Resource Management From Single-domain 5G to End-to-End 6G Network Slicing: A Survey

Ebrahimi, S., Bouali, F. & Haas, O. C. L.

DOI 10.1109/COMST.2024.3390613
ESSN 1553-877X

Publisher: Institute of Electrical and Electronics Engineers

Resource Management From Single-domain 5G to End-to-End 6G Network Slicing: A Survey

Sina Ebrahimi, Faouzi Bouali, Olivier C. L. Haas, *Senior Member, IEEE*

***Abstract*—Network Slicing (NS) is one of the pillars of the fifth/sixth generation (5G/6G) of mobile networks. It provides the means for Mobile Network Operators (MNOs) to leverage physical infrastructure across different technological domains to support different applications. This survey analyzes the progress made on NS resource management across these domains, with a focus on the interdependence between domains and unique issues that arise in cross-domain and End-to-End (E2E) settings. Based on a generic problem formulation, NS resource management functionalities (e.g., resource allocation and orchestration) are examined across domains, revealing their limits when applied separately per domain. The appropriateness of different problem-solving methodologies is critically analyzed, and practical insights are provided, explaining how resource management should be rethought in cross-domain and E2E contexts. Furthermore, the latest advancements are reported through a detailed analysis of the most relevant research projects and experimental testbeds. Finally, the core issues facing NS resource management are dissected, and the most pertinent research directions are identified, providing practical guidelines for new researchers.**

***Index Terms*—Network Slicing, End-to-End (E2E), Resource Management, Technological Domains, Radio Access Networks (RANs), Transport Networks (TNs), Core Networks (CNs), 5G/6G Networks, Orchestration, Resource Allocation (RA).**

I. Introduction

THE FIFTH generation (5G) of mobile networks is transforming connectivity, catalyzing the development of a digitized society. In this context, network Slicing (NS) has been pioneered by Research and Development (R&D) teams in both industry and academia, paving the way towards 5G's digitalization. NS enables 5G to support a wide range of requirements, including enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), and massive Machine-Type Communication (mMTC). NS divides the network into slices, each with unique features tailored to meet the heterogeneous requirements of individual users. This approach contrasts with the *one-size-fits-all* strategy of previous generations (i.e., 2G-4G) of mobile networks, leading to superior adaptability, enhanced performance, and additional opportunities for new service offerings for Mobile Network Operators (MNOs) [1]. In addition to supporting 5G services, NS allows MNOs to launch novel and unprecedented applications [2]. The NS technology is continuously evolving with the development of 5G-Advanced [3] and it is expected to be incorporated into 6G networks due to its flexibility and cost-effectiveness. The NS market is projected to grow by over fifty percent annually from 2023 to 2030 [4], offering a revenue opportunity of approximately $45 billion in 2025 and $200 billion in 2030 [2]. Furthermore, it is anticipated that 30% of potential 5G use cases will require NS as a facilitator [5].

Manuscript received August 4, 2023; revised January 13, 2024; accepted March 19, 2024.

S. Ebrahimi, F. Bouali, and O. Haas are with the Centre for Future Transport and Cities (CFTC), Coventry University, Coventry, CV1 5FB, UK. *(Corresponding author: Faouzi Bouali)* - (e-mail: {ebrahimis, ad6501, csx259}@coventry.ac.uk).

The concept of *slice* in networking was first introduced in the late 1980s [6]. Overlay networks were the initial form of NS, combining diverse resources to create virtual networks [7]. The division of the physical infrastructure into logical networks was first applied to the mobile networks by the Dedicated Core (DECOR) in the 4G standards [8]. This approach allows multiple Core Networks (CNs) to be deployed by MNOs over the same infrastructure while offering limited resource sharing and flexibility to different service consumers. Introduced in 2015 as a critical enabler for 5G [9], NS goes beyond DECOR by providing MNOs with full flexibility to support different applications over the same infrastructure. Since then, research and industry projects have focused on overcoming the obstacles associated with the development and operation of NS, as well as its impact on other communication services.

Recent R&D initiatives outline the significance of End-to-End (E2E) NS frameworks in MNO networks (see Fig. 3) [10]. However, most of the literature concentrates on overcoming the obstacles associated with a particular technological domain (e.g., CN) of telecommunication networks. To fully unlock the potential of NS, all technological domains, including Radio Access Network (RAN), Transport Network (TN), and CN, must be jointly considered within an E2E NS framework. This allows to capture the interdependence between these domains, which can substantially enhance the efficiency of resource management among other crucial functionalities [11], ultimately resulting in better support for end-user requirements. However, achieving consistent, interoperable, and secure cross-domain coordination presents challenges, as different stakeholders may control various technological domains in diverse administrative domains.

A. Motivation and Aim

The current NS solutions mostly focus on the CN segment with some initial applications to other domains (e.g., RAN and TN), while the industry necessitates NS frameworks that encompass all these domains, also known as E2E NS, to better satisfy the end-user requirements. Many MNOs have not yet offered NS-enabled 5G services due to the lack of vendor solutions supporting NS across all technological domains. Therefore, researchers in academia and industry should bridge this gap by enriching state-of-the-art frameworks with the required NS support in each technological domain and managing resources across domains to achieve the highest efficiency.

In this direction, few recent surveys have been devoted to review the progress made on NS resource management functionalities [12]–[15]. However, most of these works have focused on a single technological domain and cannot capture the issues that arise in cross-domain contexts[1]. First, the various technological domains are strongly interdependent, and so are their feasibility regions. For instance, virtualized RAN and CN functions might run on a shared pool of computing resources. This means that two acceptable RAN and CN slicing solutions might not be feasible in an E2E context due to a lack of computing resources. Second, optimizing NS across domains leads to a high number of optimization variables and constraints, leading to a larger exploration space. This means that an acceptable problem-solving methodology for a single domain might become cumbersome in an E2E context, which calls for space reduction strategies (e.g., heuristic algorithms and decomposition methods) to improve scalability. Third, due to the emergence of various technological enablers, the borders between domains have become blurry. For instance, thanks to Multi-access Edge Computing (MEC), it has become possible to relocate some of the CN functions (e.g., User Plane Function (UPF)) to run closer to the user inside the RAN. Also, the virtualization of network functions has made it possible to run latency-relaxed functions as workloads in the cloud.

The above discussion clearly shows that stitching together existing single-domain solutions would not be enough to achieve E2E NS. Resources should rather be managed from an E2E perspective, with the possibility of relocating functions between domains.

In this context, this paper aims at surveying the progress made to achieve efficient NS resource management in cross-domain and E2E contexts. This requires dissecting the core issues facing resource management functionalities in current single-domain, cross-domain, and E2E NS frameworks and identifying the most pertinent research avenues to tackle them.

B. Contributions

To the best of our knowledge, no detailed review of cross-domain and E2E NS resource management frameworks has been conducted. The main contributions of this work are:

- It reports the progress made towards E2E NS, with a focus on the interdependence between technological domains and unique issues that arise in cross-domain and E2E contexts. The limits of single- and cross-domain NS frameworks are identified, and the potential contributions of E2E NS in overcoming these are analyzed.
- It formulates a generic optimization problem that can be instantiated to cover any resource management functionality and/or cross-domain combination and critically analyzes the appropriateness of different problem-solving methodologies in solving instances of the formulated problem in cross-domain and E2E contexts.
- It examines NS resource management functionalities across domains, revealing their limits when applied separately per domain. Practical insights are provided, explaining how these functionalities should be rethought in cross-domain and E2E contexts.
- It considers practical aspects, including stakeholders, methodologies, use cases, and simulation/emulation platforms and captures the latest progress made by the R&D initiatives and testbeds across the world.
- It dissects the core issues facing NS resource management functionalities, and points to the most pertinent research directions to overcome them, providing practical guidelines for new researchers.

[1]The terms *single-domain* and *cross-domain* in this paper refer to the technological domains of the mobile network architecture (see Fig. 3).

C. Comparison with Existing Surveys

Most of the previous surveys can be categorized into four main groups (see Table I).

The first group focused on single-domain frameworks (e.g., CN [12], [13], TN [14], and RAN [15]–[17]), with a brief discussion of cross-domain aspects [18]–[20].

The second group limited their contribution to specific NS functionalities and capabilities. For example, Resource Allocation (RA) and orchestration are reviewed thoroughly in [21]–[23] and [6], [24]–[28], respectively. A recent survey about Virtual Network Function (VNF) placement in NS only examined Deep Reinforcement Learning (DRL) techniques [13]. Other works focused on Admission Control (AC) [29] and security considerations [23], [30]–[34]. However, most of the works mentioned only review single-domain NS solutions.

The third group focused on the problem-solving methodologies of NS frameworks. For instance, optimization problems supporting the RA functionality are surveyed in [21], [22], while associated algorithmic aspects of NS orchestration and RA are reviewed in [35], [36]. A growing number of reviews (e.g., [16], [15], and [37]) focus on Machine Learning (ML) approaches and their DRL subset (e.g., [13], [38]–[40]).

The final group tackled a number of miscellaneous aspects, including enabling technologies (i.e., Software-Defined Networking (SDN)/Network Function Virtualization (NFV) (e.g., [6], [24], and [25]) and MEC (e.g., [41] and [42])), standardization perspective (e.g., [6], [18], [25], and [30]), experimentation and R&D projects (e.g., [20] and [43]–[47]), and use cases (i.e., automotive/transport [18], [48], energy [18], [49], and industry 4.0 [18]).

As opposed to the aforementioned surveys, this work reviews the literature on NS resource management with a focus on domain interdependence in cross-domain and E2E contexts. Table I positions our work compared to the most relevant review papers. Readers interested in domain-specific studies are referred to the papers listed in Table I.

D. Structure

This article surveys the literature on NS resource management applied to wireless networks (in particular, 5G and beyond) from 2015 to 2023. As illustrated in Fig. 1, the paper includes three main sections (i.e., Sec. III-V). The single-domain literature (Sec. III) is examined in terms of missed opportunities compared to cross-domain solutions (Sec. IV). The

TABLE I: Comparison with other related works.

Subject	Categories					Our Approach
Technological Domain	CN [12], [13]	TN [14]	RAN [15]–[17]	Cross-domain analysis [18]–[20]		- Analysis of resource management functionalities across domains (Sec. III) - Covering cross-domain and E2E solutions (Secs. IV, V) - Focus on interdependence between domains (Secs. III, IV, V)
Functionality	RA [21]–[23]	Orchestration [6], [24]–[28]	VNF Placement [13]	Security [23], [26], [30]–[34]	AC [29]	Covering all NS resource management functionalities (Secs. II-F, III, IV, V)
Methodology	Optimization [21], [22], [35], [36]			ML [13], [15], [16], [37]–[40]		Critical analysis of problem-solving methodologies for different instances of the NS problem (Secs. II-G, III, IV, V)
Misc.	Enabling Technologies [6], [24], [25], [41], [42]		Standardization View [6], [19], [25], [30]	Experimentation View and R&D Projects [20], [25], [43]–[47]	Use Cases [18], [48], [49]	Secs. II-D, VI

*Each row is categorized differently, and the columns in each row are unrelated.

analysis of the E2E literature (Sec. V) identifies weaknesses and gaps in cross-domain solutions that E2E frameworks can overcome. In Sec. III and Sec. V, NS frameworks are classified based on supported resource management functionalities. In Sec. IV, works are categorized based on supported technological domains to outline the considered and overlooked resources in cross-domain frameworks. For all sections (i.e., Sec. III-V), the most prominent functionalities are covered in detail in the subsections, while the additional functionalities are briefly listed in the summary tables (e.g., Table III-V).

The remainder of this paper is structured as follows (see Fig. 1). Key definitions are presented in Sec. II. Single-domain, cross-domain, and E2E NS frameworks are examined in Secs. III, IV, and V, respectively. Sec. VI reviews relevant research projects and experimental platforms. Research gaps and future works are presented in Sec. VII. Finally, Sec. VIII concludes this study. The key acronyms are listed in Table II.

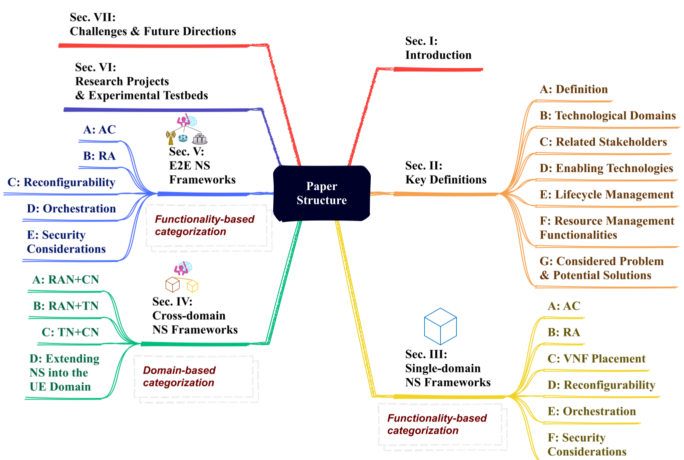

Fig. 1: Paper organization.

II. Key Definitions

This section covers essential definitions based on the 3rd Generation Partnership Project (3GPP) standards, which is the main Standards Development Organization (SDO) in the context of NS [11].

A. Definition

An NS is a logical E2E network that can be dynamically created by a slice producer and delivered to a slice consumer to support a particular service according to a given Service Level Agreement (SLA).

Once the slice is activated within the network infrastructure, it is referred to as the Network Slice Instance (NSI), and each of its logical portions is represented as a Network Slice Subnet Instance (NSSI) [11]. For example, an NSI can consist of RAN, TN, and CN NSSIs.

TABLE II: List of key acronyms.

Acronym	Definition
3GPP	Third Generation Partnership Project
5G/6G	Fifth/Sixth Generation
5G PPP	The 5G Infrastructure Public Private Partnership
5GC/EPC	5G/Evolved Packet Core
A2C	Advantage Actor-Critic
AC	Admission Control
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
API	Application Programming Interface
AR/VR/XR	Augmented/Virtual/Extended Reality
AUSF	Authentication Server Function
BS	Base Station
CN	Core Network
CNF	Cloud-native Network Functions
CSMF	Communication Service Management Function
CU/DU/RU	Centralized/Distributed/Radio Unit
DDPG/TD3	(Twin Delayed) Deep Deterministic Policy Gradient
DES	Double Exponential Smoothing
DL/SL/UL	Deep/Supervised/Unsupervised Learning
DQL/DRL	Deep Q-/Reinforcement Learning
DQN/DDQN	(Double) Deep Q-Network
eMBB	Enhanced Mobile Broadband
ETSI	European Telecommunications Standards Institute
GBR	Guaranteed Bit Rate
GSMA	Global System for Mobile Communications Association
GST	Generic Network Slice Template
GT	Game Theory
IETF	Internet Engineering Task Force
ILP	Integer Linear Programming
InP	Infrastructure Provider
IoT/IoV	Internet of Things/Vehicles
ITU	International Telecommunication Union
KPI	Key Performance Indicators
LCM	Lifecycle Management
LLM	Large Language Model
MAC	Medium Access Control
MANO	Management and Orchestration
ML/FL/RL	Machine/Federated/Reinforcement Learning
mMTC	Massive Machine-Type Communications
MNO	Mobile Network Operator
NEST	Network Slice Type
NF/NFV	Network Function (Virtualization)
NN	Neural Network
NS	Network Slice
NSI/NSSI	NS (Subnet) Instance
NSMF/NSSMF	NS (Subnet) Management Function
OSS	Operations Support System
PoP	Point of Presence
PPO	Proximal Policy Optimization
PRB	Physical Resource Block
QoS/QoE	Quality of Service/Experience
RA	Resource Allocation
RAN/C-RAN	(Centralized) Radio Access Network
RDPG	Recurrent Deterministic Policy Gradient
RIC	Radio Intelligent Controller
SBI	SouthBound Interface
SCA	Side-Channel Attack
SDN/SDR	Software-Defined Network/Radio
SDO	Standards Development Organization
SFC	Service Function Chain
SLA	Service-Level Agreement
SNC	Stochastic Network Calculus
SO/NFVO	Service/NFV Orchestrator
ST	Slice Tenant
TN	Transport Network
TS/TR	Technical Specification/Report
UAV	Unmanned Aerial Vehicles
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra-Reliable Low-Latency Communication
USRP	Universal Service Radio Peripheral
VM	Virtual Machine
VNF/VNE	Virtual Network Function/Embedding
VNFM/VIM	VNF/Virtualized Infrastructure Manager
ZSM	Zero-Touch Service Management

The created NSIs can be managed at three levels [50]. The Communication Service Management Function (CSMF) receives the service-specific requirements from the slice consumer and translates them to network-specific needs. The Network Slice Management Function (NSMF) controls NSIs from a higher perspective, and each Network Slice Subnet Management Function (NSSMF) manages its NSSIs. For example, a CN NSSMF controls the sub-slices (NSSIs) within the CN domain. Fig. 2 shows an illustrative example for these definitions.

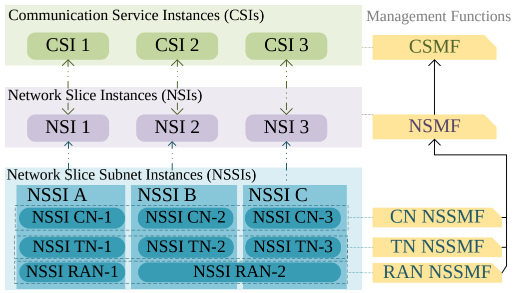

Fig. 2: Mapping NS management functions to NSIs [11], [51].

B. Technological Domains

NS operates in mobile networks, which encompass various technological domains (i.e., CN, TN, RAN, and User Equipment (UE)) (see Fig. 3). Each technological domain performs one or multiple technology-specific tasks (e.g., routing in TN). From this point forward, *domain* refers to one of the technological domains. Administrative domains will be explicitly specified when needed (see Sec. II-C). Regarding NS frameworks, single-domain solutions focus on a single domain, whereas cross-domain frameworks consider multiple domains. E2E NS solutions are a subset of cross-domain frameworks that simultaneously incorporate RAN, TN, and CN domains.

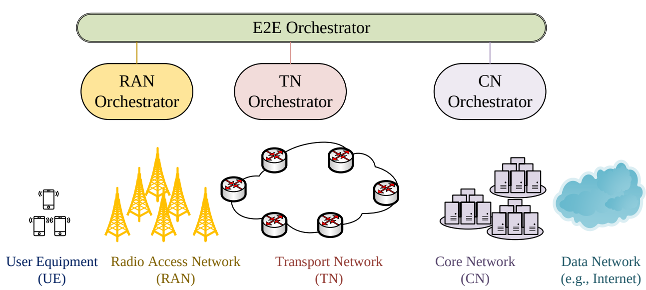

Fig. 3: Simplified mobile network architecture.

To realize slice management capabilities, a hierarchical orchestration architecture is required, where each domain is managed by its own orchestrator, and all of them are overseen by an E2E orchestrator that has an abstracted view of the whole network [20]. In the 3GPP terminology, domain orchestrators can be mapped to NSSMFs, and the E2E orchestrator consists of CSMF and NSMF [11].

1) CN: CNs facilitate serving end-users by providing functionalities such as access and mobility management, session management, and slice selection [52]. Traditionally, MNOs deployed all CN NFs in the most central part of their network. However, with the advances in NFV and MEC technologies, MNOs can deploy delay-sensitive network functions (e.g., UPF) in the edge sites closer to the end-users.

2) TN: TNs consist of the links connecting RAN and CN domains as well as their internal nodes [52]. The key factors when slicing the TN are the link capacity and latency constraints, together with the efficiency of the routing between SDN-enabled switches [19].

3) RAN: Thanks to the emergence of network function disaggregation, a traditional Base Station (BS) can be split into different logical nodes (e.g., Radio Unit (RU), Distributed Unit (DU), and Centralized Unit (CU)) that may be geographically distributed. This functional split can be performed at different layers of the RAN protocol stack depending on the deployment scenario [20]. For instance, the emerging Open RAN (O-RAN) paradigm promotes a low-level split (split 7.2x) between the RU and DU [53], which complements the standard high-level split (split 2) between DU and CU [54]. Furthermore, some of the RAN subcomponents (e.g., CU and DU) can be virtualized and integrated with MEC servers in appropriate sites. This makes them easier to manage, especially for NS.

4) UE: UE is the device providing connectivity to the end-user and serves as a termination point for the RAN, enabling communication with the network [52].

C. Related Stakeholders

Fig. 4 depicts a simple view of NS stakeholders, including the Infrastructure Provider (InP), MNO, Slice Tenants (STs), and end-users [11]. As demonstrated in Fig. 4, each stakeholder holds distinct perspectives regarding the network, potentially conflicting with other stakeholders' goals. A stakeholder can be a slice consumer and/or a slice provider; e.g., the ST can provide a slice to end-users while consuming the slice provided by the MNO. Each stakeholder within an NS framework can be translated into an administrative domain[2]. In practical scenarios, two administrative domains, such as an InP and an MNO, can manage a technological domain like TN. In addition, one stakeholder, such as an MNO, may have control over both RAN and TN resources, while another administrative domain, such as an InP, may be responsible for providing CN resources for several MNOs.

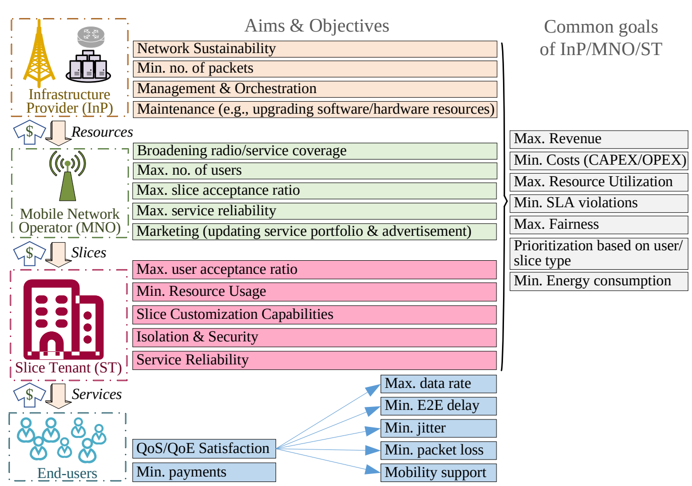

Fig. 4: NS stakeholders and their objectives.

[2]For the sake of simplicity, many NS frameworks operate only within a single administrative domain, e.g., an MNO that serves end-users.

*1) **InP**:* InP is the entity in charge of operating some resources, e.g., switching fabrics, computing servers, and radios. Data center providers and public/private cloud and edge providers are some examples of InPs.

*2) **MNO**:* MNO is the central entity to manage NSs. It may acquire infrastructure from InPs and serve the STs to deliver applications to end-users.

*3) **ST**:* ST consumes the slices provided by the MNO and ensures the end-users are served with good Quality of Service/Experience (QoS/QoE) levels. Examples of STs vary from small, medium, and large enterprises to vertical industries (e.g., automotive and manufacturing) and virtual MNOs.

*4) **End-user**:* The end-user is an entity (e.g., smartphone, Internet of Things (IoT) sensor, and (semi)-automated vehicle) consuming one or more slices provided by an ST/MNO to support one or more applications [11].

Finally, it is worth pointing out that an organization can play multiple roles simultaneously (e.g., an MNO that owns all its infrastructure also plays the role of InP).

D. Enabling Technologies

This section briefly introduces the three essential enabling technologies to realize NS.

