellite ground segment systems," in *Proc. 4th IEEE Conf. Netw. Softw. Workshops (NetSoft)*, 2018, pp. 242–246.

[81] L. Lei, Y. Yuan, T. X. Vu, S. Chatzinotas, M. Minardi, and J. F. M. Montoya, "Dynamic-adaptive AI solutions for network slicing management in satellite-integrated B5G systems," *IEEE Netw.*, vol. 35, no. 6, pp. 91–97, Nov./Dec. 2021.

[82] R. Xing, X. Ma, A. Zhou, S. Dustdar, and S. Wang, "From earth to space: A first deployment of 5G core network on satellite," *China Commun.*, vol. 20, no. 4, pp. 315–325, 2023.

[83] F. Lyu et al., "Service-oriented dynamic resource slicing and optimization for space–air–ground integrated vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7469–7483, Jul. 2022.

[84] H. Wu, J. Chen, C. Zhou, J. Li, and X. Shen, "Learning-based joint resource slicing and scheduling in space-terrestrial integrated vehicular networks," *J. Commun. Inf. Netw.*, vol. 6, no. 3, pp. 208–223, 2021.

[85] F. Lyu et al., "Characterizing urban vehicle-to-vehicle communications for reliable safety applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2586–2602, Jun. 2020.

[86] J.-H. Lee, J. Park, M. Bennis, and Y.-C. Ko, "Integrating LEO satellite and UAV relaying via reinforcement learning for non-terrestrial networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[87] T. K. Rodrigues and N. Kato, "Network slicing with centralized and distributed reinforcement learning for combined satellite/ground networks in a 6G environment," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 104–110, Feb. 2022.

[88] H. Esmat and B. Lorenzo, "LEONS: Resilient network slicing configuration and orchestration for LEO-terrestrial edge computing networks," in *Proc. IEEE ICC*, 2022, pp. 1–6.

[89] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[90] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2346–2363, Dec. 2019.

[91] H. Sun, W. Pu, X. Fu, T.-H. Chang, and M. Hong, "Learning to continuously optimize wireless resource in a dynamic environment: A bilevel optimization perspective," *IEEE Trans. Signal Process.*, vol. 70, pp. 1900–1917, 2022.

[92] Z. Chen and B. Liu, *Lifelong Machine Learning*. New York, NY, USA: Morgan & Claypool, 2018.

[93] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[94] V. Kothari, E. Liberis, and N. D. Lane, "The final frontier: Deep learning in space," in *Proc. HotMobile*, 2020, pp. 45–49.

[95] C. Han, L. Huo, X. Tong, H. Wang, and X. Liu, "Spatial anti-jamming scheme for Internet of satellites based on the deep reinforcement learning and Stackelberg game," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5331–5342, May 2020.

[96] L. Lei, E. Lagunas, Y. Yuan, M. G. Kibria, S. Chatzinotas, and B. Ottersten, "Beam illumination pattern design in satellite networks: Learning and optimization for efficient beam hopping," *IEEE Access*, vol. 8, pp. 136655–136667, 2020.

[97] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 28–35, Jun. 2018.

[98] S. Xu, X.-W. Wang, and M. Huang, "Software-defined next-generation satellite networks: Architecture, challenges, and solutions," *IEEE Access*, vol. 6, pp. 4027–4041, 2018.

[99] Y. Bi et al., "Software defined space-terrestrial integrated networks: Architecture, challenges, and solutions," *IEEE Netw.*, vol. 33, no. 1, pp. 22–28, Jan./Feb. 2019.

[100] D. De Vleeschauwer, C. Papagianni, and A. Walid, "Decomposing SLAs for network slicing," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 950–954, Mar. 2021.

[101] I. Kovacevic, A. S. Shafigh, S. Glisic, B. Lorenzo, and E. Hossain, "Multi-domain network slicing with latency equalization," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2182–2196, Dec. 2020.

[102] Q. Liu, T. Han, and N. Ansari, "Learning-assisted secure end-to-end network slicing for cyber-physical systems," *IEEE Netw.*, vol. 34, no. 3, pp. 37–43, May/Jun. 2020.

