A Comprehensive Tutorial and Survey of O-RAN: Exploring Slicing-Aware Architecture, Deployment Options, Use Cases, and Challenges

Khurshid Alam, Mohammad Asif Habibi, Matthias Tammen, Dennis Krummacker,
Walid Saad, *Fellow, IEEE*, Marco Di Renzo, *Fellow, IEEE*, Tommaso Melodia, *Fellow, IEEE*,
Xavier Costa-Pérez, *Senior Member, IEEE*, Mérouane Debbah, *Fellow, IEEE*,
Ashutosh Dutta, *Fellow, IEEE*, and Hans D. Schotten, *Member, IEEE*

Abstract—Open-radio access network (O-RAN) seeks to establish the principles of openness, programmability, automation, intelligence, and hardware-software disaggregation with interoperable and standard-compliant interfaces. It advocates for multi-vendorism and multi-stakeholderism within a cloudified and virtualized wireless infrastructure, aimed at enhancing the deployment, operation, and management of RAN architecture. These enhancements promise increased flexibility, performance optimization, service innovation, energy efficiency, and cost effectiveness across fifth-generation (5G), sixth-generation (6G), and beyond networks. A silent feature of O-RAN architecture is its support for network slicing, which entails interaction with other domains of the cellular network, notably the transport network (TN) and the core network (CN), to realize end-to-end (E2E) network slicing. The study of this feature requires exploring the stances and contributions of diverse standards development organizations (SDOs). In this context, we note that despite the ongoing industrial deployments and standardization efforts, the research and standardization communities have yet to comprehensively address network slicing in O-RAN. To address this gap, this paper provides a comprehensive exploration of network slicing in O-RAN through an in-depth review of specification documents from O-RAN Alliance and research papers from leading industry and academic institutions. The paper commences with an overview of the relevant standardization and open source contributions, subsequently delving into the latest O-RAN architecture with an emphasis on its slicing aspects. Furthermore, the paper explores O-RAN deployment scenarios, examining options for the deployment and orchestration of RAN and TN slice subnets. It also discusses the slicing of the underlying infrastructure and provides an overview of various use cases related to O-RAN slicing. Finally, it summarizes the potential research challenges identified throughout the study.

Index Terms—3GPP, 5G, 6G, disaggregation, ETSI, management and orchestration, intelligence, Near-RT RIC, network slicing, NFV-MANO, Non-RT RIC, O-cloud site, O-RAN, O-RAN alliance, open interfaces, openness, RAN architecture, RAN slicing, RIC, SMO framework, standards, TN slicing.

I. INTRODUCTORY REMARKS

THE RADIO access network (RAN) is a critical domain of cellular network, providing wireless connectivity between user equipment (UE) and base stations across a specified geographical footprint. It employs various radio access technologies (RATs) to ensure efficient bidirectional data transmission [1]. Its architecture has evolved with increasing density of UEs, diverse access technologies, performance demands (e.g., higher data rates, lower latency), and emerging trends such as virtualization and cloudification [2], [3], [4], [5]. This evolution—from fourth-generation (4G) to fifth-generation (5G) and now towards sixth-generation (6G)—has enabled a wide range of advanced services, applications, and use cases [6], [7], [8], [9].

Introduced in 4G, the distributed RAN (D-RAN) architecture separated radio and baseband functions into distinct components—the remote radio head (RRH) and baseband unit (BBU)—though both components remained co-located at the same site [2], [10], [11]. To enhance the scalability and efficiency, centralized RAN (C-RAN) decoupled the BBU from the RRH and relocating it to a centralized data center (DC) [1]. This architectural shift enabled centralized control and optimization of multiple RRHs via high-speed fronthaul (FH) interface, commonly using common public radio interface (CPRI) [10], [11], [12]. The transition from D-RAN to C-RAN marked a key milestone in the evolution of modern RAN architectures, paving the way for innovations in 4G, 5G, and beyond [1], [13].

The advent of 5G heralded a paradigm shift in RAN to address the diverse demands of emerging industrial applications and vertical markets [2], [14], [15]. To support this evolution, 3rd Generation Partnership Project (3GPP) introduced the next-generation RAN (NG-RAN), where base stations—next generation NodeBs (gNBs)—are functionally split into the centralized unit (CU) for higher layer functions and the distributed unit (DU) for lower layer processing [16], [17], [18], [19].