*1) **NFV**:* Motivated by the cost-efficiency advantages of cloud computing [55], this technology has been adopted to deploy hardware-based proprietary Network Functions (NFs) as VNFs or Cloud-native Network Functions (CNFs) on Virtual Machines (VMs) or containers, respectively [25]. The deployed VNFs/CNFs can be chained to form a Service Function Chain (SFC) in co-located or distributed cloud environments, offering network or value-added services [6].

*2) **SDN**:* SDN centralizes network control and intelligence to simplify network management and enable programmability. This is achieved by separating the control plane from the data plane in networking equipment such as switches and routers [6]. It allows scalability, flexibility, service-oriented adaption, and robustness, all of which are necessary for an NS framework [56]. For further information about SDN in the context of NS, the reader can refer to [6], [25], [57].

*3) **MEC**:* MEC exploits the flexibility offered by NFV to deploy computing and storage infrastructure in locations closer to the end-users. The local processing reduces the load on the backhaul links [58], shortens the experienced latency [42], and ensures privacy preservation of sensitive data [59].

E. Lifecycle Management

Fig. 5 presents the NS Lifecycle Management (LCM) as per 3GPP [11] (please ignore the scroll signs until the end of Sec. II-F). It involves the following phases:

- *Preparation* occurs before creating an NSI in the network. It comprises NS design, onboarding, and environment setup.
- *Commissioning* creates the NSI and allocates/configures all required resources to meet NS requirements.
- *Operation* includes the activation, supervision, reporting, modification, and deactivation of an NSI.
- *Decommissioning* terminates NSIs and relinquishes their non-shared resources.

F. Resource Management Functionalities

NS resource management frameworks perform a subset of the following functionalities:

*1) **AC**:* This is the mechanism run by the slice producer to decide whether to accept or reject slice requests [60]. It can be an initial part of the RA process to ensure that the allocated resources are available at the NS operation time. AC is always part of the preparation phase in NS LCM (see Fig. 5).

*2) **RA**:* Once AC accepts a slice request, the slice producer allocates the required resources to the slice to be created. From the LCM perspective, RA is part of the preparation and commissioning phase (see Fig. 5).

*3) **VNF placement**:* This is also known as Virtual Network Embedding (VNE) or SFC embedding. It examines the ideal placement of VNFs/CNFs at nodes (servers) and reserves the necessary interconnection capacity across them [61]. It is part of the RA functionality and is performed during the commissioning phase of the NS LCM (see Fig. 5).

*4) **Reconfigurability**:* A slice may require reconfiguration (e.g., scaling of its VNFs[3]) due to changing conditions (e.g., availability of resources or traffic demands) [62]. From the LCM point of view, reconfiguration is performed during the operation phase (see Fig. 5).

*5) **Orchestration**:* Orchestration enables interactions between the management entities in various domains and facilitates the configuration/modification of the components within domains [25], [38]. Advanced orchestration frameworks may feature LCM automation, security and trust mechanisms, and multi-stakeholder interworking procedures. Orchestration is performed in all LCM phases (see Fig. 5).

*6) **Security considerations**:* From a security perspective, the flexibility introduced by NS comes at the cost of a wider attack surface for all stakeholders. Mitigating the introduced threats across all domains is a pending issue in the NS literature. From the LCM viewpoint, security mechanisms can be run in all phases (see Fig. 5). It should be noted that this current review solely addresses security-related issues pertaining to resource management. For a more comprehensive discussion, see [34].

Summary: Fig. 5 maps the aforementioned functionalities to 3GPP's NS LCM phases [11]. Note that orchestration and security span across all phases.

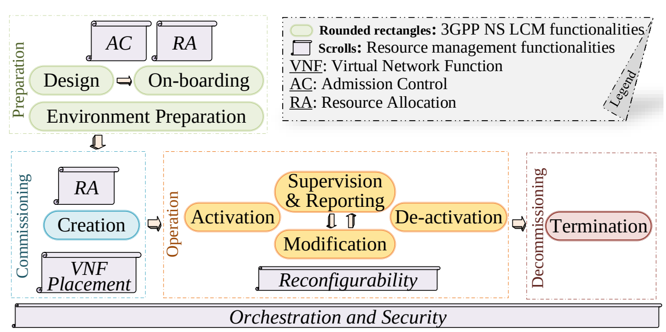

Fig. 5: Mapping of NS resource management functionalities/capabilities to NS LCM.

[3]Horizontal VNF scaling involves changing the number of active VNF instances, while vertical VNF scaling involves adding or removing resources from existing VNFs [13].

G. Considered Problem and Potential Solutions

*1) **Generic problem formulation**:* This section formulates a reference optimization problem that can be mapped to any resource management problem in the NS context. The objective function $f(x_1, \dots, x_V)$ can be designed to capture one or multiple objectives. The optimization direction in (1) can be either min, max, min - max, or max - min, depending on the considered objective function [63].

$$\begin{aligned} &\overbrace{\text{min/ max/ min - max/ max - min}}^{\textit{Optimization direction}} && f(\overbrace{x_1, \dots, x_V}^{V \textit{ Variables}}) \\ &\textit{subject to } \boldsymbol{M} \textit{ constraints:} && \begin{cases} g_1(x_1, \dots, x_V), & \dots, \\ g_M(x_1, \dots, x_V). \end{cases} \end{aligned} \tag{1}$$

Optimization variables can typically be categorized as either continuous or discrete variables [64]. Transmission power and allocated bandwidth are examples of the former, while Physical Resource Block (PRB) allocation and VNF placement variables belong to the latter. Continuous variables can be computationally intensive due to their infinite possible values. On the other hand, discrete variables (e.g., binary and integer) can result in NP-hard problems, signifying that no known algorithm can solve all instances of the problem in polynomial time. In addition, they disrupt the problem's convexity, making locating the global optimum challenging.

Whether considering a single-domain, cross-domain, or E2E scenario, the majority of constraints related to NS problems are usually expressed in the form of inequalities, classified primarily into the following types [65]. Note that the presented mathematical formulations are simplified, and the constraints may include interaction between variables (e.g., cross-product terms). Also, depending on the specific constraint of the particular problem, the mathematical operator may be $\leq$, $=$, or $\geq$.

- *Linear constraints:* These constraints are expressed as linear relationships of one or more variables. They are the most frequently employed model in NS problems and can be expressed mathematically as follows:
$$\textstyle\sum_i x_i \leq X_{\max}, \tag{2}$$
where $X_{\max}$ represents the total capacity or maximum threshold, and x_i is the resource allocated for the i-th user/slice/flow. This category of constraints may consider capacity restrictions (e.g., power and computing resources) or other SLA-related requirements (e.g., maximum tolerable E2E delay).
- *Non-linear constraints:* These can involve functions that are non-linear in variables, such as exponential, quadratic, trigonometric, or logarithmic functions. These constraints can be generally formulated as follows:
$$g(x_1, \dots, x_V) \leq c, \tag{3}$$
where c represents the maximum capacity or threshold, and $g(x_1, \dots, x_V)$ is a non-linear function of one or more variables. One example of these constraints can limit the maximum power used in beamforming scenarios. Solving optimization problems with non-linear constraints can be challenging due to non-convexity, necessitating more sophisticated techniques [64].

The objective function in (1) can be a single function or a combination of multiple functions, each affecting a subset of the defined variables. Increasing the number of objectives in $f(x_1, \dots, x_V)$ raises the problem difficulty. Multiple functions can be combined using weighted sum, weighted product, and max-min approaches [66]. The most common objectives used in NS problems based on the problem's viewpoint(s) are summarized in Fig. 4[4]. For instance, we might consider an E2E NS RA problem with two objectives: maximizing revenue (from the MNO perspective) and minimizing E2E delay (from the end-user standpoint). By employing the weighted sum approach, the two objectives can be normalized and multiplied by controllable weights to tune their relative importance.

*2) **Problem analysis**:* An E2E NS problem inherently involves optimizing multiple (i.e., continuous and discrete) variables across various technological domains. The set of feasible points (e.g., admitted users, provisioned slices, selected VNFs) in these problems is typically discrete or reducible to a discrete set. Consequently, an E2E NS problem can often be identified as a combinatorial optimization problem [67]. As shown in Fig. 6, each technological domain has a feasible region for a given variable (e.g., the binary PRB variable[5]) in an E2E NS problem. The feasible region for such a problem is the intersection of the feasible sets of all technological domains. In addition, E2E constraints (e.g., latency of slices) might further limit the feasible region of E2E NS frameworks. For instance, there is no point in assigning lots of TN resources to a given user if there are not sufficient PRB resources in the RAN. This demonstrates that the constraints and variables of each technological domain can affect the others, creating a complex, interconnected system. Therefore, simply stitching together NS solutions, each focused on a single domain such as RAN, TN, or CN, is insufficient to meet the E2E NS requirements. A holistic approach that jointly considers all domains is required to accurately reflect the interdependencies and achieve optimal results. This joint coordination among technological domains is crucial for effectively addressing E2E NS problems. However, solving such problems can be arduous due to their complexity and the large number of variables involved. Therefore, it is often more practical to aim for a near-optimal answer within a reasonable computational complexity rather than striving for complete optimality [68].

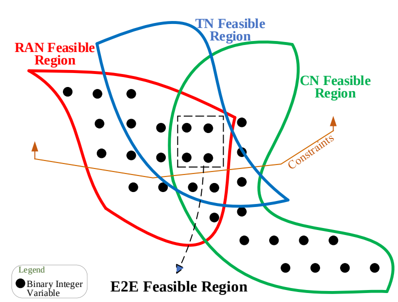

Fig. 6: An illustration of the need to ensure feasibility across all NS technological domains and satisfy E2E constraints (e.g., meeting latency requirements for different slice types).

[4]It should be noted that the list of objectives presented is not exhaustive and is provided only as an illustrative guide.

[5]Denotes the allocation of PRBs in each BS to the users.

*3) **Problem-solving methodologies:*** The following introduces a categorization of approaches for solving resource management problems in NS.

- **Optimization-based techniques** provide approximate or exact answers to optimization problems. These approaches can be classified into three categories:
  - *Closed-form approaches* provide an exact solution to the problem without the need for iterative techniques. They can be applied to problems with relatively small dimensions. Queueing Theory (QT) and Stochastic Network Calculus (SNC) are examples of this category used to model networks. For instance, E2E delay is modeled using QT in [69]. However, it is unlikely that these solutions would be sufficient to solve an NS problem since many assumptions and constraints must be taken into account.
  - *Relaxation-based approaches*, such as successive convex approximation and semi-definite relaxation, can achieve near-optimal answers for computationally intractable problems [35]. Eq. (1) can be partitioned into sub-problems and solved using relaxation techniques to achieve acceptable computational complexity (as in [70]). The quality of the approximation depends on the relaxation technique and degree of relaxation employed.
  - *(Meta)heuristic approaches* can speed up the search for answers to large-scale problems with complicated constraints and objective functions, including discrete and continuous variables [36]. While researchers develop problem-specific heuristics (e.g., [71]), metaheuristics (such as evolutionary algorithms) offer a comprehensive framework to map the problem details into their settings (e.g., [72]). These approaches are often applied to tackle large-scale combinatorial problems (e.g., E2E NS RA). For example, [72] applies a genetic algorithm to improve resource utilization and minimize fronthaul latency in RAN slicing. However, the mentioned approaches may not guarantee finding the globally optimal solution and can be sensitive to parameter settings (e.g., population size and crossover in genetic algorithms).
- **ML** is a subset of Artificial Intelligence (AI) that develops algorithms that can learn and make predictions or decisions based on data [16]. ML approaches are commonly employed in NS frameworks for network performance prediction, resource optimization, and automated decision-making [73]. While specific approaches may vary (or overlap), common approaches include:
  - *Supervised Learning (SL)* enables accurate prediction of network performance metrics based on labeled data, helping resource management decisions in NS [74]. For example, a regression tree algorithm can be used to predict PRB utilization in the RAN domain based on the temporal traffic profile of UEs [75]. Nevertheless, these techniques heavily rely on the availability of high-quality training data that adequately represent the features of NS.
  - *Unsupervised Learning (UL)* techniques can identify patterns and structures within NS data, guiding decision-making processes by revealing hidden relationships. For instance, in [76], normalized spectral clustering is utilized for slice AC to analyze similarities between incoming slice requests and currently active slices. This helps to identify the most appropriate NSIs that can fulfill requests with efficient resource usage. Nonetheless, interpreting algorithm results may require domain expertise and meticulous analysis.
  - *Reinforcement Learning (RL)* enables automated decision-making in dynamic environments. Modeling the NS environment using RL definitions (e.g., Markov Decision Process (MDP)) is straightforward and does not require pre-trained data [73]. One example is to use a Q-learning algorithm to learn nearly optimal RA policies, which can either approve or discard STs' requests [77]. However, conventional RL techniques (without using Neural Networks (NNs)) often struggle to converge in bigger environments, such as E2E NS problems.
  - *Federated Learning (FL)* allows collaborative training of shared ML models while preserving data privacy by not sharing raw data [78]. It facilitates decentralized optimization and decision-making among various technological/administrative domains in NS. For example, in [79], each BS acts as an agent where the models are trained, and a central orchestrator decides to assign PRBs and associate BSs according to these models. However, these approaches strongly depend on the reliability of the communication links between the central point and distributed FL agents.
  - *Deep Learning (DL)* techniques leverage multi-layered NNs to learn intricate data patterns and relationships. These networks can be integrated with other ML paradigms to leverage their strengths and address high-dimensional and complex NS problems [73]. Due to using NNs, DRL-based approaches (e.g., Deep Deterministic Policy Gradient (DDPG)) have better scalability and faster decision-making than traditional RL techniques, making them suitable for online NS functionalities (e.g., reconfigurability) after convergence. For example, DDPG is used in [80] to determine the continuous actions (i.e., allocation of radio and computing resources) aiming to maximize the utility of STs while ensuring end-user QoS. However, DL techniques require expertise in hyperparameter tuning and suffer from a lack of explainability.
- **Game Theory (GT)** studies the strategic interactions of rational agents using mathematical models [81]. These agents (e.g., technological/administrative domain orchestrators) aim to optimize their own utility to reach equilibrium where all parties are satisfied. Hence, GT models offer a compelling option for solving problems with multiple stakeholders in the context of NS. These tech-

niques may include cooperative and/or non-cooperative agents. For instance, in [82], virtual MNOs establish prices for RAN resources, and end-users bid for them within a non-cooperative Stackelberg game. Although GT-based approaches typically involve optimization, the distinguishing factor is the strategic interaction among multiple agents. Nevertheless, striking a balance between complexity and the number of agents involved is crucial.

It is worth point out that the above categories of problem-solving methodologies are not mutually exclusive. Each of the discussed categories aims to highlight the primary methodology employed, but different approaches may be combined to solve complex E2E NS problems. For example, heuristically-assisted DRL techniques (e.g., [83]) might combine elements of heuristics (optimization-based), RL (ML-based), and potentially GT.

III. Single-domain NS Frameworks

This category of NS frameworks exclusively covers one domain. While some papers may claim to have introduced a form of E2E NS, we classify them as single-domain solutions if their problem definition is based on a single domain. This section evaluates single-domain resource management functionalities based on their problem-solving approaches.

A. AC

NS AC analyzes abstract requirements from slice requests, maps them to necessary resources, and evaluates feasibility regarding the available resources (e.g., computing in CN, PRBs in RAN) [19]. AC frameworks (e.g., [70]) often ignore request interpretation and presume fixed slice types for STs to choose from. Evaluating slice requests against available resources is referred to as feasibility checking and can be an initial part of NS RA problems. Single-domain NS AC solutions lack coordination with other domains, resulting in a lack of information regarding other domain resources. The current literature can be categorized into two primary groups: optimization- and DRL-based approaches. In the following, we examine the advantages and limitations of these methodologies while referencing some representative works from CN-only and RAN-only NS AC frameworks.

Optimization-based approaches are commonly used to evaluate the RA problem feasibility in terms of providing resources for slice requests. Such techniques usually map the AC problem to a multiple knapsack problem, aiming to maximize the slice acceptance ratio [84]. These techniques often overlook the historical data of slice request arrivals, a crucial factor for making well-informed decisions about RA, rendering them sub-optimal. For example, [70] shows that their joint Integer Linear Programming (ILP) formulation for MNO cost minimization outperforms a disjoint formulation of CN nodes and links. AC can also be formulated as stochastic queueing models of slice requests to provide statistical analysis [85]. Fairness among operational NSIs is often overlooked in RAN AC frameworks, as in the preemption-based prioritization AC model proposed in [86].

DRL-based techniques can better adapt to changing conditions (e.g., drastic changes in request arrival time or demand), and learn from historical information about slice requests. However, they require extensive training data and lack interpretability. However, excelling in prediction accuracy necessitates integrating information on resource availability from other technological domains. Techniques such as Long Short-Term Memory (LSTM) NNs within Deep Q-Learning (DQL)-based algorithms have achieved a balance between averting Service Level Agreement (SLA) violations and resource overbooking, as shown in the CN-only AC framework proposed in [87]. DQL-based schemes can only account for discrete action spaces, but some actor-critic-based algorithms (e.g., Proximal Policy Optimization (PPO)) can also handle continuous action spaces. For instance, a joint RAN slicing framework uses two DRL agents (for AC and RA) trained with the PPO algorithm to maximize MNO revenue [88]. The proposed framework outperforms single-agent RA schemes (e.g., [89]) in terms of revenue and convergence due to considering resource feasibility in the AC agent. It allocates computing resources to VNFs in a Centralized RAN (C-RAN) environment but overlooks PRB and power allocation.

Despite the advancements in both approaches, resource verification across technological domains remains a challenge in single-domain AC, emphasizing the need for more comprehensive solutions. As shown in Fig. 6, the feasible set for CN AC (e.g., [70], [84], [87]) might be reduced when other technological domains are considered. For example, accepting a slice request without ensuring the availability of sufficient TN and RAN resources can lead to diminished QoS for end-users during the operational stage of slices (see Fig. 5). This issue is partly mitigated by cross-domain solutions that jointly consider the relevant constraints across multiple technological domains (see Sec. IV-A1). The emerging Open RAN architecture provides an opportunity to leverage user Key Performance Indicators (KPIs) to further improve the effectiveness of RAN slicing AC (e.g., [85], [86], [88]).

B. RA

RA in single-domain frameworks assigns resources to serve users with diverse requirements. The primary challenge is to optimize resource utilization while meeting QoS requirements. Although each domain evaluates different resources (e.g., computing in CN, path selection and link bandwidth allocation in TN, and PRB and transmission power in RAN), single-domain RA frameworks may cause bottlenecks due to a lack of coordination (refer to Fig. 6). Existing works can be classified into four groups: optimization-based, model-based DRL-based, model-free DRL-based, and GT-based approaches. This section assesses these groups using some representative frameworks.

Optimization-based techniques are popular for optimizing resources as they can often handle both continuous and discrete variables. Relaxation and heuristic procedures reduce complexity but may hurt optimality and increase computational time. They excel at mathematically formulating multiple objectives and constraints. For instance, the RAN slicing technique in [72] utilizes genetic algorithms to optimize C-RAN design by minimizing fronthaul delay and maximizing capacity utilization. In an Open RAN-compliant framework, [90] models PRB, power, and computing resources and uses relaxation- and greedy-based methods to solve partitioned sub-problems. Moreover, the linearized Alternating Direction Method of Multipliers (ADMM) is used to address non-convexity in a joint TN bandwidth allocation and path selection problem in [91].

Model-based DRL can be effective for single-domain RA using constrained MDPs and Deep NNs (DNNs). While this approach can improve accuracy in familiar settings, limited agent exploration may hinder its effectiveness in dynamic environments due to anticipated policies. Examples of model-based RAN slicing frameworks that optimize continuous variables (e.g., spectrum) include [92] and [93]. The latter reference extends DQL to continuous action spaces with discrete normalized advantage functions, resulting in expedited convergence. However, these frameworks only model a single BS, rendering them unsuitable for mobility-aware scenarios.

Model-free DRL approaches for single-domain RA are flexible and easy to deploy. They learn from dynamic environmental interactions, not from models. However, hyperparameter adjustments might be problematic, affecting solution performance and convergence. In addition, the curse of dimensionality may arise due to the significant number of state-action pairs involved. For instance, a Deep Q-Network (DQN) system in a fog environment with edge and core tiers has scalability challenges for diverse use cases [94]. An Advantage Actor-Critic (A2C)-based approach addresses this issue by delivering eMBB and URLLC flows through local controllers for CN nodes and links [95]. To accelerate convergence and scalability, these controllers transfer their models to other controllers via transfer learning, which requires more training episodes. LSTM is used in [96] to predict Channel State Information (CSI) for users in each RAN slice, and PRB resources are optimized using a double and dueling DQN technique to reduce action space size. In [97], a RAN RA system combines ground BSs, Unmanned Aerial Vehicles (UAVs), and Low Earth Orbit (LEO) satellites to maximize throughput, Signal-to-Interference-plus-Noise Ratio (SINR), and latency across three slice types. A centralized supervisory module and distributed modules at each layer are used in the proposed method. The DDPG algorithm in each module can handle both continuous and discrete actions. It has a critic DNN that estimates the value function of the current state-action pair and an actor DNN that maps states to actions. The authors of [98] propose a two-timescale online methodology that prioritizes users based on weight for maximum resource utilization. The problem is divided into power allocation and user admission sub-problems. First, Lyapunov optimization solves the long-term power constraint sub-problem. Second, an offline LSTM that forecasts user request arrivals assists the Trust Region Policy Optimization (TRPO) algorithm in maximizing short-term user satisfaction for the admission sub-problem. A joint network selection and routing scheme between TN nodes is modeled in [99] utilizing an online MDP and a gradient temporal-based actor-critic scheme. Overall, model-free algorithms are more flexible than model-based ones, with the latter being more accurate in familiar conditions.

GT-based approaches can model stakeholder interactions in single-domain RA frameworks to find near-optimal strategies for each actor, especially in resource pricing problems. Nonetheless, the complexity of the problem must be weighed against the number of players (e.g., BSs), as a delicate balance must be maintained. For example, the InP resources are allocated to virtual MNOs through a two-tier scheme that utilizes a federated DDPG-based algorithm, according to [82]. Interactions between the stakeholders are modeled as a Stackelberg game, where MNOs set RAN resource prices and end-users control demand. In [100], a proxy-Lagrangian two-player game is modeled to optimize energy efficiency in C-RAN using a statistical FL technique. Each pair of <CU, NSI> represents a local agent creating its own model. After that, a central orchestrator aggregates local models to allocate PRB and computing resources. A complementary study by the authors of the latter work demonstrates that channel quality is the most crucial factor in RAN slicing evaluation using Explainable AI (XAI) techniques [101].

The works presented overlook other domains. Since unexpected failures in other domains may disrupt the service, these frameworks cannot fully meet user needs. For instance, E2E delay and data rate may be affected by TN congestion or VNF overload in CN when using RAN slicing RA schemes (e.g., [72], [90], [92], [93], [96]–[98]). CN RA frameworks (e.g., [94], [95]) allocate computing and networking resources without considering virtualized RAN and TN resources. This can affect the reliability of MEC applications and hinder support for end-user mobility. TN RA solutions (e.g., [91], [99], [102]) may experience unexpected peak-hour congestion due to RAN domain demand. To overcome these limits, cross-domain frameworks orchestrate resources across several domains, making them more accurate and dependable than single-domain RA solutions (refer to Sec. IV-A1). These frameworks need to support virtualized RAN functions and incorporate the computing resources within the RAN, which allows to support future-proof Open RAN-compliant systems.