[103] W. Wang, Q. Chen, X. He, and L. Tang, "Cooperative anomaly detection with transfer learning-based hidden Markov model in virtualized network slicing," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1534–1537, Sep. 2019.

[104] V. N. Sathi, M. Srinivasan, P. K. Thiruvasagam, and S. R. Murthy, "Novel protocols to mitigate network slice topology learning attacks and protect privacy of users' service access behavior in softwarized 5G networks," *IEEE Trans. Depend. Secure Comput.*, vol. 18, no. 6, pp. 2888–2906, Nov./Dec. 2021.

[105] P. K. Thiruvasagam, A. Chakraborty, and C. S. R. Murthy, "Resilient and latency-aware orchestration of network slices using multi-connectivity in MEC-enabled 5G networks," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 3, pp. 2502–2514, Sep. 2021.

[106] F. Malandrino, C. F. Chiasserini, and G. Landi, "Service shifting: A paradigm for service resilience in 5G," *IEEE Commun. Mag.*, vol. 57, no. 9, pp. 120–125, Sep. 2019.

[107] B. Katz, S. Greenberg, N. Yarkoni, N. Blaunstien, and R. Giladi, "New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 308–319, Mar. 2007.

[108] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symp. Security Privacy*, 2010, pp. 305–316.

[109] M. Usama, I. Ilahi, J. Qadir, R. N. Mitra, and M. K. Marina, "Examining machine learning for 5G and beyond through an adversarial lens," *IEEE Internet Comput.*, vol. 25, no. 2, pp. 26–34, Mar./Apr. 2021.

[110] "Adaptable optical communications to facilitate future low-earth orbit networks." Accessed: Sep. 13, 2021. [Online]. Available: https://www.darpa.mil/news-events/2021-09-13

[111] "Spectrum and wireless innovation enabled by future technologies (SWIFT)." Accessed: May 11, 2022. [Online]. Available: https://www.nsf.gov/pubs/2022/nsf22571/nsf22571.htm

[112] "NASA." Accessed: Sep. 2021. [Online]. Available: https://www.nasa.gov/directorates/heo/scan/services/networks/near-space-network/about

[113] "INSTRUCT." Accessed: Sep. 2021. [Online]. Available: https://instruct-ipbg.uni.lu/

[114] "Demonstrator for satellite-terrestrial integration in the 5G context." Accessed: Nov. 2018. [Online]. Available: https://satis5.eurescom.eu/

[115] "Moving experimentation facility site (satellite connected vehicle)." Accessed: Jun. 2018. [Online]. Available: https://www.5g-vinni.eu/moving-experimentation-facility-site/

[116] "UNI." Accessed: Jan. 15, 2022. [Online]. Available: https://wwwfr.uni.lu/snt/research/sigcom/projects/aswell-autonomous-network-slicing-for-integrated-satellite-terrestrial-transport-networks

[117] "What is the Concept of 5G! Pagoda?" Accessed: Jul. 1, 2018. [Online]. Available: https://5g-pagoda.aalto.fi/

[118] "SLICENET." Accessed: Aug. 20, 2017. [Online]. Available: https://slicenet.eu/

[119] F. Rinaldi et al., "Non-terrestrial networks in 5G & beyond: A survey," *IEEE Access*, vol. 8, pp. 165178–165200, 2020.

[120] K. Liolis, A. Franchi, and B. Evans, "Editorial for Wiley's IJSCN special issue 'satellite networks integration with 5G,'" *Int. J. Satelliate Commun. Netw.*, vol. 39, no. 4, pp. 319–321, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sat.1413

[121] "Network function virtualisation (NFV); Management and orchestration; report on architecture options to support multiple administrative domains, v3.1.1," ETSI, Sophia Antipolis, France, Rep. GR NFV-IFA 028, 2018.

[122] Z. Mou, Y. Zhang, F. Gao, H. Wang, T. Zhang, and Z. Han, "Deep reinforcement learning based three-dimensional area coverage with UAV swarm," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3160–3176, Oct. 2021.

[123] H. Shakhatreh et al., "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48572–48634, 2019.

[124] N. Van Cuong, Y.-W. P. Hong, and J.-P. Sheu, "UAV trajectory optimization for joint relay communication and image surveillance," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10177–10192, Dec. 2022.