A defining feature of NG-RAN is its cloud-native and virtualized design, allowing the deployment of CU and DU on commercial off-the-shelf (COTS) hardware within virtualized infrastructures [19], [20]. However, NG-RAN remains

largely closed and proprietary, limiting opportunities for innovation [21]. Its components often built with tightly coupled interfaces optimized for performance, tailored to specific manufacturers, thus restricting multi-vendor interoperability [22].

To overcome the limitations of proprietary NG-RAN systems and to promote openness and interoperability, the Open RAN (O-RAN) Alliance introduced the O-RAN architecture—a transformative initiative aimed to redefine traditional RAN design principles [23], [24], [25]. Building upon the 3GPP-defined gNB, the O-RAN introduces further functional disaggregation by separating the DU into two distinct entities: the O-RAN DU (O-DU) and the O-RAN RU (O-RU). As a result, the O-RAN gNodeB (O-gNB) consists of three interoperable components—O-RAN CU (O-CU), O-DU, and O-RU—supporting a scalable and flexible architecture where a single O-CU can control multiple O-DUs, and each O-DU can connect to several O-RUs. The O-RU handles the transmission and reception of radio signals to and from UEs [20], [25], [26].

This architectural shift replaces closed and proprietary RAN solutions with open, cloud-native, interoperable, and intelligent systems. By standardizing open interfaces and specifications, O-RAN fosters multi-vendor interoperability, innovation, and flexible deployments across diverse network environments [27]. A core enabler of this vision is the integration of artificial intelligence (AI) and machine learning (ML) via RAN intelligent controllers (RICs), which enhances automation, real-time optimization, and service assurance [20], [28], [29], [30]. Achieving an open and intelligent NG-RAN necessitates the adoption of software-defined networking (SDN) and network function virtualization (NFV) to decouple control and user planes, virtualize RAN components, and implement standardized open interfaces between them [31], [32]. In addition, the O-RAN architecture underpins advanced use cases such as network slicing, dynamic spectrum sharing, and RAN resource orchestration—key for delivering next-generation services with greater flexibility, efficiency, and scalability [22], [33], [34], [35].

Network slicing enables the partitioning of a physical network into multiple virtual slices, each operating independently and configured for specific application requirements, thereby supporting differentiated quality of service (QoS) across verticals [27], [33], [36], [37]. In contrast to the rigid, *one size fits all* architectural solutions, slicing introduces a flexible, and dynamic approach for resource allocation and network optimization, aligning with the needs of emerging 5G, 6G, and beyond services [38], [39], [40], [41]. In addition, network slicing empowers mobile network operators (MNOs) to fulfill service level agreements (SLAs) with tenants [42] by addressing diverse performance and functional requirements using standardized service types as defined by 3GPP. These service types, as of this writing, encompass enhanced mobile broadband (eMBB), ultra reliable low latency communication (URLLC), massive machine type communication (mMTC), high-performance machine type communication (HMTC), and vehicle to everything (V2X) [19]. With the evolution of 5G, 6G and beyond, this set is expected to grow significantly to accommodate the expanding communication and non-communication services [6].

While 3GPP provides foundational support for RAN slicing, the O-RAN Alliance plays a complementary and critical role in enabling intelligent slicing through open interfaces and components [23], [43]. To unlock the full potential of slicing within O-RAN framework, a strong unification between O-RAN and 3GPP standards is essential. Beyond traditional RAN slicing, O-RAN introduces several key enhancements that improve customization, interoperability and automation: **(a)** Through its disaggregated architecture, O-RAN enables fine-grained control over slice specific resources. Operators can select specialized hardware and software from different vendors to optimize the performance per slice. **(b)** O-RAN's open interfaces eliminates vendor lock-in, making it possible to mix and match interoperable components. This allows operators to choose the best-in-breed solutions for each slice, leading to cost savings and innovation. **(c)** The integration of non-real-time RIC (Non-RT RIC) and near-real-time RIC (Near-RT RIC) introduces intelligent control via Non-RT RIC applications (rApps) and Near-RT RIC applications (xApps), which enable close-loop automation and optimization of slice-specific resource management [44], [45], [46]. **(d)** O-RAN further supports automation of slice lifecycle management–encompassing slice creation, configuration, and orchestration–through its interfaces and protocols, significantly reducing operational complexity and improving reliability.