C. VNF Placement

VNF placement problems fall within the realm of NS RA problems and thus require balancing end-user/ST satisfaction with resource utilization (or cost reduction). The most sophisticated research suggests numerous Points of Presence (PoPs) to coordinate VNF placement on the premises. This strategy improves VNF placement efficiency and availability but adds communication overhead. NFV and MEC architectures allow CN and RAN VNFs to be deployed at central and edge/fog sites (DU/CU), respectively. Optimization- and DRL-based techniques are used in single-domain NS VNF placement research. The pros and cons of each technique are described below.

Optimization-based techniques effectively address node and link capacity, latency, and data rate limits via simple modeling. However, the NP-hard nature of these problems causes scalability limits [103]. Adapting to dynamic network environments is also a challenge. A relaxation-based technique in [104] seeks to reduce energy consumption and slice delay but may face challenges in dynamic network conditions. This issue, along with scalability, persists in the multi-PoP model presented in [105], where the authors provide a heuristic to minimize the number of VNFs in various slices.

The most advanced VNF placement frameworks use DRL-based approaches for dynamic decision-making. These techniques often employ a variety of algorithms to solve multi-objective problems, improving online performance. However, rapid network topology changes can impact their performance. For instance, [106] proposes a federated DQN technique in Open RAN architecture that can migrate user-related RAN VNF instances between CU-UP and DU nodes in response to traffic variations. Decisions are made by the local agents at DU nodes. A near-Real-Time (Near-RT) Radio Intelligent Controller (RIC) aggregates the local models to create a global model. This federated technique outperforms local DRL but may lower QoS owing to communication overhead. DQN-based algorithms have continuous action space issues, unlike PPO, DDPG, and Asynchronous A2C (A3C) algorithms. A heuristically-assisted A3C algorithm is utilized in [83] to optimize bandwidth usage, slice acceptance ratio, and node load. Another study examines the capability of the PPO algorithm to perform online scaling of UPF CNFs in Kubernetes, explicitly focusing on online horizontal scaling [107]. A Support Vector Machine (SVM) classifier is employed to maintain learning process stability in the face of PPO's stochastic nature, but it comes at the expense of losing dynamicity. While these works addressed the InP's perspective, multi-agent DRL systems can be used to envision multiple MNOs and STs (see Fig. 4). This approach promotes competition among MNOs but poses challenges regarding VNF placement due to possible resource conflicts. The authors in [108] use a multi-agent dueling DQL mechanism, where MNOs assign ST requests to InP resources and adjust rewards based on resource conflicts. Their methodology incorporates neuron hotplugging to reduce model retraining in dynamic situations. However, multi-agent mechanisms must balance communication overhead and resource efficiency.

Prior research has not adequately addressed the need for hybrid placement and orchestration of VNFs and CNFs in slice-enabled environments [61]. RAN (e.g., [106]) and CN (e.g., [83], [104], [105], [107], [108]) VNFs need to be considered together for a more realistic NS architecture. The CN and RAN NSSMFs must collaborate with the NSMF (i.e., part of E2E orchestrator) to follow its policies. RAN VNFs can change packet size or increase user delay, making their inclusion vital for precise placement decisions.

D. *Reconfigurability*

Reconfigurations, such as hardware modifications, VNF scaling, and path re-routing, is crucial for resource optimization and QoS provisioning in NS. Reconfiguring both continuous (e.g., power) and discrete (e.g., VNF assignment) variables in a dynamic environment makes these problems more complex than RA problems. This section evaluates the optimization- and DRL-based techniques used to tackle these challenges while highlighting notable efforts.

Optimization-based approaches are commonly employed for single-domain reconfigurability schemes in NS. Relaxations and heuristics reduce problem complexity but might harm optimality. For example, [62] adjusts CN resources periodically and reconfigures flows for slices at a smaller timescale by employing approximation techniques. By finding a sequence of feasible reconfigurations, the resource reservation heuristic, proposed by [109], can reduce unnecessary reconfigurations, which results in improved runtime efficiency compared to [62]. Nevertheless, these reactive approaches may over-provision resources because they lack a prediction mechanism. A proposed TN reconfiguration technique in [110] uses a Double Exponential Smoothing (DES) algorithm to forecast incoming traffic. Based on predictions, a heuristic reallocates optical and wireless TN resources to maximize resource utilization. Furthermore, in [111], a RAN reconfiguration technique uses NNs to predict behavior when KPIs are not satisfied for a slice. It then heuristically reconfigures the AC and packet scheduler until all slices meet the SLA. However, disregarding other domains leads to sub-optimal solutions in all surveyed works.

DRL-based approaches offer greater resilience to dynamic reconfigurations. Nonetheless, traditional DQL-based systems adapt slower to environmental changes than newer techniques (e.g., DDPG). According to [112], using the dueling DQL algorithm can speed up convergence and make the DQL algorithm more stable. The Branching Dueling Q-network (BDQ) scheme uses action branching to improve DQL scalability and convergence. Using BDQ with a partial exploration policy, [113] reduces SFC reconfiguration costs in CN. Heuristics can also alleviate DQN scalability issues, as shown in the periodic TN slice embedding reconfiguration approach in [114]. The Twin Delayed DDPG (TD3) technique used in [115] enhances CU reconfiguration by adjusting RAN VNF computing capacity depending on traffic variations. TD3 resolves some of DQN's issues by reducing Q-value overestimation and handling continuous action spaces.

None of the evaluated studies take cross-domain orchestration into account, which can impact reconfiguration performance. For example, overlooking TN and RAN resources in CN reconfiguration schemes (e.g., [62], [109], [113]) may result in over- and/or under-utilization of these domains while reconfiguring the CN resources (refer to Sec. IV-A3). Furthermore, while these frameworks can automate tasks such as scaling of RAN VNFs and power adjustment, their level of automation and reconfiguration remains a design choice. The challenge lies in monitoring potential detrimental effects on active slices and ensuring performance improvement. The Open RAN architecture enables automated RAN functions, and improves KPIs (e.g., latency and availability) [53], which can be leveraged to improve RAN reconfiguration.

E. *Orchestration*

Balancing communication overhead and resource efficiency is the paramount challenge in this functionality. Furthermore, coordinating different units inside a single domain (e.g., RU, DU, and CU in RAN slicing) might face scalability issues. Some of the aforementioned single-domain CN frameworks, including [94], [95], [105], involve orchestration. The majority of the existing works in this space can be categorized into TN and RAN orchestration designs. The following discusses these two classes before highlighting open issues.

Managing scalability is easier for MNOs/InPs with effective TN orchestration. A hierarchical SDN-enabled architecture lets local controllers send abstracted topologies to a higher-level orchestrator (TN NSSMF) to enable on-demand node adjustments. However, an important issue is how to cover different TN technologies in these frameworks. In [116], an abstraction model for passive optical networks is proposed, ignoring equipment from other technologies. Segmenting the TN domain into technological (i.e., microwave, optical, and IP [117]) or topological (i.e., fronthaul, midhaul, backhaul, metro, and core [57]) sub-domains may enhance scalability. These sub-domains can improve the TN NSSMF's vision, but signaling overhead should be reduced.

Hierarchical RAN orchestration can enable Open RAN-compliant frameworks. Communication overhead and efficiency must be balanced in these schemes. For example, a federated architecture is proposed in [79] to minimize SLA violations. Each BS allocates PRB resources using multi-agent double DQN. A federated DRL orchestrator organizes these agents to enforce local model-based long-term rules.

SDN- and NFV-based technologies are not supported by all equipment in TN and RAN, making real-world implementations difficult. Achieving full programmability in the TN and RAN domains may require years due to their large scale [118]. This may limit the ability of these domain orchestrators to fully expose slice-related information to higher-level orchestrators (e.g., NSMF). If a well-correlated subset of agents cooperates, FL can enhance agent efficiency. The accuracy achieved via federated agent collaboration should outweigh the communication overhead. Finally, these mechanisms are limited to a single domain, and they must collaborate with NSMF for cross-domain orchestration (see Sec. IV-A4).

F. Security Considerations

Security mechanisms for single-domain NS resource management are examined here. Some of the most important concerns include isolating slices to reduce the impact of Distributed Denial-of-Service (DDoS) attacks, securing the Southbound Interface (SBI) in TN, and preventing jamming and Side-Channel Attacks (SCAs) in RAN. Slice isolation, a pillar in NS [6], ensures that the performance of one active slice does not affect the performance of others. In general, a balance should be maintained between isolating slices (i.e., securing them against attacks) and increasing resource usage efficiency. In the following, the most representative works in this area are grouped according to their operational domain before highlighting open issues.

In the CN domain, isolation-aware optimization techniques have been devised to prevent DDoS attacks. However, their inflexibility to network changes is a big drawback. A Mixed ILP (MILP) problem is presented in [119] to isolate inter- and intra-slice CN resources while focusing on reducing congestion and finding efficient paths. NN-based classifiers, such as the LSTM-based system in [120], can enhance resource efficiency and detection accuracy by assigning benevolent flows to the appropriate NSIs.

In the TN domain, SDN controllers and infrastructure nodes can be attacked via the SBI. Securing it may affect the performance of related RA/reconfigurability schemes. [121] proposes a quantum key distribution for SBI security, allowing secret key extraction from pre-shared keys for TN device authentication and data transmission in NS-enabled settings.

In the RAN domain, nodes might be subject to various attacks, including SCA and jamming [34], [122]. In an SCA, if two slices use separate VNF instances on the same server, a malicious VNF could extract sensitive information from the other slice [123]. To mitigate this vulnerability, [124] presented an SCA-aware VNF placement heuristic to maximize the number of slices in RAN. In [125], an adversarial Q-learning algorithm is used to jam PRBs, maximizing failed slices in a PRB allocation framework. However, the latter work only shows the attack scenario without security measures.

Securing slices in a single domain is insufficient due to the presence of attack vectors in other domains. For instance, attacking TN nodes in a framework only guarded for CN (e.g., [120]) might cause catastrophic failures, disrupting the entire service. Decentralizing single-domain orchestrators improves NS safety and resilience. For instance, Blockchain-enabled SDN architectures (e.g., [126]), still in their infancy, can enhance TN device security but increase communication overheads. Furthermore, third-party Open RAN xApps can access different KPIs, raising additional security risks in the RAN domain that need further investigation.

G. Summary

Resource management functionalities in single-domain NS solutions and their inherent limitations were discussed in this section. The 3GPP has yet to standardize RAN slicing. Therefore, the prevalence of single-domain RAN slicing frameworks will persist. In Table III, resource management functionalities are compared for single-domain NS solutions. Cross-domain NS frameworks can overcome the majority of the identified limitations of single-domain NS solutions, as discussed later in Sec. IV.

IV. CROSS-DOMAIN NS FRAMEWORKS

This section examines the cross-domain solutions depending on which technological domains they cover, i.e., RAN+CN, RAN+TN, and TN+CN. The UE domain is not part of the examined combinations, as it has not been explicitly covered by the existing cross-domain solutions.

A. RAN+CN

This section discusses the NS frameworks covering the RAN and CN domains based on their resource management functionality.

TABLE III: Summary of resource management functionalities in single-domain NS frameworks.

Functionality	Expected behavior	Methodology	Advantage	Disadvantage	Main issue	Example ref.	Methodology sub-class	Algorithm(s)	Objective(s)
AC	- Feasibility check before RA - History & behavior of slice requests can be considered	Opt.*	Reducing the complexity of the original problem	Overlooking historical data	Inaccurate decisions about slice requests due to incomplete info.	CN [84]	Heuristic	Partially adaptive greedy heuristic	Max. MNO revenue & Min. SLA violations
		ML	Learning from historical info	Lack of explainability and strenuous hyperparameter tuning		RAN [88]	DRL	Multi-agent PPO	Max. long-term MNO revenue
RA	- Assigning predetermined resources to serve end-users/STs - Balancing resource utilization with QoS satisfaction	Opt.*	Ability to handle discrete & continuous variables	Sub-optimality may occur due to relaxations and heuristics	Lack of cross-domain coordination may result in unexpected bottlenecks	RAN [90]	Relaxation & Heuristic	Iterative & greedy algorithms	Max. sum rate of all UEs
		ML	Can be adaptable to the environment changes	Hyperparameter tuning is challenging		TN [99]	DRL	Multi-task actor-critic	Max. resource utilization & QoS satisfaction
		GT	Fairer RA due to considering various stakeholders	Difficult modeling		RAN [82]	Federated DRL	Federated DDPG	Max. resource utilization & social welfare
VNF Placement	- Achieving a trade-off between minimizing SLA violations & cost reduction - Multi-PoP and MEC-aware scenarios greatly improve QoS satisfaction	Opt.*	Simple modeling	Sub-optimality may occur due to relaxations and heuristics	SLA violation may occur due to high demand from other domains	CN [105]	Heuristic	Greedy-based heuristic	Min. no. of VNFs in different slices
		ML	Topology updates can be learned using model-free DRL	Communication overhead in multi-agent DRL		RAN [106]	Federated DRL	Federated DQN	Max. resource utilization & Min. VNF reconfiguration overhead
Reconfigurability	- e.g., VNF scaling, path reconfiguration, and hardware adjustments - Achieving full automation is challenging	Opt.*	Simple modeling	Less adaptability to rapid environmental changes	Maintaining user QoS is challenging as changes in other domains are not coordinated	CN [109]	Heuristic	Divide & conquer-based	Min. SLA violations
		ML	More resilience to dynamic circumstances	Lack of explainability and strenuous hyperparameter tuning		TN [114]	DRL & Heuristic	Soft Q-Network	Min. spectrum fragmentation across optical nodes
Orchestration	- Coordinating the intra-domain resources via hierarchical abstraction	ML	Topology updates can be learned using model-free DRL	Communication overhead in multi-agent mechanisms	Sub-optimality due to not coordinating with other domains' orchestrators	RAN [79]	Federated DRL	Multi-agent double DQN	Min. SLA violations
Security	- Isolating inter-slice resources - Reducing the attack surface using levels of slice isolation	Opt.*	Simple modeling	Sub-optimality may occur due to relaxations and heuristics	Other domains are still vulnerable to attacks	RAN [124]	Heuristic	Heuristic	Max. no of RAN slices
		ML	Detecting anomalies	Data collection may be challenging		CN [120]	UL & DL	LSTM	DDoS attack detection

*__Opt.__: Optimization-based. The last three right columns refer to the cited framework in the '*Example ref.*' column.

*1) **AC**:* Performing AC across both RAN and CN necessitates the joint consideration of PRBs in the RAN alongside computing resources in CN nodes. This introduces a level of complexity that is not present in single-domain solutions. Optimization- and DRL-based approaches can be used to evaluate the feasibility of RAN+CN RA problems, as discussed in Sec. III-A. Our survey has revealed just one study centered on optimization. Therefore, this section will concentrate on the discussion of this methodology.

By splitting the problem into two domain-specific sub-problems, two-step optimization can solve AC in the RAN and CN. This approach decreases complexity but may cause sub-optimality. The authors in [127] introduce a two-step heuristic that maximizes MNO revenue while honoring slice SLAs. Joint modeling of the RAN and CN yields superior performance compared to single-domain AC solutions (e.g., [84]). However, the model also relies on many oversimplifying assumptions. For example, it assumes that all CN-related tasks for one slice are processed in one node without considering VNF placement.

Future research should integrate the main RAN and CN resources into a unified AC framework for enhanced accuracy.

*2) **RA**:* Joint RA across RAN and CN resources can be modeled as a combinatorial optimization problem (as discussed in Sec. II-G). The major challenge in RAN+CN RA problems is managing increased complexity compared to single-domain solutions by balancing efficiency and optimality. The existing literature can be classified into three groups: optimization-, GT-, and DRL-based approaches. This section discusses the advantages and limits of each of these classes before highlighting the open challenges in this space.

Combinatorial problems in RAN and CN can be addressed through optimization-based approaches. To achieve this, heuristics or relaxation techniques may be utilized, and large problems can be divided into sub-problems to reduce complexity. However, this approach may lead to sub-optimality. For instance, a combinatorial model based on Latin squares is proposed in [128] to emulate the NSMF functionality. The authors employ a heuristic to maximize resource usage and reduce wait times for slice requests, but they fail to consider the cost imposed on the MNO because they do not properly model the particular RAN (e.g., PRB) and CN (e.g., computing) resources. In [129], coverage and traffic constraints in RAN and SFC embedding in CN are modeled, where two schemes (i.e., joint and sequential) are developed. In the joint scheme, which provides better cost efficiency, each sub-problem is solved for all slices together. In the sequential method, which is simpler and faster, sub-problems are solved for slices serially. Similarly, the authors in [130] propose a joint energy-efficient scheme that outperforms their disjoint formulation by 34%. In [131], the RAN power allocation sub-problem is solved using the Lagrangian gradient method. Nevertheless, their VNF placement sub-problem, simplified into a queue-based system, only calculates incoming traffic service rates using SNC.

Unlike optimization-based techniques, GT-based approaches focus more on the interactions between NS stakeholders. However, apart from their high computational complexity, they often rely on assuming that all players are rational. For instance, RA and orchestration are optimized in [132] to jointly maximize the social welfare among STs and minimize Operational Expenditure (OPEX) for InPs. Each ST bids for different resources to maximize a local payoff function, and InPs decide on the resource prices. Nevertheless, this approach neglects TN-related resources, leading to coordination issues among InPs and STs during the auction game.

DRL-based approaches can effectively handle RAN+CN complexities. However, they encounter trade-offs between efficiency and optimality, and some techniques may struggle with uncertainties or continuous action spaces. For example, [133] presents a mobility-aware DQN scheme maximizing user access rates, where two RAN and CN sub-problems are solved using a dynamic knapsack algorithm and a link mapping heuristic, respectively. The authors demonstrate that this approach outperforms RAN-only baselines. However, DQN-based techniques lose crucial information due to discretization and do not address uncertainties such as channel conditions and user demands. In [80], a twin-actor DDPG algorithm, capable of handling high-dimensional and continuous action space, is used to solve a two-level RA problem in a MEC-enabled NS framework, enhancing performance compared to a CN-only actor-critic scheme (i.e., [134]) by generating slice-level actions before user-level actions. Additionally, in [67], the authors propose a dueling DQL scheme for combinatorial optimization of radio, computing, and storage resources under uncertain tenant demands. Still, it oversimplifies radio resources and neglects power capacity. Lastly, [135] considers both user demand and CSI uncertainties, proposing a Recurrent Deterministic Policy Gradient (RDPG) algorithm that uses a history of uncertain information in an LSTM NN to handle continuous action spaces under partial observations.

The main challenges across RAN+CN works are the complexity of the algorithms and/or the large problem size. It is also essential to "slice" the TN in these frameworks to make them more practical in real-world scenarios [136]. Finally, finding the best trade-off between resource utilization and timescales for both RAN (e.g., allocation of PRBs) and CN (e.g., VNF placement) still requires further research.

*3) **Reconfigurability**:* Reconfiguring the resources in RAN and CN requires orchestration across these domains. This is the main difference between cross- and single-domain reconfiguration schemes reviewed in Sec. III. The main challenge is balancing the coordination overhead with the efficiency of reconfigurations. This section first reviews the progress made using mostly ML- and optimization-based approaches. Then, the open challenges are identified and discussed.

ML-based reconfiguration schemes possess the key ability to operate under unknown dynamic conditions. While traditional DQN-based approaches may struggle to adapt to dynamically changing environments, newer techniques (e.g., A2C, DDPG, and LSTM) are better equipped to handle uncertain circumstances. For example, an A2C scheme in [137] optimally triggers VNF migration in response to user mobility but faces scalability issues that can be mitigated by partial exploration policy [113] or DDPG [80]. An LSTM NN in [138] collects testbed data to enhance resource scalability and assurance in NS LCM automation using Intent-based Networking (IBN) [139]. This framework outperforms average throughput accuracy compared to a RAN reconfiguration scheme (i.e., [115]) due to considering the RAN and CN domains together.

Optimization-based techniques, while useful for known network changes, require a complete network model, which is often impractical. For instance, a heuristic scheme is proposed by [140], where slices are reconfigured online across RAN and CN domains using a vector graph. The Dijkstra-based algorithm, used for resource remapping, reduces reconfiguration delay by approximately 50% compared to a CN reconfiguration scheme [109].

Proper integration of the RAN and CN resources is paramount, as a reconfiguration decision in the RAN may require another reconfiguration in the CN and vice versa. For most of the cited works (e.g., [137] and [140]), the most critical RAN resources (i.e., PRBs) are not considered. Moreover, a reconfiguration in the TN domain can impact the accuracy of decisions made in these two domains, urgently calling for E2E NS reconfiguration systems.

*4) **Orchestration**:* Most RAN+CN orchestration progress has exploited DRL to utilize resources more efficiently in the long term. The performance of these DRL frameworks hinges on the choice of states, actions, reward functions, and hyperparameters. A few representative works are first analyzed before identifying the open challenges.

Many DRL schemes in this context use the DQL algorithm, which is simple to implement but faces difficulty adapting swiftly to topology changes because it relies on a target network—a fixed replica of the estimated value function used for stabilization. This is evident in [141], which innovatively customizes the combinatorial resources across RAN and CN but struggles with large state-action pairs. Additionally, [142] splits the SLA-guaranteed latency into RAN and CN components, and uses a double DQN scheme and a pointer network for SFC mapping to minimize each of these components, respectively. [143] applies a DDPG-based algorithm for an Internet of Vehicles (IoV) scenario, outperforming an optimization-based baseline in the long run.

Albeit these works achieve good performance, they do not consider the TN domain and the characteristics of forwarding nodes (e.g., routing and bandwidth), which can cause unprecedented events, such as congestion in TN nodes. Furthermore, capturing the different timescales associated with each domain can make cross-domain orchestration of resources more efficient. Efficiently accounting for all related resources across domains would empower the orchestration modules to make more accurate decisions regarding resource (re)allocation for NSs, as will be discussed later in Sec. V-D.

*5) **Security considerations**:* This section reviews security mitigation schemes associated with RAN+CN NS resource management. The references identified use optimization-based techniques. Therefore, we evaluate the optimization-based approaches before discussing open challenges, although other methodologies may also be useful.

Optimization-based techniques can model DDoS-aware resource management restrictions. However, a trade-off between DDoS detection accuracy and resource efficiency should be found. For instance, [144] first identifies DDoS attacks using ML-assisted proximal gradient estimation, and then breaks down utility maximization into node-specific sub-problems using ADMM.

Although NS still faces challenges in considering DDoS attacks, it must also address other security issues associated with RAN and CN domains (e.g., VNF/CNF manipulation attacks, privacy concerns of STs, and slice configuration attacks [32]). Furthermore, neglecting the TN domain exposes the created slices to potential attacks on TN nodes.

B. RAN+TN

Most existing RAN+TN solutions focus on the RA and reconfiguration functionalities.

*1) **RA**:* Coordination among the NSSMFs of RAN and TN is necessary for RA. The main issue in these problems is balancing coordination overhead with efficiency. The works in this area can be classified into combined (i.e., AC+RA) and standalone (i.e., RA) approaches. Each class is evaluated before discussing the associated open challenges.