[125] S. R. Chowdhury et al., "Dedicated protection for survivable virtual network embedding," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 4, pp. 913–926, Dec. 2016.

[126] L. Gao and G. N. Rouskas, "Virtual network reconfiguration with load balancing and migration cost considerations," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2018, pp. 2303–2311.

[127] Z. Zhou, T. Lin, and K. Thulasiraman, "Survivable cloud network design against multiple failures through protecting spanning trees," *J. Lightw. Technol.*, vol. 35, no. 2, pp. 288–298, Jan. 15, 2017.

[128] N. Shahriar et al., "Reliable slicing of 5G transport networks with bandwidth squeezing and multi-path provisioning," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 3, pp. 1418–1431, Sep. 2020.

[129] Q. Zhang, O. Ayoub, J. Wu, F. Musumeci, G. Li, and M. Tornatore, "Progressive slice recovery with guaranteed slice connectivity after massive failures," *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, pp. 826–839, Apr. 2022.

[130] P. Sun, Z. Guo, J. Li, Y. Xu, J. Lan, and Y. Hu, "Enabling scalable routing in software-defined networks with deep reinforcement learning on critical nodes," *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, pp. 629–640, Apr. 2022.

[131] Y. Yuan, L. Lei, T. X. Vu, Z. Chang, S. Chatzinotas, and S. Sun, "Adapting to dynamic LEO-B5G systems: Meta-critic learning based efficient resource scheduling," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9582–9595, Nov. 2022.

[132] C.-Q. Dai, J. Luo, S. Fu, J. Wu, and Q. Chen, "Dynamic user association for resilient backhauling in satellite–terrestrial integrated networks," *IEEE Syst. J.*, vol. 14, no. 4, pp. 5025–5036, Dec. 2020.

[133] J. Yang, D. Li, X. Jiang, S. Chen, and L. Hanzo, "Enhancing the resilience of low earth orbit remote sensing satellite networks," *IEEE Netw.*, vol. 34, no. 4, pp. 304–311, Jul./Aug. 2020.

[134] W. Cao, Y. Zou, Z. Yang, and J. Zhu, "Relay selection for improving physical-layer security in hybrid satellite-terrestrial relay networks," *IEEE Access*, vol. 6, pp. 65275–65285, 2018.

[135] J. Mitola and G. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[136] X. Zhang et al., "Outage performance of NOMA-based cognitive hybrid satellite-terrestrial overlay networks by amplify-and-forward protocols," *IEEE Access*, vol. 7, pp. 85372–85381, 2019.

[137] C. Zhang, C. Jiang, J. Jin, S. Wu, L. Kuang, and S. Guo, "Spectrum sensing and recognition in satellite systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2502–2516, Mar. 2019.

[138] G. Araniti, I. Bisio, M. De Sanctis, A. Orsino, and J. Cosmas, "Multimedia content delivery for emerging 5G-satellite networks," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 10–23, Mar. 2016.

[139] C. Niephaus, J. Mödeker, and G. Ghinea, "Toward traffic offload in converged satellite and terrestrial networks," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 340–346, Jun. 2019.

[140] G. Giambene, D. K. Luong, T. de Cola, V. A. Le, and M. Muhammad, "Analysis of a packet-level block coding approach for terrestrial-satellite mobile systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8117–8132, Aug. 2019.

[141] Y. Shi, Y. Cao, J. Liu, and N. Kato, "A cross-domain SDN architecture for multi-layered space-terrestrial integrated networks," *IEEE Netw.*, vol. 33, no. 1, pp. 29–35, Jan./Feb. 2019.

[142] Z. Mou, F. Gao, J. Liu, and Q. Wu, "Resilient UAV swarm communications with graph convolutional neural network," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 393–411, Jan. 2022.

[143] T. Nagatsuma and A. Kasamatsu, "Terahertz communications for space applications," in *Proc. Asia–Pac. Microw. Conf. (APMC)*, 2018, pp. 73–75.