The AC module helps the RA strategy optimize resource usage and QoS. However, it requires complex heuristics and iterative processes. For example, [71] proposes a two-tier architecture, where RAN/MEC servers and central office/TN nodes are located in the lower and upper tiers, respectively, and uses a heuristic to minimize resource over-provisioning subject to user latency constraints. It first checks resource availability, then iteratively adjusts node and link capacity and traffic allocation until user latency requirements are met. It outperforms TN-only proposals (e.g., [99]) in terms of resource utilization. Similarly, [145] synchronizes resource provisioning in RAN and backhaul to maximize QoS satisfaction, outperforming a RAN-only scheme presented in [146]. However, their DQN algorithm is bound by discrete actions and suffers from the curse of dimensionality.

The standalone approach does not perform feasibility check before allocating resources to reduce the ST/end-user waiting time before slice provisioning. For instance, a heuristic proposed in [147] maps RAN slices to a given isolation level for different RAN functional splits and TN constraints while minimizing the number of TN wavelength channels and active CU/DUs. It is shown that higher isolation levels result in lower resource utilization and higher costs. Furthermore, control and traffic tasks are handled in two BS tiers in [148] to maximize the long-term utility of end-users in the presence of backhaul constraints. This is achieved by first activating certain traffic BSs to heuristically assign resources, and then minimizing the loss function in the DDPG algorithm. Although DDPG is shown to be more complex than DQN, it outperforms DQN in terms of RAN coverage and resource utilization.

The aforementioned works have achieved good performance under a variety of conditions but face some limits. First, they do not integrate the VNF placement in the CN domain, which can ensure more reliable QoS guarantees to end-users. Second, they do not support the latest TN features (e.g., segment routing), which can help prioritize and isolate incoming traffic from different slices [136]. Finally, they do not comply with the emerging Open RAN standards [149], which could significantly simplify the management of fronthaul and midhaul in complex NS scenarios.

*2) **Reconfigurability**:* The RAN+TN joint reconfiguration is still in its infancy. It requires orchestration between the two domains, in which its communication overhead must be balanced with the efficiency of reconfigurations. The only reference found in this category uses optimization-based techniques, although alternative approaches may also be useful.

Optimization-based relaxation techniques can be used to enhance the efficiency of reconfiguration schemes in RAN and TN. However, it is important to strike a balance between efficiency and optimality. For example, [150] proposes a two-timescale framework, where RA is performed on the long timescale to maximize the expected utility of STs, while activated slices are reconfigured on a shorter timescale in response to dynamic user traffic and channel state changes. Regularization is used to resolve the problem's non-convexity using a reservation-based heuristic. This approach outperforms a similar benchmark operating only in the CN (i.e., [62]) in terms of revenue and backhaul bandwidth. However, this approach does not consider the history of slice requests and therefore cannot cope with sudden traffic fluctuations.

Future work in this area should focus on the pending issues, such as introducing slice priorities and ensuring fairness across slices. In addition, considering more RAN-specific resources (e.g., power and PRB), as well as modeling the VNF scaling in the CN domain, can significantly enhance the accuracy and performance of these frameworks; see Sec. V-C.

C. TN+CN

The main areas of interest in this area include VNF placement across physical servers (CN), efficient routing (TN), and an efficient coordination mechanism to allocate resources across these domains. Existing works can be classified into SNC-, optimization-, and DRL-based approaches. The following discusses the advantages and limits of each class before outlining the open challenges.

SNC frameworks, a subset of probabilistic optimization-based approaches, are advantageous for TN+CN slicing because they can analyze the interplay between resources, delay, and traffic demand. They often ignore RAN resources, leading to inaccurate E2E latency estimates. [151] employs SNC to reduce packet processing and transmission per VNF, outperforming TN slicing (i.e., [152]) in supported slices per wavelength. According to [153], an SNC-based latency estimation heuristic can be used to adjust slice resources under dynamic but known traffic conditions.

Optimization-based heuristics can solve the NS design problem, which is performed during the preparation phase of the slice LCM (see Fig. 5). These techniques should hit a balance between optimality and efficiency. Disregarding RAN resources makes the framework impractical, as competition for these resources can greatly affect slice performance in real-world scenarios. For instance, a multi-objective Particle Swarm Optimization (PSO)-based heuristic is proposed in [154] to accommodate heterogeneous 5G services (i.e., eMBB, mMTC, and URLLC) while maximizing resource utilization across TN and CN nodes regardless of RAN resources.

Some DRL algorithms (e.g., PPO) are capable of handling continuous and discrete action spaces to maximize resource efficiency. However, hyperparameter tuning in these schemes is challenging. For instance, [155] presents a three-tier MEC system that utilizes two (i.e., independent and joint) PPO-based schemes to handle TN constraints. The joint scheme uses centralized critics and global information, while the independent scheme updates each slice individually, making it faster but less accurate. The joint method outperforms the CN-only RA (i.e., [156]) in terms of resource utilization.

Given the wide range of constraints in the TN and CN domains, action/state spaces in these frameworks can be very large and intractable. Some works (e.g., [154] and [155]) have developed optimization- and DRL-based schemes to reduce the state space and the associated runtime. However, they have emphasized the MNO-ST relationship and thus cannot directly capture the end-user requirements. This makes it difficult to extend these solutions to E2E NS frameworks. Moreover, other critical functionalities (e.g., reconfiguration) have not yet been fully addressed in the existing TN+CN NS frameworks.

D. Extending NS into the UE Domain

Extending the NS architecture to the UE domain can foster improved decision-making in the NS LCM but also poses additional MNO security risks and UE privacy concerns. Architectural extension of the UE into the CN is discussed in [157] and [158]. Some RAN slicing frameworks (e.g., [159]–[162]) also incorporate ML-based approaches to improve RA decisions by considering UE feedback (e.g., Channel Quality Indicators (CQIs)). While 5G networks can handle up to 8 Single-Network Slice Selection Assistance Information (S-NSSAIs) per UE [52], further study is needed on privacy and user-level isolation among various S-NSSAIs.

E. Summary

Table IV summarizes the cross-domain frameworks discussed in this section. The table is not exhaustive, as our goal is to identify the shortcomings of the most mature works. Most of the identified limits of cross-domain NS solutions can be overcome by E2E NS frameworks; see Sec. V. Most reviewed papers cover the RAN and CN domains. This is because these works have striven to be compliant with the main SDO in NS (i.e., 3GPP [11]), which delegates the standardization of TN slicing to another SDO (i.e., Internet Engineering Task Force (IETF) [136]). Future cross-domain solutions may use hierarchical multi-agent DRL techniques (agents per NSSMF, node, or slice) to address the pending issues (e.g., orchestration).

V. E2E NS Frameworks

Recall that E2E NS frameworks are a subset of cross-domain solutions that comprehensively cover RAN, TN, and CN domains [11]. In most of the proposed frameworks, each domain has its own controller/orchestrator that manages and controls the infrastructure within its territory using SDN and NFV technologies. In E2E NS, these orchestrators are controlled by an upper-layer orchestrator that can deploy an E2E slice and orchestrate heterogeneous resources from various domains (e.g., see Fig. 3). The contributions made towards E2E NS can be classified into:

- **Generic (i.e., inter-domain) frameworks,** which only account for high-level coordination among multiple domains but neglect the specific attributes of each domain.
- **E2E solutions,** which incorporate domain-specific resources and cross-domain orchestration, allowing for more accurate network modeling. Both Inter- and intra-domain slicing are performed in this category.

This section discusses the progress made by these contributions for each of the relevant NS functionalities. The methodologies used for each functionality are evaluated, and the references are thoroughly analyzed and compared to their counterparts in cross- and single-domain solutions.

A. AC

Compared to single- and cross-domain AC, E2E AC is more complex because it considers the feasibility of RAN, TN, and CN resources as a whole (refer to Sec. II-G). This section analyzes the progress made on generic and E2E AC schemes, followed by a discussion of the open issues.

1) ***Generic frameworks****:* Cross-domain orchestration, based on abstract domain models, can lead to inaccurate slice request approvals in AC problems if the underlying resources are disregarded.

Optimization-based techniques, while efficient, may cause sub-optimality. For example, the SLA decomposition problem in [165] minimizes slice rejection risk using domain-specific risk models. An E2E orchestrator partitions the E2E slice SLA into portions for each domain controller, which provides a parameter-free risk model of its resources. This abstraction improves AC decision-making compared to a RAN+CN cross-domain benchmark (i.e., [127]). The online AC framework in [60] focuses on slice prioritization and fairness using a multi-queue heuristic similar to a RAN-only benchmark (i.e., [85]), but it is fairer due to jointly considering resources across domains.

2) ***E2E solutions****:* E2E AC mechanisms should cope with the high dimensionality of the problem due to the consideration of heterogeneous resource types. The literature in this area can be divided into RL- and optimization-based frameworks.

RL-based approaches can account for the history of slice arrivals. For instance, the State-Action-Reward-State-Action (SARSA) algorithm is suitable for online interactions with the environment but only handles discrete action spaces. The framework proposed in [76] uses a SARSA scheme for online cross-slice AC and congestion control to maximize the MNO's revenue. It strikes a balance between maximizing resource utilization efficiency and minimizing the likelihood of dropping high-priority slices by reducing the resources of elastic slices to admit new inelastic ones. This yields better outcomes than single-domain solutions (e.g., RAN-only AC [111]) due to its E2E visibility. However, the efficiency and convergence of SARSA can be improved using NNs. Furthermore, power allocation in RAN is not considered since SARSA cannot work in continuous action spaces.

Optimization techniques can be used to increase the efficiency of solving joint AC+RA problems. However, oversimplifying assumptions or relaxations can lead to sub-optimality. For instance, the authors in [166] formulate a multi-MNO max-min optimization problem to maximize accepted slices and minimize MNO costs jointly. After linearizing the formulated problem, it is solved using heuristic algorithms (i.e., joint and sequential), both of which outperform a similar cross-domain solution (i.e., RAN+CN RA in [129]) in terms of acceptance ratio due to the combined consideration of resources and constraints across all domains. Nevertheless, the RAN model is oversimplified because power resources are not considered, which can affect the energy efficiency and accuracy of the RAN model.

3) ***Open challenges****:* The lack of available information on the AC module to predict future failures can lead to the performance degradation of operational slices. By collecting and processing more detailed information from all domains, online AC schemes can enhance end-user QoS satisfaction. This is partly done by [76], but can be further improved by considering more resources.

TABLE IV: A summary of the literature on cross-domain NS solutions.

Category	Reference	Functionality: AC	Functionality: RA	Functionality: VNF P.*	Functionality: Reconf.*	Functionality: Orch.*	Functionality: Security	Stak.*: MNO-ST	Stak.*: MNO-User	Stak.*: InP-ST	Methodology: Opt.*: Closed	Methodology: Opt.*: Relaxation	Methodology: Opt.*: Heuristic	Methodology: ML: RL	Methodology: ML: DL	Methodology: ML: SL	Methodology: ML: UL	Methodology: ML: FL	Methodology: GT-based	Algorithm(s)	Distinguishing feature(s)	Objective(s)	Var.*: Continuous	Var.*: Discrete	Con.*: Linear	Con.*: Non-linear	Online	Simulation Tools	Improvement against single-domain solutions	Shortcomings
RAN+CN	[127]	★★		★					✓			✓	✓							ILP relaxation, heuristics	Two-step AC	Max. MNO revenue		✓	✓		✗	Matlab, Mosek, CVX	- More reliable AC [84]	- Scalability - Oversimplifying CN assumptions
RAN+CN	[129]		★★	★★★		★★		✓	✓			✓								Column generation	Mobility- & coverage-aware	Min. MNO costs		✓	✓		✗	Matlab, CPLEX	- Mobility support [95] - More reliable resource provisioning [98]	- Possible congestion due to ignored TN resources
RAN+CN	[67]	★★	★★			★		✓						✓	✓					Dueling DQL	Demand uncertainty-aware	Max. MNO utilization & slice satisfaction		✓	✓		✗	TensorFlow	- Better slice satisfaction ratio [70]	- Oversimplifying RAN assumptions (e.g., ignored power)
RAN+CN	[132]		★★★			★★				✓		✓							✓	Iterative auction game	Auction among STs	Max. social welfare among STs & Min. OPEX for InPs	✓		✓		✓	Matlab	- Better resource utilization [72] - Better InP-ST coordination [82]	- Possible miscoordination among stakeholders - Potential bottleneck in TN
RAN+CN	[135]		★★★						✓					✓	✓					RDPG	Demand & CSI uncertainty-aware	Max. MNO utility	✓	✓	✓	✓	✗	PyTorch, TensorFlow	- More accurate RAN behavior [105]	- Scalability - Potential bottleneck in TN
RAN+CN	[137]		★★		★★			✓						✓	✓					DQN, A2C	Mobility-aware	Min. latency & reconfiguration overheads	✓	✓	✓		✓	Python, PyTorch	- Considering mobility in RAN [109]	- Scalability - Unnecessary reconfigurations due to disregarding TN
RAN+CN	[142]		★	★★		★★★		✓					✓	✓	✓					Double DQN	SLA decomposition among CN & RAN	Max. no. of users & Min. hop count of SFCs		✓	✓		✓	Python, PyTorch	- More accurate delay model [93]	- No URLLC support - No multi-cell support
RAN+CN	[144]		★★			★	★★		✓			✓				✓				Proximal gradient, ADMM	DDoS mitigation	Max. NS utility		✓	✓		✓	Testbed based on OAI	- Better resource utilization [119]	- Security vulnerabilities in TN
RAN+TN	[71]		★★			★		✓					✓							Heuristic	MEC-based two-tier architecture	Min. resource over-provisioning		✓	✓		✗	Python, Testbed	- No congestion in TN [99]	- Scalability - CN resources disregarded
RAN+TN	[145]		★★★			★			✓				✓	✓	✓					Heuristic-assisted DQN	Fairness among slices	Min. resource utilization & Max. QoS satisfaction		✓	✓		✓	Python, Tensor Flow	- No unexpected congestion [114]	- Neglected computing resources in RAN - Possible QoS degradation because of overlooked CN
RAN+TN	[147]		★★			★	★★		✓				✓							Heuristic	MEC & WDM	Min. no. of established wavelength channels & no. of active CU/DUs	✓	✓	✓		✗	sklearn, Keras	More reliable isolation of slices [82]	- Oversimplified TN design - Less accurate delay due to disregarding CN
RAN+TN	[148]		★★★		★	★★			✓				✓	✓	✓					DDPG, heuristic	Two-tier RAN design	Max. long-term utility for various slice types	✓		✓	✓	✓	Python, TensorFlow	- No potential bottleneck in TN [90]	- Less accurate delay due to disregarding CN
RAN+TN	[150]		★★		★★★	★★		✓	✓			✓	✓							ADMM & other techniques	Two-timescale reconfiguration	Max. expected utility of STs	✓	✓	✓	✓	✓	N/A	- Higher revenue [62] - Less unexpected reconfigurations [114]	- Mobility not supported - Power resources not considered
TN+CN	[153]		★★★		★	★		✓			✓		✓				✓			SNC-based heuristic, clustering	Online adjustment of slices	Min. sum of processing packets in all VNFs	✓	✓	✓	✓	✓	ns-3, Hadoop	- No unexpected congestion in TN [105]	- Mobility not supported
TN+CN	[154]		★★★			★				✓			✓							PSO	Traffic uncertainty-aware	Max. resource utilization for various services		✓	✓		✗	Matlab, ROME [163]	- Better utilization & execution time [164]	- Fairness among different slices overlooked - Mobility not supported
TN+CN	[155]		★★		★			✓					✓	✓	✓					PPO	MEC	Max. resource efficiency & Max. QoS satisfaction	✓	✓	✓		✓	Python, PyTorch	- Better utilization of resources [156]	- Scalability - Inaccurate resource utilization due to ignoring RAN

***VNF P.**: VNF Placement, **Reconf.**: Reconfigurability, **Orch.**: Orchestration, **Stak.**: Stakeholders, **Opt.**: Optimization-based, **Var./Con.**: Variables and Constraints categorization. *Significance of solutions in their supported functionalities are ranked from ★ (basic) to ★★★ (advanced).

B. RA

Considering resources acrosss all domains makes E2E resource provisioning more intricate and susceptible to scalability issues. This section examines the progress made on generic and E2E RA before discussing the open challenges.

*1) **Generic frameworks**:* This category presents a generic model for network resources across domains, with nodes and links having finite capacities. This section critically evaluates optimization- and ML-based methodologies.

Optimization-based approaches using complex network theory can generically model network elements and resources, enabling efficient placement and chaining of VNFs for diverse service deployments. However, developing effective heuristics to assist RA can be complex and may not capture the specificities of each domain due to the generic modeling. For instance, [167] models slices across domains and develops different heuristics for each service type (i.e., eMBB, URLLC, and mMTC). A subsequent study by the same group analyzes slice performance and proposes schemes to enhance RA by scaling infrastructure and node capacity [168]. The latter study proposes two heuristic schemes for adding more servers (i.e., scaling out) and increasing the capacity of forwarding nodes (i.e., scaling up).

ML-based techniques can leverage historical data to make RA decisions. DL approaches offer more scalability through the use of NNs. For example, [169] uses graph NNs in a network digital twin to predict the E2E latency of slices, which are composed of ordered VNFs. The trained model supports a heuristic scheme for RA.

The solutions discussed above ignore the specific resources in each domain (e.g., PRB allocation in RAN nodes) and mathematically model domain-specific generic resources. While this approach is theoretically effective, it cannot be used in practical deployments.

*2) **E2E solutions**:* Optimizing both intra- and inter-domain RA across all domains is more complex and poses scalability and efficiency challenges that can be addressed using optimization- and ML-based approaches.

Optimization-based (e.g., closed-form, relaxation, and heuristic) methodologies can find near-optimal solutions to complex E2E RA problems. For instance, QT can be used to estimate E2E delays of slices, as demonstrated in [69] for an industrial private 5G scenario. Time-Sensitive Network (TSN)-based Ethernet switches are used between nodes to provide more reliable smart factory services. A relaxed uncertainty-aware ILP formulation of InP revenue maximization is proposed by [170] in the presence of legacy (i.e., non-NS) services. In addition, a non-convex RA problem is introduced by [171] to minimize energy consumption in E2E NS. The authors transform their problem into a convex quadratic programming problem that can be solved using state-of-the-art optimization tools (e.g., CVX [64]). Their scheme outperforms a RAN slicing benchmark (i.e., [172]) by 20% in energy consumption and 30% in bandwidth usage. Nevertheless, this work presupposes the existence of only one backhaul link, which is not applicable in extensive deployments.

ML-based techniques can capture the hidden features and complexities of E2E resources. However, dealing with the high dimensionality of E2E RA problems is arduous. ML-based traffic forecasting can help optimize RA problems. For instance, in [173], a soft Gated Recurrent Unit (GRU) predicts slice traffic and optimizes E2E RA under non-convex SLA constraints. Using the prediction model, a heuristic balances slice isolation and resource over-provisioning by adjusting the Lagrange multiplier radius of the constraints. Model-free DRL approaches dynamically account for environmental changes but are sensitive to hyperparameter tuning and state/action determination. In [174], radio, power, transport, and bandwidth resources are jointly considered to minimize network usage and SLA violations. The E2E orchestration agent and domain agents use the PPO algorithm to allocate both continuous and discrete resources, outperforming DDPG agents used by cross-domain solutions (e.g., RAN+TN RA in [148]). Performance can be further improved by capturing mutual effects between resources across domains using FL-based methodologies.

*3) **Open challenges**:* Controlling and reducing the state and action spaces are essential for improving the tractability of E2E RA schemes. Several techniques can be used to achieve this, including (i) splitting the problem into sub-problems (e.g., [133]), (ii) constraint relaxation (e.g., [166]) in optimization problems, (iii) experience replay memory (e.g., [142]), and (iv) deep dueling architecture (e.g., [175]) in DRL-based problems. Nevertheless, this should not affect the quality of the solutions, but rather, a trade-off should be sought between reducing complexity (i.e., shortening response time) and increasing accuracy (i.e., exploring most of the action/state space). The effectiveness of E2E RA mechanisms hinges on factors such as training data quality and quantity, as well as hyperparameter tuning, which are still challenging for researchers.

Finally, E2E VNF placement, which is strongly coupled with RA, still requires many enabling technologies (e.g., NFV, MEC, and SDN) to mature across all domains. For instance, the recent virtualization of RAN functions in the Open RAN paradigm [53] calls for a more efficient RA and a simpler orchestration architecture for E2E deployments.

C. Reconfigurability

E2E NS RA frameworks can be augmented with a *reconfigurability* capability, which would make them more efficient and practical. Hierarchical orchestration could be a key feature to achieve that; see Sec. (IV-A3) for cross-domain frameworks. Balancing efficiency and optimality is the main challenge in these frameworks. This section reviews the progress made in this area before outlining the open issues.

*1) **Generic frameworks**:* Optimization-based heuristics can be used to dynamically reconfigure cross-domain resources. For instance, domains and their resources are generically modeled in [176] to perform three tasks: (i) initial intra-domain RA, (ii) inter-domain delay budget redefinition, and (iii) inter-domain reconfiguration. While inter-domain interactions are well covered, intra-domain reconfiguration is not considered, which reduces efficiency.

*2) **E2E solutions**:* Dynamic E2E reconfiguration has not yet been thoroughly investigated in the literature. Optimization- and DRL-based schemes can be used for this purpose. For instance, an online joint AC+RA mechanism proposed in [166] uses heuristics to reconfigure intra- and inter-domain resources. A decision period is assumed to reconfigure the resources previously allocated to other slices to cope with the dynamicity of slice requests. The proposed E2E methodology outperforms single-domain reconfigurability frameworks (i.e., CN [62] and RAN [111]) in terms of slice acceptance rate.

*3) **Open Challenges**:* Considering all domains can lead to high reconfiguration times for the operational slices, affecting their QoS levels. Reducing the problem action/state space by using a partial exploration policy [113] and DDPG [80] can help solve these problems. Still, their impact on reconfiguration efficiency should be minimized. A trade-off between increasing resource usage efficiency and meeting end-user requirements is needed. Advanced AI/ML techniques (e.g., baseline switching in imitation learning [177] and XAI [101]) can help strike a balance, but training should be done in a test environment to avoid unprecedented failures.

D. Orchestration

Orchestration between domains is necessary for E2E NS frameworks to coordinate their resource management functionalities (e.g., RA). Since configuring an E2E slice requires adjusting many variables across all domains, utilizing traditional model-based or closed-form solutions to orchestrate resources can be complicated. This section evaluates the methodologies used to achieve that before discussing the open issues.

*1) **Generic frameworks**:* To achieve highly interoperable systems, orchestration architectures that connect multiple technical and administrative domains are highly desirable. A balance between the level of abstraction, coordination overhead, and security should be maintained to enable efficient resource provisioning. For example, individual orchestrators associated with each of the considered domains can interact in a federated manner, as in [178], where Open-Source MANO (OSM) achieves interoperability between different sites by means of standard interfaces of NFV Orchestrator (NFVO) and Service

Orchestrator (SO) [179]–[181]. A recent proposal constructs a MANO platform for multi-InP environments using a Permissioned Distributed Ledger (PDL) Blockchain architecture [182]. This platform connects SOs of different InPs, introducing competition, transparency, and resource redundancy, resulting in CAPEX/OPEX savings for MNOs. While Blockchain-based orchestration brings the aforementioned advantages, it can delay decision-making due to coordination overhead. The reviewed architectures, albeit useful, still need to specify how each domain manager can abstract and expose its resources to the cross-domain orchestrator.

*2) **E2E solutions**:* E2E orchestration frameworks, which support both intra- and inter-domain coordination, inherently face scalability challenges. Therefore, effective strategies are needed to address the massive size of their problem. Multi-agent DRL algorithms can be augmented with other approaches to alleviate this problem. The authors in [177] envision a hierarchical orchestration framework where each domain orchestrator (i.e., NSSMF) is a DRL agent that reports to the high-level orchestrator (i.e., NSMF). In addition to multi-agent DRL, other techniques are utilized to overcome the following issues in DRL-based E2E orchestration: (i) *safety DRL* is used to address free action space exploration and performance requirements; (ii) *imitation learning* is devised to accelerate the online learning process by imitating decisions from a baseline heuristic when the model has not yet converged. Distributed DRL techniques (e.g., [177]) are promising for partitioning the problem space but impose a communication overhead between agents that should be balanced with their efficiency and response time. Thus, there is still room to make these frameworks more practical.

*3) **Open challenges**:* Existing solutions have made good progress toward E2E orchestration but still face some open challenges. Orchestration of resources across different administrative domains is explored in a few works (e.g., [182]), but it should be effectively integrated with the RA modules associated with each of the technological domains. Stakeholders need to collaboratively orchestrate resources by enabling visibility to each other through Application Programming Interfaces (APIs), which is not straightforward due to the resulting signaling overhead and security concerns. To overcome these issues, a promising direction is to standardize a common and secure protocol for communication among stakeholders.

E. Security Considerations

This section reviews the progress on E2E NS security in the context of resource management, together with the associated challenges. All the works found are generic trust frameworks, and there is a lack of E2E solutions that operate in both intra- and inter-domain interactions in the literature.

*1) **Generic frameworks**:* Given that E2E slices may span various technological and/or administrative (sub-)domains, trusted relationships between stakeholders should be established through SLAs. Blockchain technology supports the negotiation and provision of E2E slices and the management of SLAs in multi-stakeholder scenarios. For example, in the trust architecture proposed in [183], a resource selection problem is formulated as an ILP to minimize the cost and maximize the reputation of InPs. Based on this, a smart contract algorithm is designed to manage stakeholder negotiations. In [184], a Blockchain consensus scheme is proposed to ensure security and information consistency across nodes from different domains. Their scheme involves a bilateral evaluation mechanism based on GT that suppresses malicious behavior during orchestration between consensus nodes. While these works contribute to securing inter-domain and multi-stakeholder NS orchestration, their lack of support for non-virtualized resources (e.g., power in RAN) among other specific resources in each domain limits their applicability to actual NS deployments.

*2) **Open challenges**:* Blockchain-based orchestration platforms can significantly increase security and reliability levels by involving a large number of nodes in their consensus algorithm, but this comes at the cost of increased coordination overhead, which can severely impact the performance of critical NS functionalities (e.g., RA and orchestration). The balance between security and performance is still an ongoing challenge that requires more practical implementation.

F. Summary

There are fewer works presented in this section compared to Secs. III and IV, indicating that research on E2E NS frameworks is still in its infancy. Many of the cross-domain solutions described in Sec. IV claim a form of E2E NS support. However, some domains are either not modeled or oversimplified (e.g., TN in [80] and [129]). Fig. 7 summarizes the shortcomings and open issues discussed in Secs. III, IV, and V, respectively. Many of the challenges associated with E2E NS frameworks still apply to cross-domain solutions. In addition, as the enabling technologies (e.g., SDN, NFV, MEC, TN slicing [136], and RAN virtualization [53]) mature and become more standardized, it will be easier to implement more reliable E2E NS frameworks with more features.

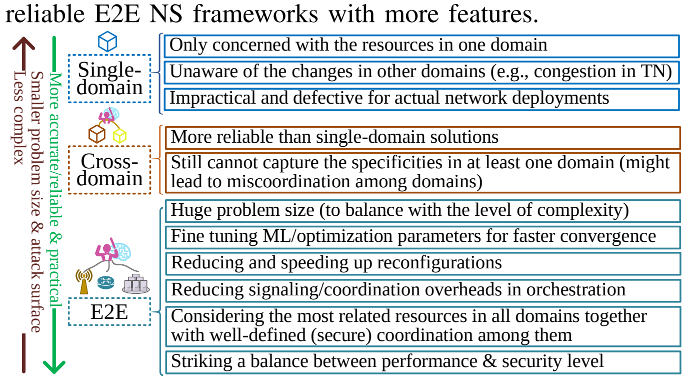

Fig. 7: Shortcomings and open challenges in NS frameworks.

Table V provides a summary of the E2E solutions discussed in this section. The main findings are that no existing work has addressed the need for automation from an E2E perspective, and reconfigurability and security considerations still need further development. In summary, future research should account for the available resources across all domains (i.e., RAN, TN, and CN) to achieve the most effective and efficient E2E NS framework, regardless of the specific functionality being targeted.

VI. Research Projects and Experimental Testbeds

This section summarizes the latest related research projects and experimental frameworks in the context of NS.

TABLE V: A summary of the literature on E2E NS solutions

Reference	Functionality						Stak.*			Methodology									Algo-rithm(s)	Focus	Specific use case	Objective(s)	Var.*		Con.*		Online	Simu-lation Tools	Scope		Improvement vs. single- & cross-domain solutions	Shortcomings	
										Opt.*			ML																			Ignored resources & oversimplifying assumptions	Open Challenges
	AC	RA	VNF P.*	Reconf.*	Orch.*	Security	MNO-ST	MNO-User	InP-ST	Closed	Relaxation	Heuristic	RL	DL	SL	UL	FL	GT-based					Continuous	Discrete	Linear	Non-linear			Intra-domain	Inter-domain			
[165]	★★				★★		✓				✓								Sequential quadratic programming	SLA decomposition	N/A	Min. overall risk of slice rejection	✓		✓	✓	✗	Java [185]	○	●	- More reliable AC [127]	- RAN: no PRBs & power resources - CN: no computing capacity	- Predicting future failures using ML models to perform more accurate AC
[60]	★★★				★★		✓	✓				✓							Multi-queue heuristic	Fairness	Elastic & inelastic slices	Max. fairness among slices		✓		✓	✓	C++	○	●	- Fairer slice AC [85]	- RAN: no PRBs & power resources - CN: no computing capacity	
[76]	★★		★★★	★	★★				✓		✓		✓			✓			SARSA	Slice prioritization		Max. revenue & Min. slice rejection ratio		✓	✓		✓	N/A	◐	●	- Lower rejection ratio [111]	- All domains: simplified NFs - RAN: no PRBs & power resources	
[166], [170]	★★★	★★★		★	★			✓	✓		✓	✓							Heuristics	Uncertainty in slice requirements	Video streaming	Max. no. of accepted slices & Min. total cost		✓	✓	✓	✓	Matlab, CPLEX	●	●	- More reliable AC [129] & cost estimate [127] - Less reconfigurations [137]	- RAN: no power resources	
[167], [168]			★★		★		✓				✓	✓							Heuristics	Complex network theory	eMBB, URLLC, mMTC	Max. utilization for eMBB/mMTC & Min. path length for URLLC		✓	✓		✓	Matlab, CPLEX	○	●	- Distinction of services [151] - More accurate resource utilization [155]	- RAN: no PRBs - Unbounded latency	- Trade-off between complexity & problem size - Fine-tuning ML/optimization parameters - More mature enabling technologies
[169]		★★	★★		★★		✓					✓		✓	✓				GNN-based heuristic	Digital twin	NSFNET topology	Min. delay & Max resource efficiency	✓	✓	✓	✓	✗	Tensor-Flow, Stellar-graph	○	●	- Better delay prediction [186]	- RAN: no PRBs & spectrum allocation	
[69]		★★			★				✓	✓	✓								QT-based analyzer [187]	Private 5G	Smart factory	Max. reliability of E2E delay estimation	✓		✓		✓	N/A	●	●	- Better delay prediction [186] - Can provide E2E delay bound [151]	- TN: no isolation in nodes	
[171]	★	★★	★★★		★★			✓			✓								Convex quadratic programming	Energy Efficiency	mMTC, eMBB	Min. total energy consumption	✓		✓	✓	✗	Matlab, CVX	●	●	- Better bandwidth consumption [172] - More accurate delay & utilization [147], [154]	- RAN: no PRBs - TN: oversimplified backhaul	
[173]		★★			★	★	✓				✓	✓		✓	✓				GRU, heuristic	Traffic forecasting	Heterogeneous web apps	Max. resource over-provisioning	✓	✓	✓	✓	✗	Python, Tensor-Flow	◐	●	- Insight about future traffic [71]	- RAN: no PRBs	
[174], [177]		★★★			★★★			✓				✓	✓	✓					Multi-agent PPO	Hierarchical distributed DRL	Heterogeneous mobile apps	Min. utilization & Min. SLA violations	✓	✓	✓		✓	Python, PyTorch, Testbed	●	●	- Faster convergence [148] - No bottlenecks in TN [132], [142]	- CN: no storage capacity	
[176]		★		★★	★★		✓				✓	✓							Heuristic	Fairness in achieved latencies across slices	URLLC	Min. resource utilization & total delay	✓		✓	✓	✓	Matlab	◐	●	- Less unnecessary reconfigurations [111]	- RAN: no PRBs & power resources - CN: no computing & memory resources	- Standard & secure APIs among stakeholders - Reducing signaling overheads - Balancing security level & performance
[182]					★★	★	✓		✓			✓							Quorum Block-chain	PDL architecture	Smart manufacturing	Ensuring trust among InPs/MNOs	N/A				✓	Quorum Block-chain	○	●	- Usability in multi-stakeholder scenarios [143]	- RAN: no PRBs & power resources	
[183]		★			★	★			✓		✓							✓	ILP, heuristic	Smart contracts	URLLC, eMBB	Ensuring SLA among stakeholders		✓	✓		✓	Gurobi, Ethereum	◐	●	- Usability in multi-stakeholder scenarios [142]	- RAN: no PRBs & power resources - CN: no computing capacity - TN: oversimplified backhaul	
[184]					★★	★★			✓			✓						✓	GT-based heuristic	Fairness, Block-chain	Data center	Min. resource capacity gap	N/A				✓	CCF, Open-Enclave	○	●	- Orchestration & security considered [82]	- RAN: no PRBs - Non-cooperative agents not considered	

*VNF P.: VNF Placement, **Reconf.**: Reconfigurability, **Orch.**: Orchestration, **Stak.**: Stakeholders, **Opt.**: Optimization-based, **Var./Con.**: Variables and Constraints categorization.
Significance of solutions in their supported functionalities are ranked from ★ (basic) to ★★★ (advanced).
For intra- and inter-domain support, ○, ◐, and ● represent '*not supported*', '*partially supported*', and '*fully supported*', respectively.

TABLE VI: Summary of ongoing/recent NS projects.

Initiative	Country	Project Title	Dom.* CN	Dom.* TN	Dom.* RAN	Functionality AC	Functionality RA	Functionality VNF P.*	Functionality Reconf.*	Functionality Orch.*	Functionality Security	Stak.* MNO-ST	Stak.* MNO-User	Stak.* InP-ST	Relevance to NS	Use case(s)	Status	Relevant Publications
5G PPP Phase 3 (2021-24)	European countries	5G ZORRO: Zero-tOuch secuRity & tRust for ubiquitous cOmputing & connectivity in 5G networks	✓	✓		★★	★			★★	★	✓		✓	Cross-domain security & trust between stakeholders	Smart contracts	Finished (Q4 2022)	[188]–[190]
		MonB5G: Distributed management of network slices in beyond 5G	✓	✓			★★	★★		★★	★			✓	Zero-touch orchestration & administration of slices	Tactile Intenet	Finished (Q4 2022)	[83], [115], [182], [151], [191]–[194]
		TeraFlow: Secured autonomic traffic management for a Tera of SDN flows		✓			★★★		★	★★★	★	✓		✓	Integrating heterogeneous TN resources in a multi-layer SDN controller	IoV	Finished (Q3 2023)	[195]–[197]
		5GMediaHUB: 5G experimentation environment for 3rd party media services		✓	✓	★	★★		★	★★			✓		Cross-domain orchestration of slices in testbed sites to validate various apps	Immersive AR/VR/XR	Ongoing - ends Q2 2024	[198]–[200]
		6G BRAINS: Bringing Reinforcement learning Into Radio Light Network for Massive Connections	✓	✓	✓	★	★★	★★	★★	★★★		✓	✓	✓	Implementing IoT testbeds within an NS framework	Smart factory	Finished (Q4 2023)	[201]–[203]
		MARSAL: ML-based networking & computing infrastructure resource management of 5G & beyond intelligent networks		✓	✓		★			★★	★★			✓	A suitable SDN control plane for NS, along with a security & trust framework for multi-tenant infrastructure	Content distribution	Ongoing - ends Q2 2024	[204]
		VITAL-5G: Vertical Innovations in Transport And Logistics over 5G experimentation facilities	✓	✓	✓		★★		★	★★		✓		✓	A slice orchestration platform to support three testbed sites	Transport & logistics	Ongoing - ends Q2 2024	[205], [206]
		B5G-OPEN: Beyond 5G - OPtical nEtwork coNtinuum		✓			★★		★★	★				✓	Enabling slicing control in the Multi-Band & optical TN domains	Video streaming	Ongoing - ends Q4 2024	[207]
		SEMANTIC: E2E Slicing and data-drivEn autoMAtion of Next-generation cellular neTworks with mobIle edge Clouds	✓		✓		★			★★	★		✓		Zero-touch NS automation with a focus on MEC offloading with FL	eMBB	Ongoing - ends Q3 2024	[173], [208]
6G SNS Phase 1 (2023-25)		6Green: Green Technologies For 5/6G Service-Based Architectures	✓		✓			★★★	★★	★★		✓			Promoting energy efficiency in cross-domain slices between stakeholders	AR	Initial - ends Q4 2025	[209]
		PRIVATEER: Privacy-First Security Enablers For 6G Networks	✓	✓	✓		★	★	★	★★	★★			✓	Privacy-aware slicing & orchestration in 6G	Logistics & smart city	Initial - ends Q4 2025	[210]
		Hexa-X-II: A holistic flagship towards the 6G network platform & system	✓	✓	✓	★★	★★★	★★	★★	★★	★★	✓	✓	✓	A flagship project for 6G development, including NS as a key enabler	XR & collaborative robots	Initial - ends Q3 2025	[27]
		6G-XR: 6G eXperimental Research infrastructure to enable next-generation XR services	✓		✓		★★			★★				✓	Building a multi-site 6G testbed with E2E slicing capabilities	Immersive XR	Initial - ends Q4 2025	[211]
		TrialsNet: TRials Supported By Smart Networks Beyond 5G	✓		✓				★	★★		✓		✓	Innovative 6G apps on network digital twins based on dynamic NS management	eHealth, smart transport	Initial - ends Q4 2025	[212]
		6G-BRICKS: Building Reusable testbed Infrastructures for validating Cloud-to-device breaKthrough technologieS	✓	✓	✓		★★	★★	★★	★★★	★		✓	✓	Open RAN-compliant NS testbed	Metaverse, Smart factory	Initial - ends Q4 2025	[101], [213], [214]
		ACROSS: Automated zero-touch CROSS-layer provisioning framework for 5G and beyond vertical services	✓	✓	✓		★★	★★★	★★	★★★	★	✓		✓	Automated and trusted orchestration among multiple administrative/technological domains	eHealth, etc.	Initial - ends Q4 2025	[215], [216]
PAWR	US	POWDER: Platform for Open Wireless Data-driven Experimental Research	✓	✓	✓		★★★			★★★		✓	✓		Open RAN-compliant xApp development for RAN slicing	Research testbed	Available since 2019	[217], [218]
		Colosseum: The World's Most Powerful Wireless Network Emulator			✓		★★★	★★	★	★★		✓	✓		Open RAN-compliant NS testbed	Research testbed	Available since 2022	[219]–[221]
National Projects	Germany	6G-ANNA: 6G Access, Network of Networks, Automation, and Simplification			✓	★	★★	★★	★★	★★	★★	✓	✓	✓	AI-enabled Cloud-based RAN architecture supporting NS	Smart factory	Initial - ends Q3 2025	[222]
		6GEM: 6G research hub for open, efficient and secure mobile communications systems		✓	✓	★	★★	★★	★★	★★★	★★★	✓	✓	✓	Providing various testfields to evaluate NS frameworks	Port logistics, Smart hospital	Ongoing - ends Q4 2025	[223]–[227]
	Spain	6GDAWN: Decentralized AI and Architectures for Massive Wireless NS Scalability and Sustainability	✓		✓	★	★★★		★★	★★	★★★	✓	✓		NS utilization and security with explainable AI	Monitoring	Ongoing - ends Q4 2024	[191], [228]
		6G-CHRONOS: AI-assisted beyond 5G-6G arCHitectuRe with deterministic netwOrking for iNdustrial communicatiOnS		✓	✓	★	★★		★★	★	★★		✓		Exploring the usage of NS using 5G & TSN technologies	Smart factory	Ongoing - ends Q4 2024	[229], [230]
	UK	BEACON-5G: Building REconfigurable, Agile, SeCure, & TrustwOrthy Systems for OpeNness in 5G	✓		✓		★★		★	★★	★	✓	✓	✓	Developing a RAN slicing solution based on Open RAN and integrating it with existing NS solutions	Smart city, IoV, & smart healthcare	Finished (Q3 2023)	[231]
	Brazil	SFI2: Slicing Future Internet Infrastructures	✓			★	★★	★★★	★★	★★	★★★		✓		Integrating several national projects into an automated NS testbed	IoV	Ongoing - ends Q2 2024	[39], [122], [232]–[234]

***Dom.**: Domains, **VNF P.**: VNF Placement, **Reconf.**: Reconfigurability, **Orch.**: Orchestration, **Stak.**: Stakeholders.
Significance of solutions in their supported functionalities are ranked from ★ (basic) to ★★★ (advanced).

A. Research Projects

The state-of-the-art research projects on NS are briefly presented in this section.

In Europe, some projects in the third phase of the 5G Infrastructure Public Private Partnership (5G PPP) of the European Union's Horizon 2020 program are still ongoing [235]. More recent European projects are defined under the European Smart Networks and Services (6G SNS) Joint Undertaking, a part of the Horizon Europe funding program [236]. The architectural landscape of 5G/6G networks collected by these initiatives can be found in [10], [237]. AI/ML algorithms are harnessed in 5GZORRO, MonB5G, 6G BRAINS, MARSAL, SEMANTIC, and ACROSS projects to enhance the RA and orchestration functionalities in NS frameworks with multiple technological and administrative domains [173], [188], [192], [202], [204], [215]. Enabling a trusted, secure relationship among various stakeholders in the 5G ecosystem to use cross-domain NSs is presented in 5GZORRO and PRIVATEER projects [188], [210]. The integration of different TN technologies (e.g., wireless and optical) into NS by monitoring, programming, and orchestration is being investigated in the TeraFlow and B5G-OPEN projects [195], [207]. In addition, 5GMediaHUB, 6G-XR, 6G-BRICKS, and TrialsNet are implementing NS-enabled testbeds to investigate the feasibility of deploying novel use cases (e.g., Augmented/Virtual/Extended Reality (AR/VR/XR)) [198], [211]–[213].

Platforms for Advanced Wireless Research (PAWR) is an initiative funded by the National Science Foundation (NSF), a U.S.-based agency that provides four city-scale testbeds for 5G and beyond networks [238]; one of them is called

TABLE VII: Components used in state-of-the-art NS testbeds

Domain	Module	Software/Hardware components	Used in
CN (&NFV)	NFVO & VNFM	OSM [239]	[240]–[244]
		ONAP [245]	[246]
	VIM	OpenStack [247]	[240], [241], [244]
		LXC and Docker	[174], [248], [249]
	CN VNFs	OpenAirInterface (OAI)-CN [250]	[144], [174], [240]
		Open5GS [251]	[71]
TN	SDN Controller	ONOS [252]	[249], [253]
		OpenDaylight [254]	[174], [177]
		Ryu [255]	[248]
	Emulation	OVSs as emulated switches [256] & the OpenFlow SBI [257]	[174], [244], [248], [253]
RAN	Radio Front-end	Software-Defined Radio (SDR) e.g., Universal Service Radio Peripheral (USRP) and Zynq	[138], [144], [174], [242], [243], [248], [249]
	RAN NSSMF	FlexRAN [258]	[174], [242], [249]
	Emulation	OAI [250]	[144], [242], [244]
		UERANSIM [259]	[260]

POWDER [217], which includes a set of O-RAN xApps that support NS. Another PAWR platform, Colosseum, consists of the OpenRAN Gym project, which supports RAN slicing in an Open RAN-compliant framework [219].

Several national R&D projects around the world, such as 6G-ANNA [222], 6GEM [223], 6G-CHRONOS [230], 6GDAWN [191], and SFI2 [122], investigate NS-related challenges. These initiatives emphasize AI/ML-based automation, XAI, security, and smart factories in NS.

Table VI provides a brief summary of the relevant ongoing (and some recently finished) projects. Since 6G SNS projects are in the initial stages, their findings have not been widely published. As seen in Table VI, orchestration is covered by all recent projects due to its importance in cross-domain and E2E NS. For further studies, already-finished NS-related projects are reviewed in [18], [47]. Although R&D projects promote E2E NS frameworks, relevant SDOs each focus on a particular scope (e.g., IETF focuses on TN slicing [136]), and harmonization across them is urgently required [19].

B. Experimental Testbeds

This section presents the most mature experimental NS testbeds. The state-of-the-art components of the presented testbeds are briefly listed. Then, the experimental works covering single- and cross-domain NS are described before reviewing the few available E2E testbeds.

Table VII shows some of the mostly used software and hardware components in NS testbeds. However, the list is not exhaustive, and the reader is referred to [44] for a comprehensive list of NS experiments.

Most of the available testbeds support one single domain. For instance, in the CN domain, the experimentation facility of [240] implements resource forecasting and dynamic scaling of VNFs. In the TN domain, an experimental NS framework for optical metro networks that utilizes ONOS is proposed in [241]. In the RAN domain, the testbed constructed in [243] introduces a virtualization system where STs can either (i) choose one of the available virtual baseband units or (ii) specify a set of available VNFs for a desired NS functionality. In addition, [219] introduces an NS-enabled Open RAN-compliant testbed for the development of ML-assisted xApps. However, as other domains are not considered in these solutions, they are not practical for more realistic NS scenarios.

Several testbeds offer a form of support for cross-domain NS. For instance, in a smart factory use case with time-critical applications, an NS orchestration system that uses a TSN control plane within an Xhaul scenario is implemented and analyzed [242]. A two-level spectrum allocation mechanism is proposed in [244], where a master Medium Access Control (MAC) optimizes the number of spectrum resources among different slices, and multiple slave MACs are in charge of scheduling the resource in their associated slices. A demonstration with proprietary hardware and open-source software is conducted without integrating an allocation scheme. Furthermore, the usage of the Open Network Automation Platform (ONAP) orchestrator in a RAN+CN NS testbed is studied in [246] from an architectural perspective, but no numerical analysis is presented. A Mininet-based Python package is introduced in [261] to facilitate KPI management and monitoring in NS simulation. Moreover, a Cloud-native Lightweight Slice Orchestration (CLiSO) framework is proposed in [214] to enable container-based interaction of RAN, edge, and CN.