[144] T. S. Rappaport et al., "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[145] S. Nie and I. F. Akyildiz, "Channel modeling and analysis of inter-small-satellite links in Terahertz band space networks," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8585–8599, Dec. 2021.

[146] I. F. Akyildiz, J. M. Jornet, and C. Han, "TeraNets: Ultra-broadband communication networks in the terahertz band," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 130–135, Aug. 2014.

[147] R. Kumar and S. Arnon, "SNR optimization for LEO satellite at sub-THz frequencies," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 4449–4458, Jun. 2022.

[148] Y. Xing and T. S. Rappaport, "Terahertz wireless communications: Co-sharing for terrestrial and satellite systems above 100 GHz," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3156–3160, Oct. 2021.

[149] N. Deng and M. Haenggi, "A fine-grained analysis of millimeter-wave device-to-device networks," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4940–4954, Nov. 2017.

[150] F. Zhao, W. Hao, H. Guo, G. Sun, Y. Wang, and H. Zhang, "Secure energy efficiency for mmWave-NOMA cognitive satellite terrestrial network," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 2226–2229, Jun. 2021.

[151] Q. Zhang, K. An, X. Yan, and T. Liang, "Coexistence and performance limits for the cognitive broadband satellite system and mmWave cellular network," *IEEE Access*, vol. 8, pp. 51905–51917, 2020.

[152] H. Lin, C. Zhang, Y. Huang, R. Zhao, and L. Yang, "Fine-grained analysis on downlink LEO satellite-terrestrial mmWave relay networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1871–1875, Sep. 2021.

[153] Q. Wang, H. Zhang, J.-B. Wang, F. Yang, and G. Y. Li, "Joint beamforming for integrated mmWave satellite-terrestrial self-backhauled networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9103–9117, Sep. 2021.

[154] S. He et al., "A survey of millimeter-wave communication: Physical-layer technology specifications and enabling transmission technologies," *Proc. IEEE*, vol. 109, no. 10, pp. 1666–1705, Oct. 2021.

[155] Z. Xiao et al., "A survey on millimeter-wave beamforming enabled UAV communications and networking," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 557–610, 1st Quart., 2022.

[156] J. P. Sterbenz et al., "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Comput. Netw.*, vol. 54, no. 8, pp. 1245–1265, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128610000824

[157] P. Cholda, J. Tapolcai, T. Cinkler, K. Wajda, and A. Jajszczyk, "Quality of resilience as a network reliability characterization tool," *IEEE Netw.*, vol. 23, no. 2, pp. 11–19, Mar. 2009.

[158] M. Chiesa, A. Kamisiński, J. Rak, G. Rétvári, and S. Schmid, "A survey of fast-recovery mechanisms in packet-switched networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1253–1301, 2nd Quart., 2021.

[159] M. Rasti, S. K. Taskou, H. Tabassum, and E. Hossain, "Evolution toward 6G multi-band wireless networks: A resource management perspective," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 118–125, Aug. 2022.

[160] B. Lorenzo, F. J. Gonzalez-Castano, and Y. Fang, "A novel collaborative cognitive dynamic network architecture," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 74–81, Feb. 2017.

[161] F. A. Awin, Y. M. Alginahi, E. Abdel-Raheem, and K. Tepe, "Technical issues on cognitive radio-based Internet of Things systems: A survey," *IEEE Access*, vol. 7, pp. 97887–97908, 2019.

[162] M. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1917–1937, Jun. 2005.

[163] J. Zhang, E. Modiano, and D. Hay, "Enhancing network robustness via shielding," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2209–2222, Aug. 2017.

[164] J. Zhou, W. Zhao, and S. Chen, "Dynamic network slice scaling assisted by prediction in 5G network," *IEEE Access*, vol. 8, pp. 133700–133712, 2020.

[165] H. Ding, X. Li, Y. Cai, B. Lorenzo, and Y. Fang, "Intelligent data transportation in smart cities: A spectrum-aware approach," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2598–2611, Dec. 2018.

[166] M. A. Khan, "Intelligent environment enabling autonomous driving," *IEEE Access*, vol. 9, pp. 32997–33017, 2021.