Very few testbeds support E2E NS. In [248], an E2E orchestrator, called Hyperstrator, is implemented with well-defined domain orchestrators, each of which interacts with the domain managers (i.e., PyLXD controller for CN, Ryu controller for TN, and OpenWifi controller for RAN) to access the underlying resources. This work also includes an analysis of NS deployment time and an empirical evaluation of the optimal resource utilization ratios in each domain to yield better performance in terms of throughput and delay. However, their testbed generalizes the TN components with Open vSwitch (OVS) containers, which cannot emulate optical switches used in large-scale networks.

Table VIII provides a summary of the experimental solutions discussed in this section, including some papers marked as testbeds in Tables IV and V (i.e., [71], [138], [144]). Compared to the other studied testbeds, Hyperstrator is more modular and can be used in future studies due to its open-source access. As seen in Table VIII, only a few testbeds provide open-source implementations with limited documentation for further research.

VII. Challenges and Future Directions

This section highlights the ongoing challenges in E2E NS frameworks and possible future directions to overcome them.

A. E2E Orchestration Across Technological and Administrative Domains

Orchestrating and interworking among diverse administrative domains continues to pose a challenge that requires further investigation as it incorporates multiple parties who must all trust one another and communicate through common interfaces. Governments may need to intervene in selecting/developing these interfaces to ensure consensus among all stakeholders.

As investigated in Secs. III, IV, and V, many studies fail to integrate all available resources associated with slices across

TABLE VIII: A summary of experimental NS testbeds.

Reference	Dom.* CN	Dom.* TN	Dom.* RAN	Functionality AC	Functionality RA	Functionality VNF P.*	Functionality Reconf.*	Functionality Orch.*	Functionality Security	Stak.* MNO-ST	Stak.* MNO-User	Stak.* InP-ST	Project	Focus	Use Case	Shortcomings	Orchestrator	Used RF	Open-Source?
Dynamic CN scaling [240]	✓				★★		★	★		✓			N/A	Predictive scaling of CN VNFs	DC Traffic Dataset	Only supporting NFV-enabled resources	OSM	✗	✗
METRO-HAUL [241], [253]		✓			★	★		★★				✓	METRO-HAUL	Optical metro networks	Crowdsourced live video streaming	No direct end-user support	OSM	✗	✗
AIRTIME [243]			✓		★	★★		★★	★	✓			ORCA	- Virtualizing RAN - Flexibility for slice requests	eMBB, IoT	Possible congestion in TN nodes	OSM	USRP B210	✓
OpenRAN Gym [219], [220]			✓		★★★	★★	★	★★		✓	✓		Colosseum	Open RAN compliant RAN slicing	RAN slice scheduling xApp	No integration with TN & CN nodes	OrchestRAN [221]	USRP X310 & B210	✓ ✓
TSN-based NS [242]		✓	✓		★★			★★			✓		5G-Victori	Achieving more reliability using TSN techniques	eMBB, URLLC in a smart factory	- Considering few slices - Only suitable for private 5G networks	JOX slice orchestrator [262]	USRP (Undeclared)	✗
Two-level RA of NSs [244]	✓		✓		★★		★	★			✓	✓	N/A	Two-level spectrum allocation	eMBB, URLLC, mMTC	Not guaranteeing performance isolation	Proprietary	Proprietary	✗
NS Mgmt. w/ ONAP [246]	✓		✓					★★	★	✓			N/A	NS LCM	Best-effort	No results	ONAP	Undeclared	✗
Slicenet [261]	✓		✓	★	★			★★	★	✓			N/A	Monitoring	eMBB, URLLC, mMTC	Disregarding TN resources	Slicenet	Undeclared	✓
CLiSO [214]	✓		✓		★★			★★★	★	✓			6G-BRICKS	Cloud-native orchestration	eMBB, URLLC, mMTC	Disregarding legacy VNFs	Kubernetes-based	USRP B210	✗
IBN-based NS LCM [138]	✓		✓		★★		★	★★		✓			N/A	- NS LCM automation - IBN	IoT, GBR*, Non-GBR slices	Scalability issues	OSM	USRP B210	✗
ML-assisted secure NS [144]	✓		✓		★			★	★★		✓		N/A	Balancing isolation & utilization	N/A	Only supporting hardware-based isolation	ML-assisted resource orchestrator	USRP B210	✗
2-tier resource slicer [71]	✓		✓		★★			★		✓			5TONIC	Latency analysis in C-RAN	MEC-based streaming	- Scalability issues - No routing support	Tacker	Proprietary	✗
NexRAN [218]		✓	✓	★★	★★			★★		✓	✓		POWDER	RAN slicing xApp for O-RAN	Various RIC use cases	CN NFs not integrated	O-RAN RIC	X310 & B210	✓
Arena [263]		✓	✓	★	★★				★	✓	✓		Arena	Spectrum sharing	IoT	RAN computing resources not considered	N/A	X310 & N210	✗
Hyperstrator [248]	✓	✓	✓		★★			★★★		✓			ORCA	- Hierarchical Orchestration - Microservices	N/A	Not supporting optical TN nodes	Hyperstrator	Zynq SDR	✓
OnSlicing [174], [177]	✓	✓	✓		★★★			★★★			✓		N/A	Online DRL-based policy configuration	Heterogeneous mobile apps	- Scalability issues - Few slices considered	Distributed DRL-based Orchestrator	USRP B210	✗
UWS NS manager [201], [203]	✓	✓	✓	★	★★	★★★	★	★★★		✓		✓	6G BRAINS	Topology-aware orchestration	Industrial IoT	Scalability issues	ONAP	USRP X310	✗
Dynamic NS mgmt. [205], [206]	✓	✓	✓		★★		★	★★		✓		✓	VITAL-5G	LCM & dynamic orchestration of 3 testbed sites	Transport & logistics	Not using ML-based NS mgmt.	VITAL-5G Orchestrator	Undeclared	✗

*Dom.: Domains, VNF P.: VNF Placement, Reconf.: Reconfigurability, Orch.: Orchestration, Stak.: Stakeholders, GBR: Guaranteed Bit Rate. Significance of solutions in their supported functionalities are ranked from ★ (basic) to ★★★ (advanced).

technological domains (i.e., RAN, TN, and CN). Resources can be allocated/reconfigured in different timescales, and the E2E orchestrator should account for this heterogeneity (see Sec. IV-A4). Additionally, as 5G-Advanced and 6G networks increasingly incorporate new nodes (e.g., UAVs) and segments (e.g., Non-Terrestrial Networks (NTNs)) [97], NS orchestrators must effectively integrate all these elements. Advanced capabilities, such as automation, security mechanisms, and reconfigurability, can be leveraged to support E2E orchestration.

B. Integrating Open RAN into E2E NS

The Open RAN paradigm has recently emerged to transform the RAN into an open, intelligent, virtualized, and interoperable system. In the context of E2E NS, Open RAN could increase the flexibility and efficiency of resource management [264]. For example, with its disaggregated and modular functionalities, this architecture can accommodate the RAN slicing requirements for isolation, scaling, and dedicated processing [90], [220]. Remaining challenges include interoperability with state-of-the-art orchestrators, optimizing performance, KPI data provisioning, and ensuring security in open and multi-vendor environments [265]. Furthermore, to establish the optimal policy for parameter adjustment, there is a need to enhance conflict mitigation among different xApps (e.g., PRB scheduler and interference mitigator) and rApps (e.g., higher-level SLA assurance) operating in the near-RT and non-Real-Time (non-RT) RICs, respectively. Future research should prioritize these facets, alongside effectively integrating

Open RAN with other key technologies such as AI/ML, MEC, and Blockchain for efficient and secure E2E NS.

C. Scaling up NS Frameworks

It is paramount for an NS framework to scale properly since the problem state and action spaces can be huge. Most of the reviewed papers provide evaluations based on simulations but cannot scale up in actual network deployments. Moreover, even if the scalability of these frameworks is improved, unforeseen circumstances can always arise in real large-scale networks. One promising way to overcome such challenges is to build a network digital twin across domains that enables virtual testing of E2E NS solutions before deploying them in production networks [46], [169], [266].

D. AI/ML-driven Automation

As presented in Tables III, IV, and V, there are only a few works focused on automating the lifecycle of NSs [115], [138], which do not consider all technological domains. The Zero-Touch Service Management (ZSM) architecture recently standardized by ETSI enables automation in NFV deployments, but it is still not fully supported across all technological domains. Besides, integrating ETSI ZSM [267] with Open RAN into a single NS automation framework can significantly increase efficiency and thus should be investigated in future works.

To introduce automation into NS frameworks, data should be collected and integrated from various nodes/NFs (e.g., O-RAN RIC, Network Data Analytics Function (NWDAF) [268], NFVO, NSMF, and NSSMFs), which is a significant challenge MNOs still face. Once the data collection and integration processes mature, the AI/ML-driven automation of the E2E NS lifecycle can become more efficient.

E. Large Language Model-assisted NS Configuration and Instantiation

With the recent emergence of Large Language Models (LLMs), such as GPT-4 [269], both low-level network configurations and higher-level service instantiations/modifications can be automated to a certain extent, making them more human-friendly [270]. LLMs can understand complex patterns in data, which can be leveraged to optimize network configurations for specific services. Given the practical complexities of deploying E2E NS, MNOs/STs require a highly skilled workforce to maintain NS-related services. Automating the NS lifecycle is not only desirable but also inevitable. A few vendors already offer limited NS automation. The use of customized LLMs can assist MNOs/STs in transitioning toward full automation of E2E NS. For instance, a customized LLM could be employed to translate the service requirements of STs received in the MNO's Operations Support System (OSS) during the NS preparation phase. The LLM could suggest a slice template to the ST and, upon approval, initiate the onboarding process.

F. The Role of Multi-agent DRL in E2E NS

DRL algorithms, adept at navigating intricate and dynamic environments, have gained prominence in NS for resource management [73]. Multi-agent DRL, where multiple agents learn to make decisions from unstructured input data, has displayed significant potential [271]. In the context of E2E NS, multi-agent DRL has the ability to enable efficient RA and orchestration across diverse technological domains without needing centralized control [177]. E2E NS suggests that these agents should function cooperatively. Nonetheless, interworking among various stakeholders may incorporate competitive agents. Balancing the above-mentioned trade-off, as well as developing a hybrid multi-agent framework[6] that can incorporate both technological and administrative domains, can be pursued in future research.

G. Distributed/Federated Learning for E2E NS

Distributed learning involves training models across multiple nodes, where each node learns from its local data. FL, a subset of distributed learning, takes this a step further by training a global model across multiple nodes, where each node learns from its local data and shares only the model updates, thereby preserving data privacy [78].

In the context of E2E NS, these techniques can facilitate efficient resource management across various technological domains without requiring centralized control or data exchange. This approach can enhance scalability, decrease communication overhead, and improve privacy preservation [79], [161]. However, addressing challenges such as managing data heterogeneity across different network nodes, ensuring resilience against malicious or straggling nodes, and optimizing the balance between global model performance and communication efficiency is crucial. Future research may concentrate on creating innovative algorithms and frameworks for these learning paradigms, taking into account the unique qualities and necessities of E2E NS.

H. Application of Imitation/Transfer Learning for E2E NS

Imitation learning can be particularly beneficial in E2E NS, where the complexity and dynamism of the network environment can make traditional learning approaches inefficient [272]. For instance, an NS agent can learn optimal slicing policies by observing and mimicking the actions of an expert agent (e.g., a heuristically guided agent), thereby reducing the exploration space and expediting convergence [177]. This is particularly useful in managing resources across various technological domains within an E2E context, where the number of state-action pairs may be exponential.

Transfer learning can boost the performance of E2E NS frameworks by leveraging knowledge acquired in one technological domain to improve learning in another [273], [274]. For instance, optimal RA strategies learned in one RAN area can be transferred to another, improving performance and accelerating convergence, especially in dynamic networks with

[6] A hybrid (or mixed) multi-agent framework deals with both competitive and cooperative agents.

fluctuating conditions [95]. Future works could explore the development of novel imitation and transfer learning algorithms specifically tailored for E2E NS.

I. Hierarchical Slicing

In hierarchical slicing, an ST is both a slice consumer and a producer. It initially receives a slice (i.e., NSI) from an MNO and then splits it into a few sub-slices that are offered to its own STs (see Fig. 8). In this case, the aggregated requirements of the created sub-slices should match the capabilities of the original slice. This problem, also known as hierarchical or recursive slicing [57], [136], has been barely investigated due to the initial focus on maturing NS as a technology. For instance, [275] proposes a form of hierarchical slicing for the RAN without support of other domains (i.e., TN and CN). To fully achieve hierarchical slicing, each slice request should have a separate slice orchestrator while ensuring harmonization between all orchestrators.

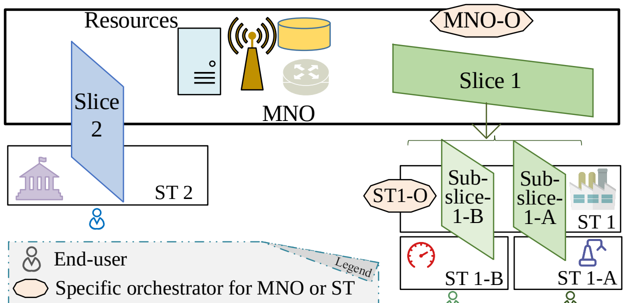

Fig. 8: An illustrative example of hierarchical slicing.

J. UE Slicing

Most of the current slicing mechanisms (e.g., [135]) do not account for more than one active slice for a given UE, as opposed to the standardized 8 active slices (see Sec. IV-D). The resource management and signaling overhead among multiple slices within each UE should be balanced with QoS provisioning and fairness between all UEs. Furthermore, achieving privacy and user-level isolation together (i.e., between multiple S-NSSAIs) may be elusive and require further research.

K. Promoting Open-Source NS Frameworks and Datasets

Leveraging existing work is necessary to advance research in NS. Using open-source simulation and experimentation facilities allows researchers to enhance E2E NS performance instead of creating new frameworks from scratch [276]. This strategy, similar to the approach adopted by the O-RAN Software Community (OSC), promotes efficiency and collaboration [264]. In our survey, we found that only a limited number of simulation platforms (e.g., [92], [129], [135], [142]) and experimental testbeds (e.g., [218], [220], [243], [248]) have publicly released their source code.

It is important to note that many of the AI/ML-based algorithms reviewed rely on synthetic, simulated data that may not fully capture the intricacies of real-world scenarios. The datasets currently used, which are often outdated and limited to the CN domain, do not adequately reflect the intricacies of E2E NS solutions. As a result, there is an urgent need for more comprehensive, up-to-date, and context-specific datasets to facilitate future research in E2E NS.

L. Harmonization of efforts

As introduced in Sec. VI, there is a need for SDOs and industry forums to work more collaboratively than before. In this context, IETF and O-RAN Alliance have been recently working on TN slicing [136] and RAN slicing [149], aligned with the 3GPP specifications [52], respectively. More effort in this direction will lead to more harmonized definitions and NS solutions, which will promote multi-vendor interoperability and facilitate the widespread adoption of NS functionalities.

VIII. Conclusions

This work has presented a survey of NS resource management with a focus on domain inter-dependence in cross-domain and E2E contexts. In addition to highlighting the shortcomings of single- and cross-domain NS frameworks, the open challenges facing E2E NS solutions are discussed, with a reflection on the most promising directions to overcome them.

The main lessons learned from our analysis include: (i) resource management should be conducted across heterogeneous technological domains throughout the slice creation and operation phases; (ii) relevant resources for NS in each domain (e.g., PRBs in RAN; link throughput in TN; computing capacity in CN) should be jointly considered; (iii) promoting self-optimization and self-(re)configuration in E2E NS is required through intelligent resource orchestration.

By evaluating different NS frameworks, we realize that the more domains a framework spans, the more comprehensive and reliable it is, but the more intricate and challenging it is to develop and deploy. In this context, AI/ML-assisted functionalities are increasingly needed to deal with the increased level of complexity. Due to its inherent flexibility, NS is likely to be a key pillar in 5G-Advanced and 6G networks. However, further improvements in cross-administrative domain orchestration, security mechanisms, and automation capabilities are needed to overcome the limitations of the current 5G NS frameworks.

Acknowledgments

This work was supported by the Centre for Future Transport and Cities (CFTC), Coventry University, Coventry, UK. The authors express their gratitude to the anonymous reviewers who contributed to enhancing the quality of the manuscript.

References

[1] GSMA, "Network Slicing Use Case Requirements Whitepaper," Apr. 2018.

[2] Ericsson, "Network slicing: A go-to-market guide to capture the high revenue potential," Ericsson, Tech. Rep., Mar. 2021. [Online]. Available: www.ericsson.com/49c844/assets/local/networks-slicing/docs/network-slicing-value-potential.pdf

[3] 3GPP, "Study on enhancement of network slicing; Phase 3 (Release 18)," 3GPP, TR 23.700-41, Dec. 2022.

[4] Fortune Business Insights, "Network Slicing Market Size, Share & COVID-19 Impact Analysis, By Enterprise Type, By End-User, and Regional Forecast, 2023-2030," Fortune Business Insights, Tech. Rep., Mar. 2023. [Online]. Available: www.fortunebusinessinsights.com/segmentation/network-slicing-market-107303

[5] Ericsson, "Network slicing: Top 10 use cases to target," Ericsson, Tech. Rep., Mar. 2021. [Online]. Available: www.ericsson.com/49c6d0/assets/local/networks-slicing/docs/top-network-slicing-use-cases.pdf

[6] I. Afolabi *et al.*, "Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.
[7] ITU, "ITU-T Technical Specification; FG NET-2030 – Focus Group on Technologies for Network 2030; Network 2030 Architecture Framework," Jun. 2020.
[8] 3GPP, "Architecture enhancements for dedicated core networks; Stage 2 (Release 13)," 3GPP, TR 23.707, Dec. 2014.
[9] NGMN, "5G White Paper," Feb. 2015.
[10] A. Gavras *et al.*, "5G PPP Architecture Working Group - View on 5G Architecture, Version 4.0," Oct. 2021.
[11] 3GPP, "Management and orchestration; Concepts, use cases and requirements (Release 18)," 3GPP, TS 28.530, Jan. 2024.
[12] V.-G. Nguyen, A. Brunstrom, K.-J. Grinnemo, and J. Taheri, "SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1567–1602, 2017.
[13] N. Saha, M. Zangooei, M. Golkarifard, and R. Boutaba, "Deep Reinforcement Learning Approaches to Network Slice Scaling and Placement: A Survey," *IEEE Communications Magazine*, vol. 61, no. 2, pp. 82–87, 2023.
[14] S. Huang, B. Guo and Y. Liu, "5G-Oriented Optical Underlay Network Slicing Technology and Challenges," *IEEE Communications Magazine*, vol. 58, no. 2, pp. 13–19, 2020.
[15] M. Zangooei, N. Saha, M. Golkarifard, and R. Boutaba, "Reinforcement Learning for Radio Resource Management in RAN Slicing: A Survey," *IEEE Communications Magazine*, vol. 61, no. 2, pp. 118–124, 2023.
[16] Y. Azimi, S. Yousefi, H. Kalbkhani, and T. Kunz, "Applications of Machine Learning in Resource Management for RAN-Slicing in 5G and Beyond Networks: A Survey," *IEEE Access*, vol. 10, pp. 106 581–106 612, 2022.
[17] S. E. Elayoubi, S. B. Jemaa, Z. Altman, and A. Galindo-Serrano, "5G RAN Slicing for Verticals: Enablers and Challenges," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 28–34, 2019.
[18] Y. Wu *et al.*, "A Survey of Intelligent Network Slicing Management for Industrial IoT: Integrated Approaches for Smart Transportation, Smart Energy, and Smart Factory," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1175–1211, 2022.
[19] M. Chahbar *et al.*, "A Comprehensive Survey on the E2E 5G Network Slicing Model," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 49–62, 2021.
[20] J. Ordonez-Lucena *et al.*, "On the Rollout of Network Slicing in Carrier Networks: A Technology Radar," *Sensors*, vol. 21, no. 23, 2021.
[21] A. Banchs, G. de Veciana, V. Sciancalepore, and X. Costa-Perez, "Resource Allocation for Network Slicing in Mobile Networks," *IEEE Access*, vol. 8, pp. 214 696–214 706, 2020.
[22] R. Su *et al.*, "Resource Allocation for Network Slicing in 5G Telecommunication Networks: A Survey of Principles and Models," *IEEE Network*, vol. 33, no. 6, pp. 172–179, 2019.
[23] F. Debbabi, R. Jmal, L. C. Fourati, and R. L. Aguiar, "An Overview of Interslice and Intraslice Resource Allocation in B5G Telecommunication Networks," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 5120–5132, 2022.
[24] J. Ordonez-Lucena *et al.*, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.
[25] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges," *Computer Networks*, vol. 167, p. 106984, 2020.
[26] F. Javed *et al.*, "Distributed Ledger Technologies for Network Slicing: A Survey," *IEEE Access*, vol. 10, pp. 19 412–19 442, 2022.
[27] M. A. Habibi *et al.*, "Toward an Open, Intelligent, and End-to-End Architectural Framework for Network Slicing in 6G Communication Systems," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1615–1658, 2023.
[28] E. Coronado *et al.*, "Zero Touch Management: A Survey of Network Automation Solutions for 5G and 6G Networks," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2535–2578, 2022.
[29] M. O. Ojijo and O. E. Falowo, "A Survey on Slice Admission Control Strategies and Optimization Schemes in 5G Network," *IEEE Access*, vol. 8, pp. 14 977–14 990, 2020.
[30] R. M. Dhanasekaran, J. Ping and G. P. Gomez, "End-to-End Network Slicing Security Across Standards Organizations," *IEEE Communications Standards Magazine*, vol. 7, no. 1, pp. 40–47, 2023.
[31] R. F. Olimid and G. Nencioni, "5G Network Slicing: A Security Overview," *IEEE Access*, vol. 8, pp. 99 999–100 009, 2020.
[32] R. Dangi *et al.*, "ML-Based 5G Network Slicing Security: A Comprehensive Survey," *Future Internet*, vol. 14, no. 4, 2022.
[33] F. Salahdine, Q. Liu and T. Han, "Towards Secure and Intelligent Network Slicing for 5G Networks," *IEEE Open Journal of the Computer Society*, vol. 3, pp. 23–38, 2022.
[34] C. D. Alwis *et al.*, "A Survey on Network Slicing Security: Attacks, Challenges, Solutions and Research Directions," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 534–570, 2024.
[35] S. Vassilaras *et al.*, "The Algorithmic Aspects of Network Slicing," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 112–119, 2017.
[36] 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. 162 748–162 762, 2020.
[37] H. P. Phyu, D. Naboulsi and R. Stanica, "Machine Learning in Network Slicing—A Survey," *IEEE Access*, vol. 11, pp. 39 123–39 153, 2023.
[38] J. A. Hurtado Sánchez, K. Casilimas and O. M. Caicedo Rendon, "Deep Reinforcement Learning for Resource Management on Network Slicing: A Survey," *Sensors*, vol. 22, no. 8, p. 3031, 2022.
[39] A. Donatti, S. L. Correa, J. S. B. Martins, and T. C. Carvalho, "Survey on Machine Learning-Enabled Network Slicing: Covering the Entire Life Cycle," *IEEE Transactions on Network and Service Management*, vol. 21, no. 1, pp. 994–1011, 2024.
[40] C. Ssengonzi, O. P. Kogeda and T. O. Olwal, "A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization," *Array*, vol. 14, p. 100142, 2022.
[41] S. D. A. Shah, M. A. Gregory and S. Li, "Cloud-Native Network Slicing Using Software Defined Networking Based Multi-Access Edge Computing: A Survey," *IEEE Access*, vol. 9, pp. 10 903–10 924, 2021.
[42] L. Cominardi *et al.*, "MEC Support for Network Slicing: Status and Limitations from a Standardization Viewpoint," *IEEE Communications Standards Magazine*, vol. 4, no. 2, pp. 22–30, 2020.
[43] A. Esmaeily and K. Kralevska, "Small-Scale 5G Testbeds for Network Slicing Deployment: A Systematic Review," *Wireless Communications and Mobile Computing*, vol. 2021, no. 6655216, pp. 1–26, 2021.
[44] L. Bonati *et al.*, "Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead," *Computer Networks*, vol. 182, p. 107516, 2020.
[45] D. Lake, N. Wang, R. Tafazolli, and L. Samuel, "Softwarization of 5G Networks–Implications to Open Platforms and Standardizations," *IEEE Access*, vol. 9, pp. 88 902–88 930, 2021.
[46] F. Tusa and S. Clayman, "End-to-end slices to orchestrate resources and services in the cloud-to-edge continuum," *Future Generation Computer Systems*, vol. 141, pp. 473–488, 2023.
[47] S. Zahoor *et al.*, "Comprehensive Analysis of Network Slicing for the Developing Commercial Needs and Networking Challenges," *Sensors*, vol. 22, no. 17, 2022.
[48] S. Wijethilaka and M. Liyanage, "Survey on Network Slicing for Internet of Things Realization in 5G Networks," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957–994, 2021.
[49] X. Xia *et al.*, "A Survey on 5G Network Slicing Enabling the Smart Grid," in *2019 IEEE 25th International Conference on Parallel and Distributed Systems*, 2019, pp. 911–916.
[50] 3GPP, "Management and orchestration; Study on management and orchestration of network slicing for next generation network (Release 15)," 3GPP, TR 28.801, Jan. 2018.
[51] L. Deng, H. Deng and A. Mayer, "Harmonizing Open Source and Standards: A Case for 5G Slicing," *ONAP*, 2020.
[52] 3GPP, "System architecture for the 5G System (5GS); Stage 2 (Release 18)," 3GPP, TS 23.501, Dec. 2023.
[53] O-RAN, "O-RAN Work Group 1 (Use Cases and Overall Architecture); O-RAN Architecture Description," O-RAN Alliance, TS O-RAN.WG1.OAD-R003-v11.00, Feb. 2024.
[54] 3GPP, "NR; NR and NG-RAN Overall description; Stage-2 (Release 18)," 3GPP, TS 38.300, Jan. 2024.
[55] ETSI, "Network Functions Virtualisation; An Introduction, Benefits, Enablers, Challenges & Call for Action," Oct. 2012.
[56] W. Xia *et al.*, "A Survey on Software-Defined Networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2015.
[57] ONF, "TR-526; Applying SDN Architecture to 5G Slicing," Apr. 2016.
[58] ETSI, "ETSI GR MEC 017 V1.1.1; Mobile Edge Computing (MEC): Deployment of mobile edge computing in an NFV environment," Feb. 2018.
[59] R. Ou, D. Ayepah–Mensah and G. Liu, "MEC-enabled Federated Learning for Network Slicing," in *2022 International Conference on Computing, Communication, Perception and Quantum Technology*, 2022, pp. 249–254.

[60] M. Dai *et al.*, "PSACCF: Prioritized Online Slice Admission Control Considering Fairness in 5G/B5G Networks," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 6, pp. 4101–4114, 2022.
[61] W. Attaoui, E. Sabir, H. Elbiaze, and M. Guizani, "VNF and CNF Placement in 5G: Recent Advances and Future Trends," *IEEE Transactions on Network and Service Management*, vol. 11, pp. 39 123–39 153, 2023.
[62] G. Wang *et al.*, "Reconfiguration in Network Slicing—Optimizing the Profit and Performance," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 591–605, 2019.
[63] F. Peres and M. Castelli, "Combinatorial Optimization Problems and Metaheuristics: Review, Challenges, Design, and Development," *Applied Sciences*, vol. 11, no. 14, 2021.
[64] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[65] S. Nayak, *Fundamentals of optimization techniques with algorithms*. Academic Press, 2020.
[66] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective Signal Processing Optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, 2014.
[67] N. Van Huynh, D. Thai Hoang, D. N. Nguyen, and E. Dutkiewicz, "Optimal and Fast Real-Time Resource Slicing With Deep Dueling Neural Networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1455–1470, 2019.
[68] D. Bertsekas, *Network optimization: continuous and discrete models*. Athena Scientific, 1998, vol. 8.
[69] L. Chinchilla-Romero *et al.*, "5G Infrastructure Network Slicing: E2E Mean Delay Model and Effectiveness Assessment to Reduce Downtimes in Industry 4.0," *Sensors*, vol. 22, no. 1, 2022.
[70] S. Ebrahimi, A. Zakeri, B. Akbari, and N. Mokari, "Joint Resource and Admission Management for Slice-enabled Networks," in *2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1–7.
[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 Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2079–2091, 2020.
[72] A. A. Khan *et al.*, "An End-to-End (E2E) Network Slicing Framework for 5G Vehicular Ad-Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 7103–7112, 2021.
[73] D. T. Hoang *et al.*, *Deep Reinforcement Learning for Wireless Communications and Networking: Theory, Applications and Implementation*. Wiley-IEEE Press, 2023.
[74] J. Wang, J. Liu, J. Li, and N. Kato, "Artificial Intelligence-Assisted Network Slicing: Network Assurance and Service Provisioning in 6G," *IEEE Vehicular Technology Magazine*, vol. 18, no. 1, pp. 49–58, 2023.
[75] N. Salhab *et al.*, "Autonomous Network Slicing Prototype Using Machine-Learning-Based Forecasting for Radio Resources," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 73–79, 2021.
[76] G. Dandachi, A. De Domenico, D. T. Hoang, and D. Niyato, "An Artificial Intelligence Framework for Slice Deployment and Orchestration in 5G Networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 858–871, 2020.
[77] D. Bega *et al.*, "Optimising 5G infrastructure markets: The business of network slicing," in *IEEE Conference on Computer Communications*, 2017, pp. 1–9.
[78] X. Liu, Y. Deng, A. Nallanathan, and M. Bennis, "Federated Learning and Meta Learning: Approaches, Applications, and Directions," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 571–618, 2024.
[79] F. Rezazadeh *et al.*, "On the Specialization of FDRL Agents for Scalable and Distributed 6G RAN Slicing Orchestration," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 3473–3487, 2023.
[80] Z. Wang, Y. Wei, F. R. Yu, and Z. Han, "Utility Optimization for Resource Allocation in Multi-Access Edge Network Slicing: A Twin-Actor Deep Deterministic Policy Gradient Approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 5842–5856, 2022.
[81] M. Jiang, M. Condoluci and T. Mahmoodi, "Network slicing in 5G: An auction-based model," in *2017 IEEE International Conference on Communications*, 2017, pp. 1–6.
[82] R. Ou *et al.*, "Two-Tier Resource Allocation for Multi-Tenant Network Slicing: A Federated Deep Reinforcement Learning Approach," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp. 20 174–20 187, 2023.
[83] J. J. Alves Esteves, A. Boubendir, F. Guillemin, and P. Sens, "A Heuristically Assisted Deep Reinforcement Learning Approach for Network Slice Placement," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4794–4806, 2022.
[84] R. Challa *et al.*, "Network Slice Admission Model: Tradeoff Between Monetization and Rejections," *IEEE Systems Journal*, vol. 14, no. 1, pp. 657–660, 2020.
[85] B. Han *et al.*, "Multiservice-Based Network Slicing Orchestration With Impatient Tenants," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 5010–5024, 2020.
[86] Y. Adou, E. Markova and Y. Gaidamaka, "Modeling and Analyzing Preemption-Based Service Prioritization in 5G Networks Slicing Framework," *Future Internet*, vol. 14, no. 10, 2022.
[87] S. Saxena and K. M. Sivalingam, "DRL-Based Slice Admission Using Overbooking in 5G Networks," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 29–45, 2023.
[88] M. Sulaiman *et al.*, "Coordinated Slicing and Admission Control using Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1110–1124, 2023.
[89] X. Wang and T. Zhang, "Reinforcement Learning Based Resource Allocation for Network Slicing in 5G C-RAN," in *2019 Computing, Communications and IoT Applications*, 2019, pp. 106–111.
[90] M. Karbalaee Motalleb, V. Shah-Mansouri, S. Parsaeefard, and O. L. Alcaraz López, "Resource Allocation in an Open RAN System Using Network Slicing," *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 471–485, 2023.
[91] J. Wang *et al.*, "Joint Bandwidth Allocation and Path Selection in WANs with Path Cardinality Constraints," *Journal of Communications and Information Networks*, vol. 6, no. 3, pp. 237–250, 2021.
[92] J. J. Alcaraz, F. Losilla, A. Zanella, and M. Zorzi, "Model-Based Reinforcement Learning With Kernels for Resource Allocation in RAN Slices," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 486–501, 2023.
[93] C. Qi *et al.*, "Deep Reinforcement Learning With Discrete Normalized Advantage Functions for Resource Management in Network Slicing," *IEEE Communications Letters*, vol. 23, no. 8, pp. 1337–1341, 2019.
[94] A. Nassar and Y. Yilmaz, "Deep Reinforcement Learning for Adaptive Network Slicing in 5G for Intelligent Vehicular Systems and Smart Cities," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 222–235, 2022.
[95] F. Mason, G. Nencioni and A. Zanella, "Using Distributed Reinforcement Learning for Resource Orchestration in a Network Slicing Scenario," *IEEE/ACM Transactions on Networking*, vol. 31, no. 1, pp. 88–102, 2023.
[96] S. Choi *et al.*, "Deep Reinforcement Learning for Scalable Dynamic Bandwidth Allocation in RAN Slicing with Highly Mobile Users," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 576–590, 2024.
[97] G. Zhou *et al.*, "Multi-objective Optimization of Space-Air-Ground Integrated Network Slicing Relying on a Pair of Central and Distributed Learning Algorithms," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8327–8344, 2024.
[98] Y. Cai *et al.*, "Deep Reinforcement Learning for Online Resource Allocation in Network Slicing," *IEEE Transactions on Mobile Computing*, 2023.
[99] T. Dong *et al.*, "Intelligent Joint Network Slicing and Routing via GCN-Powered Multi-Task Deep Reinforcement Learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 1269–1286, 2022.
[100] H. Chergui, L. Blanco and C. Verikoukis, "Statistical Federated Learning for Beyond 5G SLA-Constrained RAN Slicing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 2066–2076, 2022.
[101] S. Roy, F. Rezazadeh, H. Chergui, and C. Verikoukis, "Joint Explainability and Sensitivity-Aware Federated Deep Learning for Transparent 6G RAN Slicing," in *IEEE International Conference on Communications*, 2023, pp. 1238–1243.
[102] S. S., S. Mishra and C. Hota, "Joint QoS and energy-efficient resource allocation and scheduling in 5G Network Slicing," *Computer Communications*, vol. 202, pp. 110–123, 2023.
[103] J. J. A. Esteves, A. Boubendir, F. Guillemin, and P. Sens, "Location-based Data Model for Optimized Network Slice Placement," in *2020 6th IEEE Conference on Network Softwarization*, 2020, pp. 404–412.
[104] W.-K. Chen *et al.*, "Towards Efficient Large-Scale Network Slicing: An LP Dynamic Rounding-and-Refinement Approach," *IEEE Transactions on Signal Processing*, vol. 71, pp. 615–630, 2023.
[105] R. A. Addad *et al.*, "Optimization Model for Cross-Domain Network Slices in 5G Networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1156–1169, 2020.

[106] E. Amiri, N. Wang, M. Shojafar, and R. Tafazolli, "Edge-AI Empowered Dynamic VNF Splitting in O-RAN Slicing: A Federated DRL Approach," *IEEE Communications Letters*, vol. 28, no. 2, pp. 318–322, 2024.

[107] H. T. Nguyen, T. Van Do and C. Rotter, "Scaling UPF Instances in 5G/6G Core With Deep Reinforcement Learning," *IEEE Access*, vol. 9, pp. 165 892–165 906, 2021.

[108] R. Han *et al.*, "Multi-SP Network Slicing Parallel Relieving Edge Network Conflict," *IEEE Transactions on Parallel and Distributed Systems*, vol. 34, no. 11, pp. 2860–2875, 2023.

[109] M. Pozza *et al.*, "On Reconfiguring 5G Network Slices," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1542–1554, 2020.

[110] S. Yin *et al.*, "Prediction-Based End-to-End Dynamic Network Slicing in Hybrid Elastic Fiber-Wireless Networks," *Journal of Lightwave Technology*, vol. 39, no. 7, pp. 1889–1899, 2021.

[111] B. Khodapanah *et al.*, "Framework for Slice-Aware Radio Resource Management Utilizing Artificial Neural Networks," *IEEE Access*, vol. 8, pp. 174 972–174 987, 2020.

[112] W. Guan, H. Zhang and V. C. Leung, "Slice Reconfiguration based on Demand Prediction with Dueling Deep Reinforcement Learning," in *2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[113] K. Tokuda, T. Sato and E. Oki, "Network slice reconfiguration with deep reinforcement learning under variable number of service function chains," *Computer Networks*, vol. 224, p. 109636, 2023.

[114] S. S. Johari *et al.*, "DRL-Assisted Reoptimization of Network Slice Embedding on EON-enabled Transport Networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 800–814, 2023.

[115] F. Rezazadeh, H. Chergui, L. Alonso, and C. Verikoukis, "Continuous Multi-objective Zero-touch Network Slicing via Twin Delayed DDPG and OpenAI Gym," in *2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[116] C. Zhang *et al.*, "Architectural Design and Life Cycle Management of Network Slicing for Software-Defined Optical Access Networks," *IEEE Network*, vol. 36, no. 2, pp. 82–88, 2022.

[117] L. Contreras *et al.*, "iFUSION: Standards-based SDN Architecture for Carrier Transport Network," in *2019 IEEE Conference on Standards for Communications and Networking*, 2019, pp. 1–7.

[118] A. Zakeri *et al.*, "Digital Transformation VIA 5G: Deployment Plans," in *2020 ITU Kaleidoscope: Industry-Driven Digital Transformation*, 2020, pp. 1–8.

[119] D. Sattar and A. Matrawy, "Towards Secure Slicing: Using Slice Isolation to Mitigate DDoS Attacks on 5G Core Network Slices," in *2019 IEEE Conference on Communications and Network Security*, 2019, pp. 82–90.

[120] N. A. E. Kuadey *et al.*, "DeepSecure: Detection of Distributed Denial of Service Attacks on 5G Network Slicing—Deep Learning Approach," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 488–492, 2022.

[121] P. Wright *et al.*, "5G network slicing with QKD and quantum-safe security," *Journal of Optical Communications and Networking*, vol. 13, no. 3, pp. 33–40, 2021.

[122] J. S. B. Martins *et al.*, "Enhancing Network Slicing Architectures With Machine Learning, Security, Sustainability and Experimental Networks Integration," *IEEE Access*, vol. 11, pp. 69 144–69 163, 2023.

[123] Y. Zhang, A. Juels, A. Oprea, and M. K. Reiter, "HomeAlone: Co-residency Detection in the Cloud via Side-Channel Analysis," in *2011 IEEE Symposium on Security and Privacy*, 2011, pp. 313–328.

[124] Y. Li *et al.*, "Side Channel Attack-Aware Resource Allocation for URLLC and eMBB Slices in 5G RAN," *IEEE Access*, vol. 8, pp. 2090–2099, 2020.

[125] Y. Shi, Y. E. Sagduyu, T. Erpek, and M. C. Gursoy, "Jamming Attacks on NextG Radio Access Network Slicing with Reinforcement Learning," in *2022 IEEE Future Networks World Forum*, 2022, pp. 397–402.

[126] P. Alemany *et al.*, "Evaluation of the abstraction of optical topology models in blockchain-based data center interconnection," *Journal of Optical Communications and Networking*, vol. 14, no. 4, pp. 211–221, 2022.

[127] K. Noroozi, M. Karimzadeh-Farshbafan and V. Shah-Mansouri, "Service Admission Control for 5G Mobile Networks with RAN and Core Slicing," in *2019 IEEE Global Communications Conference*, 2019, pp. 1–6.

[128] D. Gligoroski and K. Kralevska, "Expanded Combinatorial Designs as Tool to Model Network Slicing in 5G," *IEEE Access*, vol. 7, pp. 54 879–54 887, 2019.

[129] Q.-T. Luu, S. Kerboeuf, A. Mouradian, and M. Kieffer, "A Coverage-Aware Resource Provisioning Method for Network Slicing," *IEEE/ACM Transactions on Networking*, vol. 28, no. 6, pp. 2393–2406, 2020.

[130] S. K. Taskou, M. Rasti and E. Hossain, "End-to-End Resource Slicing for Coexistence of eMBB and URLLC Services in 5G-Advanced/6G Networks," *IEEE Transactions on Mobile Computing*, pp. 1–18, 2023.

[131] B. Yu, X. Chi and X. Liu, "Martingale-Based Bandwidth Abstraction and Slice Instantiation Under the End-to-End Latency-Bounded Reliability Constraint," *IEEE Communications Letters*, vol. 26, no. 1, pp. 217–221, 2022.

[132] J. Khamse-Ashari, G. Senarath, I. Bor-Yaliniz, and H. Yanikomeroglu, "An Agile and Distributed Mechanism for Inter-Domain Network Slicing in Next Generation Mobile Networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 10, pp. 3486–3501, 2022.

[133] T. Li, X. Zhu and X. Liu, "An End-to-End Network Slicing Algorithm Based on Deep Q-Learning for 5G Network," *IEEE Access*, vol. 8, pp. 122 229–122 240, 2020.

[134] J. S. Pujol Roig, D. M. Gutierrez-Estevez and D. Gündüz, "Management and Orchestration of Virtual Network Functions via Deep Reinforcement Learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 304–317, 2020.

[135] A. Gharehgoli *et al.*, "AI-based Resource Allocation in End-to-End Network Slicing under Demand and CSI Uncertainties," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3630–3651, 2023.

[136] A. Farrel *et al.*, "A Framework for Network Slices in Networks Built from IETF Technologies," RFC 9543, IETF, Tech. Rep. 9543, Mar. 2024.

[137] R. A. Addad, D. L. C. Dutra, T. Taleb, and H. Flinck, "Toward Using Reinforcement Learning for Trigger Selection in Network Slice Mobility," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2241–2253, 2021.

[138] K. Abbas, T. A. Khan, M. Afaq, and W.-C. Song, "Network Slice Lifecycle Management for 5G Mobile Networks: An Intent-Based Networking Approach," *IEEE Access*, vol. 9, pp. 80 128–80 146, 2021.

[139] A. Clemm, L. Ciavaglia, L. Z. Granville, and J. Tantsura, "Intent-Based Networking - Concepts and Definitions," IETF RFC 9315, Oct. 2022.

[140] H. Yang *et al.*, "Data-Driven Network Slicing From Core to RAN for 5G Broadcasting Services," *IEEE Transactions on Broadcasting*, vol. 67, no. 1, pp. 23–32, 2021.

[141] W. Guan, H. Zhang and V. C. M. Leung, "Customized Slicing for 6G: Enforcing Artificial Intelligence on Resource Management," *IEEE Network*, vol. 35, no. 5, pp. 264–271, 2021.

[142] H. Bai, Y. Zhang, Z. Zhang, and S. Yuan, "Latency Equalization Policy of End-to-End Network Slicing Based on Reinforcement Learning," *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 88–103, 2022.

[143] W. Wu *et al.*, "AI-Native Network Slicing for 6G Networks," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 96–103, 2022.

[144] Q. Liu, T. Han and N. Ansari, "Learning-Assisted Secure End-to-End Network Slicing for Cyber-Physical Systems," *IEEE Network*, vol. 34, no. 3, pp. 37–43, 2020.

[145] S. R. A. Sebakara *et al.*, "SNAF: DRL-based Interdependent E2E Resource Slicing Scheme for a Virtualized Network," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 9069–9084, 2023.

[146] R. Mahindra, M. A. Khojastepour, H. Zhang, and S. Rangarajan, "Radio Access Network sharing in cellular networks," in *2013 21st IEEE International Conference on Network Protocols*, 2013, pp. 1–10.

[147] H. Yu *et al.*, "Isolation-Aware 5G RAN Slice Mapping Over WDM Metro-Aggregation Networks," *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1125–1137, 2020.

[148] Y. Wang *et al.*, "Deep Reinforcement Learning-Based Optimization for End-to-End Network Slicing With Control- and User-Plane Separation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 179–12 194, 2022.

[149] O-RAN, "O-RAN Work Group 1 (Use Cases and Overall Architecture); Slicing Architecture," O-RAN Alliance, TS O-RAN.WG1.Slicing-Architecture-R003-v12.00, Feb. 2024.

[150] N. Reyhanian and Z.-Q. Luo, "Data-Driven Adaptive Network Slicing for Multi-Tenant Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 113–128, 2022.

[151] H. Yu, T. Taleb, J. Zhang, and H. Wang, "Deterministic Latency Bounded Network Slice Deployment in IP-Over-WDM Based Metro-Aggregation Networks," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 2, pp. 596–607, 2022.

[152] A. Bhattacharya, H. Kumbhakar, M. Chatterjee, and D. Saha, "Controlling Traffic Hops - A New Approach for Traffic Grooming in WDM Optical Networks," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies*, 2020, pp. 1–5.

[153] Q. Xu, J. Wang and K. Wu, "Learning-Based Dynamic Resource Provisioning for Network Slicing with Ensured End-to-End Performance Bound," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 28–41, 2020.

[154] R. Wang, A.-H. Aghvami and V. Friderikos, "Service-Aware Design Policy of End-to-End Network Slicing for 5G Use Cases," *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 962–975, 2022.

[155] Y. Kim and H. Lim, "Multi-Agent Reinforcement Learning-Based Resource Management for End-to-End Network Slicing," *IEEE Access*, vol. 9, pp. 56 178–56 190, 2021.

[156] B. Xiang, J. Elias, F. Martignon, and E. Di Nitto, "Joint Network Slicing and Mobile Edge Computing in 5G Networks," in *2019 IEEE International Conference on Communications*, 2019, pp. 1–7.

[157] A. Rabitsch *et al.*, "Integrated Network and End-Host Policy Management for Network Slicing," in *18th IFIP International Conference on Network and Service Management*, 2023.

[158] S. Chen *et al.*, "5G UE Network Slicing based on SDK Implementation and Field Test," in *2022 3rd International Conference on Computing, Networks and Internet of Things*, 2022, pp. 151–157.

[159] P. Gorla, V. Chamola, V. Hassija, and D. Niyato, "Network Slicing for 5G with UE State Based Allocation and Blockchain Approach," *IEEE Network*, vol. 35, no. 3, pp. 184–190, 2021.