[167] S. Yoon, O. Soysal, M. Demirbas, and C. Qiao, "Coordinated locomotion and monitoring using autonomous mobile sensor nodes," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 10, pp. 1742–1756, Oct. 2011.

[168] Y. Xu, G. Yu, X. Wu, Y. Wang, and Y. Ma, "An enhanced Viola-Jones vehicle detection method from unmanned aerial vehicles imagery," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1845–1856, Jul. 2017.

[169] D. S. Hochbaum and P. Baumann, "Sparse computation for large-scale data mining," in *Proc. IEEE Trans. Big Data*, vol. 2, 2016, pp. 354–363.

[170] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 310–323, Jan. 2019.

[171] H. Lee, S. Eom, J. Park, and I. Lee, "UAV-aided secure communications with cooperative jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9385–9392, Oct. 2018.

[172] Z. Zhu, S. Wan, P. Fan, and K. B. Letaief, "Federated multiagent actor–critic learning for age sensitive mobile-edge computing," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1053–1067, Jan. 2022.

[173] L. Zhang, W. Abderrahim, and B. Shihada, "Heterogeneous traffic offloading in space–air–ground integrated networks," *IEEE Access*, vol. 9, pp. 165462–165475, 2021.

[174] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar./Apr. 2021.

[175] B. Jiang, Y. Yan, L. You, J. Wang, W. Wang, and X. Gao, "Robust secure transmission for satellite communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 1598–1612, Apr. 2023.

[176] D.-H. Jung, J.-G. Ryu, and J. Choi, "When satellites work as eavesdroppers," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2784–2799, 2022.

[177] I. Ahmad, J. Suomalainen, P. Porambage, A. Gurtov, J. Huusko, and M. Höyhtyä, "Security of satellite-terrestrial communications: Challenges and potential solutions," *IEEE Access*, vol. 10, pp. 96038–96052, 2022.

Haitham H. Esmat (Member, IEEE) received the B.S. and M.S. degrees from Helwan University, Cairo, Egypt, in 2011 and 2016, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Massachusetts Amherst, Amherst, MA, USA.

His current research interests include B5G and 6G mobile communication technologies, device-to-device communications, Internet of Things, dynamic network slicing, satellite–terrestrial edge computing networks, and AI-based applications in wireless communication systems.

Beatriz Lorenzo (Senior Member, IEEE) received the M.Sc. degree in telecommunication engineering from the University of Vigo, Vigo, Spain, in 2008, and the Ph.D. degree from the University of Oulu, Oulu, Finland, in 2012.

She is currently an Assistant Professor and the Director of the Network Science Lab, Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. She has published more than 50 papers and coauthored two books on advanced wireless networks. The latest book *Artificial Intelligence and Quantum Computing for Advanced Wireless Networks* (Wiley, 2022), covers the enabling technologies for the definition, design, and analysis of incoming 6G/7G systems. Her research interests include AI for wireless networks, B5G and 6G network architectures and protocol design, mobile computing, optimization, and network economics.

Dr. Lorenzo was the recipient of the Fulbright Visiting Scholar Fellowship with the University of Florida from 2016 to 2017. She is also an Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. She was the General Co-Chair for WiMob Conference in 2019 and serves regularly in the TPC of top IEEE/ACM conferences.

Weisong Shi (Fellow, IEEE) received the B.S. degree in computer engineering from Xidian University, Xi'an, China, in 1995, and the Ph.D. degree in computer engineering from Chinese Academy of Sciences, Beijing, China, in 2000.

He is a Professor and the Chair with the Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA, where he directs the Mobile and Internet Systems Laboratory and Connected and Autonomous Driving Laboratory, investigating performance, reliability, power and energy efficiency, and trust and privacy issues of networked computer systems, and applications. He is one of the world leaders in the Edge Computing Research Community and published the first book on Edge computing. His paper titled *Edge Computing: Vision and Challenges* has been cited more than 1700 times.

Dr. Shi served as a Lead Guest Editor for the edge computing special issue on the Prestigious Proceedings of the IEEE in 2019. He is the Founding Steering Committee Chair of the ACM/IEEE Symposium on Edge Computing and the IEEE/ACM Connected Health: Applications, Systems and Engineering. He is an ACM Distinguished Scientist.