[160] S. Bakri, P. A. Frangoudis, A. Ksentini, and M. Bouaziz, "Data-Driven RAN Slicing Mechanisms for 5G and Beyond," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4654–4668, 2021.

[161] Y.-J. Liu *et al.*, "Device Association for RAN Slicing Based on Hybrid Federated Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 731–15 745, 2020.

[162] A. Papa *et al.*, "User-Based Quality of Service Aware Multi-Cell Radio Access Network Slicing," *IEEE Transactions on Network and Service Management*, vol. 19, no. 1, pp. 756–768, 2022.

[163] J. Goh and M. Sim, "Robust Optimization Made Easy with ROME," *Operations Research*, vol. 59, no. 4, pp. 973–985, 2011.

[164] S. O. Oladejo and O. E. Falowo, "Profit-Aware Resource Allocation for 5G Sliced Networks," in *2018 European Conference on Networks and Communications*, 2018, pp. 43–9.

[165] D. De Vleeschauwer, C. Papagianni and A. Walid, "Decomposing SLAs for Network Slicing," *IEEE Communications Letters*, vol. 25, no. 3, pp. 950–954, 2021.

[166] Q.-T. Luu, S. Kerboeuf and M. Kieffer, "Admission Control and Resource Reservation for Prioritized Slice Requests With Guaranteed SLA Under Uncertainties," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 3136–3153, 2022.

[167] W. Guan *et al.*, "A service-oriented deployment policy of end-to-end network slicing based on complex network theory," *IEEE Access*, vol. 6, pp. 19 691–19 701, 2018.

[168] W. Guan, H. Zhang and V. C. M. Leung, "Analysis of Traffic Performance on Network Slicing Using Complex Network Theory," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 188–15 199, 2020.

[169] H. Wang, Y. Wu, G. Min, and W. Miao, "A Graph Neural Network-Based Digital Twin for Network Slicing Management," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1367–1376, 2022.

[170] Q.-T. Luu, S. Kerboeuf and M. Kieffer, "Uncertainty-Aware Resource Provisioning for Network Slicing," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 79–93, 2021.

[171] M. Masoudi, O. T. Demir, J. Zander, and C. Cavdar, "Energy-Optimal End-to-End Network Slicing in Cloud-Based Architecture," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 574–592, 2022.

[172] Y. Sun *et al.*, "Service Provisioning Framework for RAN Slicing: User Admissibility, Slice Association and Bandwidth Allocation," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3409–3422, 2021.

[173] H. Chergui and C. Verikoukis, "Offline SLA-Constrained Deep Learning for 5G Networks Reliable and Dynamic End-to-End Slicing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 350–360, 2020.

[174] Q. Liu, N. Choi and T. Han, "OnSlicing: Online End-to-End Network Slicing with Reinforcement Learning," in *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 141–153.

[175] Y. Xie *et al.*, "Resource allocation for network slicing in dynamic multi-tenant networks: A deep reinforcement learning approach," *Computer Communications*, vol. 195, pp. 476–487, 2022.

[176] I. Kovacevic *et al.*, "Multi-Domain Network Slicing With Latency Equalization," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2182–2196, 2020.

[177] Q. Liu, N. Choi and T. Han, "Deep Reinforcement Learning for End-to-End Network Slicing: Challenges and Solutions," *IEEE Network*, vol. 37, no. 2, pp. 222–228, 2023.

[178] J. Ordonez-Lucena, C. Tranoris, J. Rodrigues, and L. M. Contreras, "Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility," in *2020 European Conference on Networks and Communications*, 2020, pp. 40–45.

[179] ETSI, "ETSI GS NFV 002 V1.2.1; Network Functions Virtualisation (NFV); Architectural Framework," Dec. 2014.

[180] TM Forum, "GB992 Open API Map R18.0.1," Sep. 2018.

[181] Metro Ethernet Forum (MEF), "MEF 55; Lifecycle Service Orchestration (LSO): Reference Architecture and Framework," Mar. 2016.

[182] E. Zeydan *et al.*, "Blockchain-Based Service Orchestration for 5G Vertical Industries in Multicloud Environment," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4888–4904, 2022.

[183] S. Ben Saad, A. Ksentini and B. Brik, "An end-to-end trusted architecture for network slicing in 5G and beyond networks," *SECURITY AND PRIVACY*, vol. 5, no. 1, p. e186, 2021.

[184] G. He *et al.*, "NetChain: A Blockchain-Enabled Privacy-Preserving Multi-Domain Network Slice Orchestration Architecture," *IEEE Transactions on Network and Service Management*, vol. 19, no. 1, pp. 188–202, 2022.

[185] N. Torkzaban, C. Papagianni and J. S. Baras, "Trust-Aware Service Chain Embedding," in *2019 Sixth International Conference on Software Defined Systems*, 2019, pp. 242–247.

[186] S. K. Taskou, M. Rasti and P. H. J. Nardelli, "Minimizing Energy Consumption for End-to-End Slicing in 5G Wireless Networks and Beyond," in *2022 IEEE Wireless Communications and Networking Conference*, 2022, pp. 860–865.

[187] W. Whitt, "The queueing network analyzer," *The Bell System Technical Journal*, vol. 62, no. 9, pp. 2779–2815, 1983.

[188] *Overview of the security and trust mechanisms in the 5GZORRO project*. Zenodo, Jun. 2022.

[189] V. Theodorou *et al.*, "Blockchain-based Zero Touch Service Assurance in Cross-domain Network Slicing," in *2021 Joint European Conference on Networks and Communications & 6G Summit*, 2021, pp. 395–400.

[190] D. Breitgand *et al.*, "Dynamic Slice Scaling Mechanisms for 5G Multi-domain Environments," in *2021 IEEE 7th International Conference on Network Softwarization (NetSoft)*, 2021, pp. 56–62.

[191] F. Rezazadeh, H. Chergui and J. Mangues-Bafalluy, "Explanation-Guided Deep Reinforcement Learning for Trustworthy 6G RAN Slicing," in *2023 IEEE International Conference on Communications Workshops*, 2023, pp. 1026–1031.

[192] S. Kukliński *et al.*, "MonB5G: AI/ML-Capable Distributed Orchestration and Management Framework for Network Slices," in *2021 IEEE International Mediterranean Conference on Communications and Networking*, 2021, pp. 29–34.

[193] Z. Shu and T. Taleb, "A Novel QoS Framework for Network Slicing in 5G and Beyond Networks Based on SDN and NFV," *IEEE Network*, vol. 34, no. 3, pp. 256–263, 2020.

[194] H. Chergui *et al.*, "Zero-Touch AI-Driven Distributed Management for Energy-Efficient 6G Massive Network Slicing," *IEEE Network*, vol. 35, no. 6, pp. 43–49, 2021.

[195] R. Vilalta *et al.*, "TeraFlow: Secured Autonomic Traffic Management for a Tera of SDN flows," in *2021 Joint European Conference on Networks and Communications & 6G Summit*, 2021, pp. 377–382.

[196] P. Alemany *et al.*, "Blockchain-Based Connectivity Provisioning in Multiple Transport SDN Domains," in *2021 International Conference on Optical Network Design and Modeling*, 2021, pp. 1–3.

[197] S. Barguil *et al.*, "Packet Optical Transport Network Slicing with Hard and Soft Isolation," *Applied Sciences*, vol. 11, no. 13, 2021.

[198] A. Siokis *et al.*, "5G experimentation environment for third party media services: the 5GMediaHUB project," in *2021 IEEE International Mediterranean Conference on Communications and Networking*, 2021, pp. 1–6.

[199] B. Ojaghi, F. Adelantado, A. Antonopoulos, and C. Verikoukis, "Impact of Network Densification on Joint Slicing and Functional Splitting in 5G," *IEEE Communications Magazine*, vol. 60, no. 7, pp. 30–35, 2022.

[200] A. Dalgkitsis, A. Chawla, A.-M. Bosneag, and C. Verikoukis, "SafeSCHEMA: Multi-domain Orchestration of Slices based on SafeRL for B5G Networks," in *2022 IEEE Global Communications Conference*, 2022, pp. 3435–3440.
[201] A. M. Escolar *et al.*, "Adaptive Network Slicing in Multi-Tenant 5G IoT Networks," *IEEE Access*, vol. 9, pp. 14 048–14 069, 2021.
[202] E. Chirivella-Perez *et al.*, "Intent-Based E2E Network Slice Management for Industry 4.0," in *2021 Joint European Conference on Networks and Communications & 6G Summit*, 2021, pp. 353–358.
[203] J. Fonseca *et al.*, "6G BRAINS Topology-aware Industry-grade Network Slice Management and Orchestration," in *2023 Joint European Conference on Networks and Communications & 6G Summit*, 2023, pp. 418–423.
[204] M. Wang *et al.*, "SDN East–West cooperation in a converged fixed-mobile optical access network: enabling 5G slicing capabilities," *Journal of Optical Communications and Networking*, vol. 14, no. 7, pp. 540–549, 2022.
[205] V. Charpentier *et al.*, "Dynamic and Quality-aware Network Slice Management in 5G Testbeds," in *2023 Joint European Conference on Networks and Communications & 6G Summit*, 2023, pp. 611–616.
[206] M. Iordache *et al.*, "Advanced 5G Open Testbed for Network Applications Experiments," in *2023 Joint European Conference on Networks and Communications & 6G Summit*, 2023, pp. 526–531.
[207] E. Kosmatos *et al.*, "Real-time orchestration of QoS-aware end-to-end slices across a converged Metro and Access network exploiting burst-mode technology," *Journal of Optical Communications and Networking*, vol. 15, no. 1, pp. 1–15, 2023.
[208] S. Roy, H. Chergui, L. Sanabria-Russo, and C. Verikoukis, "A Cloud Native SLA-Driven Stochastic Federated Learning Policy for 6G Zero-Touch Network Slicing," in *IEEE International Conference on Communications*, 2022, pp. 4269–4274.
[209] R. Bolla *et al.*, "Machine-Learning-Based 5G Network Function Scaling via Black- and White-Box KPIs," in *2023 21st Mediterranean Communication and Computer Networking Conference*, 2023, pp. 143–150.
[210] D. Masouros *et al.*, "Towards privacy-first security enablers for 6G networks: the PRIVATEER approach," in *2023 International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation (SAMOS XXIII)*. Springer, 2023.
[211] 6G-XR, "6G eXperimental Research," Mar. 2024. [Online]. Available: https://www.6g-xr.eu
[212] N. Apostolakis *et al.*, "Digital Twins for Next-Generation Mobile Networks: Applications and Solutions," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 80–86, 2023.
[213] I. P. Chochliouros *et al.*, "6G-BRICKS: Developing a Modern Experimentation Facility for Validation, Testing and Showcasing of 6G Breakthrough Technologies and Devices," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2023, pp. 17–31.
[214] S. Arora, A. Ksentini and C. Bonnet, "Cloud native Lightweight Slice Orchestration (CLiSO) framework," *Computer Communications*, vol. 213, pp. 1–12, 2024.
[215] D. Giannopoulos *et al.*, "ACROSS: Automated zero-touch cross-layer provisioning framework for 5G and beyond vertical services," in *2023 Joint European Conference on Networks and Communications & 6G Summit*, 2023, pp. 735–740.
[216] A. Mozo *et al.*, "A Machine-Learning-Based Cyberattack Detector for a Cloud-Based SDN Controller," *Applied Sciences*, vol. 13, no. 8, 2023.
[217] J. Breen *et al.*, "POWDER: Platform for Open Wireless Data-Driven Experimental Research," in *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH'20. New York, NY, USA: Association for Computing Machinery, 2020, p. 17–24.
[218] D. Johnson, D. Maas and J. Van Der Merwe, "NexRAN: Closed-Loop RAN Slicing in POWDER -A Top-to-Bottom Open-Source Open-RAN Use Case," in *15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization*, ser. WiNTECH'21. New York, NY, USA: Association for Computing Machinery, 2021, p. 17–23.
[219] L. Bonati *et al.*, "OpenRAN Gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms," *Computer Networks*, vol. 220, p. 109502, 2023.
[220] M. Polese *et al.*, "ColO-RAN: Developing Machine Learning-based xApps for Open RAN Closed-loop Control on Programmable Experimental Platforms," *IEEE Transactions on Mobile Computing*, vol. 22, no. 10, pp. 5787–5800, 2023.
[221] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network Automation through Orchestrated Intelligence in the Open RAN," in *IEEE Conference on Computer Communications*, 2022, p. 270–279.
[222] M. Hoffmann *et al.*, "A Secure and Resilient 6G Architecture Vision of the German Flagship Project 6G-ANNA," *IEEE Access*, vol. 11, pp. 102 643–102 660, 2023.
[223] A. Nota *et al.*, "Context-based Latency Guarantees Considering Channel Degradation in 5G Network Slicing," in *2022 IEEE Real-Time Systems Symposium*, 2022, pp. 253–265.
[224] D. Overbeck, N. A. Wagner, F. Kurtz, and C. Wietfeld, "Proactive Resource Management for Predictive 5G Uplink Slicing," in *2022 IEEE Global Communications Conference*, 2022, pp. 1000–1005.
[225] C. Bektas, S. Böcker and C. Wietfeld, "The Cost of Uncertainty: Impact of Overprovisioning on the Dimensioning of Machine Learning-based Network Slicing," in *2022 IEEE Future Networks World Forum*, 2022, pp. 652–657.
[226] D. Overbeck, F. Kurtz, S. Böcker, and C. Wietfeld, "Design of a 5G Network Slicing Architecture for Mixed-Critical Services in Cellular Energy Systems," in *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, 2022, pp. 90–95.
[227] F. Kurtz, R. Wiebusch, D. Overbeck, and C. Wietfeld, "Predictive 5G Uplink Slicing for Blockchain-driven Smart Energy Contracts," in *2022 IEEE International Conference on Communications Workshops*, 2022, pp. 19–24.
[228] S. Barrachina-Muñoz *et al.*, "Deploying cloud-native experimental platforms for zero-touch management 5G and beyond networks," *IET Networks*, 2023.
[229] O. Adamuz Hinojosa, P. Ameigeiras, P. Muñoz, and J. Lopez, "UE Blocking Probability Model for Planning 5G Guaranteed Bit Rate (GBR) RAN Slices," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 124–12 138, 2023.
[230] O. Adamuz-Hinojosa, P. Muñoz, P. Ameigeiras, and J. M. Lopez-Soler, "Potential-Game-Based 5G RAN Slice Planning for GBR Services," *IEEE Access*, vol. 11, pp. 4763–4780, 2023.
[231] Toshiba, "Toshiba-led BEACON-5G celebrates innovations in Open RAN for 5G Diversification," Sep. 2023. [Online]. Available: www.beacon-5g.com/news
[232] E. S. Xavier, N. Agoulmine and J. S. B. Martins, "On Modeling Network Slicing Communication Resources with SARSA Optimization," in *ADVANCE'2023 - International Workshop on ADVANCEs in ICT Infrastructures and Services*, 2023, pp. 1–8.
[233] J. F. Santos *et al.*, "Optimal Embedding of Heterogeneous RAN Slices for Secure and Technology-Agnostic RANaaS," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 5284–5298, 2022.
[234] S. Clayman *et al.*, "The NECOS Approach to End-to-End Cloud-Network Slicing as a Service," *IEEE Communications Magazine*, vol. 59, no. 3, pp. 91–97, 2021.
[235] 5G PPP, "5G PPP Phase 3 Projects," Mar. 2024. [Online]. Available: https://www.5g-ppp.eu/5g-ppp-phase-3-projects
[236] 6G SNS JU, "SNS Phase 1," Mar. 2024. [Online]. Available: https://smart-networks.europa.eu/sns-phase-1/
[237] B. Khorsandi *et al.*, "The 6G Architecture Landscape - European perspective, Version 6.0," Feb. 2023.
[238] PAWR, "The Platforms for Advanced Wireless Research Program," Mar. 2024. [Online]. Available: www.advancedwireless.org
[239] ETSI, "Open Source MANO (OSM)," Mar. 2024. [Online]. Available: https://osm.etsi.org/
[240] C.-N. Nhu and M. Park, "Dynamic Network Slice Scaling Assisted by Attention-Based Prediction in 5G Core Network," *IEEE Access*, vol. 10, pp. 72 955–72 972, 2022.
[241] A. S. Muqaddas *et al.*, "NFV orchestration over disaggregated metro optical networks with end-to-end multi-layer slicing enabling crowd-sourced live video streaming," *Journal of Optical Communications and Networking*, vol. 13, no. 8, pp. D68–D79, 2021.
[242] S. Bhattacharjee *et al.*, "Network Slicing for TSN-Based Transport Networks," *IEEE Access*, vol. 9, pp. 62 788–62 809, 2021.
[243] M. Kist *et al.*, "AIRTIME: End-to-End Virtualization Layer for RAN-as-a-Service in Future Multi-Service Mobile Networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 8, pp. 2701–2717, 2022.
[244] X. Li *et al.*, "End-to-End Network Slicing in Radio Access Network, Transport Network and Core Network Domains," *IEEE Access*, vol. 8, pp. 29 525–29 537, 2020.
[245] Linux Foundation, "Open Network Automation Platform," Mar. 2024. [Online]. Available: www.onap.org

[246] V. Q. Rodriguez, F. Guillemin and A. Boubendir, "5G E2E Network Slicing Management with ONAP," in *2020 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops*, 2020, pp. 87–94.

[247] The OpenInfra Project, "OpenStack," Mar. 2024. [Online]. Available: www.openstack.org

[248] J. F. Santos *et al.*, "Breaking Down Network Slicing: Hierarchical Orchestration of End-to-End Networks," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 16–22, 2020.

[249] E. H. Bouzidi *et al.*, "Online based learning for predictive end-to-end network slicing in 5G networks," in *2020 IEEE International Conference on Communications*, 2020, pp. 1–7.

[250] N. Nikaein *et al.*, "OpenAirInterface: A Flexible Platform for 5G Research," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, p. 33–38, Oct. 2014.

[251] Open5GS, "Open5GS Project," Mar. 2024. [Online]. Available: www.open5gs.org

[252] ONF, "Open Network Operating System (ONOS)," Mar. 2024. [Online]. Available: www.opennetworking.org/onos

[253] B. Shariati *et al.*, "Demonstration of latency-aware 5G network slicing on optical metro networks," *Journal of Optical Communications and Networking*, vol. 14, no. 1, pp. A81–A90, 2022.

[254] Linux Foundation, "OpenDaylight," Mar. 2024. [Online]. Available: www.opendaylight.org

[255] Ryu, "Ryu SDN Framework," Mar. 2024. [Online]. Available: www.ryu-sdn.org

[256] Linux Foundation, "Open vSwitch," Mar. 2024. [Online]. Available: www.openvswitch.org

[257] N. McKeown *et al.*, "OpenFlow: Enabling Innovation in Campus Networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, p. 69–74, mar 2008.

[258] X. Foukas *et al.*, "FlexRAN: A Flexible and Programmable Platform for Software-Defined Radio Access Networks," in *Proceedings of the 12th International on Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 427–441.

[259] A. Gungor, "UERANSIM Project: An Open-Source State-of-the-Art 5G UE and RAN (gNodeB) Simulator," Mar. 2024. [Online]. Available: www.github.com/aligungr/UERANSIM

[260] M. S. Khan *et al.*, "SliceSecure: Impact and Detection of DoS/DDoS Attacks on 5G Network Slices," in *2022 IEEE Future Networks World Forum*, 2022, pp. 639–642.

[261] V. KumarSkandPriya, A. Dandoush and G. Diaz, "Slicenet: a Simple and Scalable Flow-Level Simulator for Network Slice Provisioning and Management," in *2023 International Symposium on Networks, Computers and Communications*, 2023, pp. 1–6.

[262] K. Katsalis, N. Nikaein and A. Huang, "JOX: An event-driven orchestrator for 5G network slicing," in *2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1–9.

[263] L. Bertizzolo *et al.*, "Arena: A 64-antenna SDR-based ceiling grid testing platform for sub-6 GHz 5G-and-Beyond radio spectrum research," *Computer Networks*, vol. 181, p. 107436, 2020.

[264] M. Polese *et al.*, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[265] J. Thaliath *et al.*, "Predictive Closed-Loop Service Automation in O-RAN Based Network Slicing," *IEEE Communications Standards Magazine*, vol. 6, no. 3, pp. 8–14, 2022.

[266] M. Yaqoob, R. Trestian, M. Tatipamula, and H. X. Nguyen, "Digital Twin-driven End-to-End Network Slicing Towards 6G," *IEEE Internet Computing*, pp. 1–6, 2024.

[267] ETSI, "ETSI GS ZSM 003 V1.1.1; Zero-touch network and Service Management (ZSM); End-to-end management and orchestration of network slicing," Jun. 2021.

[268] 3GPP, "5G System; Network Data Analytics Services; Stage 3 (Release 18)," 3GPP, TS 29.520, Dec. 2023.

[269] S. Bubeck, V. Chandrasekaran, H. Palangi, and Y. Zhang, "Sparks of Artificial General Intelligence: Early experiments with GPT-4," 2023.

[270] C. Wang *et al.*, "Making Network Configuration Human Friendly," 2023.

[271] S. Li *et al.*, "Robust Multi-Agent Reinforcement Learning via Minimax Deep Deterministic Policy Gradient," in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. AAAI Press, 2019.

[272] B. Zheng *et al.*, "Imitation Learning: Progress, Taxonomies and Challenges," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–16, 2022.

[273] H. Zhou, M. Erol-Kantarci and H. V. Poor, "Knowledge Transfer and Reuse: A Case Study of AI-Enabled Resource Management in RAN Slicing," *IEEE Wireless Communications*, vol. 30, no. 5, pp. 160–169, 2023.

[274] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer Learning in Deep Reinforcement Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 344–13 362, 2023.

[275] Y. Sun, M. Peng, S. Mao, and S. Yan, "Hierarchical Radio Resource Allocation for Network Slicing in Fog Radio Access Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3866–3881, 2019.

[276] E. Bjornson, "Reproducible Research: Best Practices and Potential Misuse [Perspectives]," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 106–123, 2019.

Sina Ebrahimi received the B.Sc. and M.Sc. degrees in software engineering from the University of Zanjan, Iran, and Tarbiat Modares University, Tehran, Iran, in 2016 and 2020, respectively. He is currently pursuing his Ph.D. degree at the Centre for Future Transport and Cities, Coventry University, Coventry, UK. His research interests include network slicing, 5G/6G enabling technologies (e.g., NFV, SDN, and MEC), resource allocation, and telco architecture.

Faouzi Bouali received a Ph.D. in Signal Theory and Communications from the University Politecnica de Catalunya (UPC), Spain in 2013. He is an Assistant Professor (Senior Lecturer) with the Centre for Future Transport and Cities, Coventry University, UK. His research interests span the field of communications and networking, with a focus on radio resource management, architectural innovations for next-generation networks (e.g., network slicing and Open RAN), 5G/6G experimental research and testbeds, and connected and automated industries.

Olivier C. L. Haas (Senior Member, IEEE, 2022) received a DUT (Diplôme Universitaire de Technologie) from the Université Joseph Fourier, Grenoble, France, B.Eng., M.Sc., and Ph.D. from Coventry University, Coventry, UK. He is an Associate Professor with 25 years' experience, 28 PhD completions. He heads the Intelligent Transportation Systems and 5G, Centre for Future Transport and Cities, Coventry University. pureportal.coventry.ac.uk/en/persons/olivier-haas