Note that, while both O-RAN and 3GPP address key aspects of RAN slicing, neither framework explicitly focuses on the virtualization aspects. Thus, a broader harmonization—encompassing O-RAN, 3GPP, and European Telecommunications Standards Institute (ETSI)—is required to support both the physical and virtual infrastructure. Such alignment is anticipated to deliver unified and scalable solutions for the realization of flexible, service-specific O-RAN slices [47].

A. Literature Review and Research Gap Analysis

The research community has made significant contributions to diverse research challenges associated with O-RAN. These efforts span multiple domains, encompassing the management & orchestration (M&O) of O-RAN slicing, the design of intelligent applications such as xApps and rApps, and the resolution of numerous optimization problems. In addition, several comprehensive survey and overview papers have been published by prominent academic institutions. Table I presents a curated and up-to-date list of these works, highlighting their major contributions to the broader O-RAN landscape, and where applicable, to the specific domain of O-RAN slicing.

While several of the works referenced in Table I offer valuable insights into the architectural components of O-RAN and explore aspects of network slicing, **they fall short in delivering a comprehensive analysis of slicing mechanisms tailored to the O-RAN framework**. In particular, **these surveys lack a detailed examination of the current landscape of O-RAN slicing**, including the exploration of various deployment strategies proposed in the literature and the complexities involved in tight integration of O-RAN components necessary to support intelligent and flexible slicing.

TABLE I
COMPARING THE CONTRIBUTIONS OF OUR PAPER TO THE MOST RECENT STATE-OF-THE-ART OVERVIEW AND SURVEY PAPERS RELATED TO O-RAN

Ref.	Year	Network Architecture	Deployment Scenarios	Open Source Initiatives	Network Slicing	Management & Orchestration	Use Cases & Examples
[18]	2020	✓		✓			
[48]	2022	✓					
[31]	2022	✓					
[49]	2022	✓				✓	✓
[50]	2022	✓					
[22]	2022	✓					
[21]	2023	✓	✓	✓			✓
[51]	2023	✓					
[52]	2023	✓					
[53]	2023	✓					
[27]	2023	✓					
[12]	2023	✓	✓				
[5]	2024	✓		✓	✓		
[54]	2024	✓	✓	✓			
[55]	2024	✓					✓
[56]	2024	✓				✓	✓
[57]	2024	✓					✓
This paper	2025	✓	✓	✓	✓	✓	✓

Note: For each column, a ✓ indicates that the aspect is discussed in detail, while a blank space signifies that the aspect is absent.

Moreover, **there is a significant gap in the existing literature regarding the seamless integration and interoperability of O-RAN's disaggregated components and open interfaces**, a crucial factor for the successful development and deployment of diverse O-RAN slice types. The absence of thorough discussion on these vital aspects hampers a complete understanding of how O-RAN can unlock the full potential of slicing across 5G, 6G, and beyond. This is particularly important when considering the challenges of achieving **interoperability and scalability in multi-vendor environments**, where cohesive integration across different components is essential for ensuring efficient network performance.

Therefore, **the lack of a detailed and cohesive analysis of the entire O-RAN and underlying infrastructure slicing**, from deployment scenarios to seamless integration of disaggregated components, reveals a critical gap in the literature. This underscores the need for further research to bridge these gaps and advance the development of optimized, interoperable network slices within the O-RAN framework.

B. Goals and Contributions

To address the identified gap in the literature, this paper provides a comprehensive exploration of slicing-aware O-RAN architecture. It offers a detailed tutorial and critical overview of network slicing within the context of the O-RAN architecture. The key contributions of this paper are as follows:

- An exploration of **open source initiatives, standardization efforts, and the design of experimental platforms** supporting the development and validation of O-RAN.

- A focused analysis of **the latest O-RAN architecture** as defined by the O-RAN Alliance, with particular emphasis on both **theoretical advancements from academic research and practical deployment by industry leaders**.
- A comprehensive and **holistic view of the network slicing paradigm within O-RAN**, including **the functional components and open interfaces essential for implementing slicing capabilities** across O-RAN.
- An exploration of key **deployment scenarios for O-RAN slicing and the service management and orchestration (SMO) framework**, alongside **detailed insight of several high-level use cases** that O-RAN is expected to support.
- An in-depth discussion of **the network functions (NFs) and transport network (TN) elements, along with their M&O**—particularly FH and midhaul (MH)—that collectively constitute the O-RAN slice subnet. Additionally, this paper examines **the aspects of the underlying infrastructure associated with the TN** and the slicing of its resources.

Finally, Table I, presents a comparison between the contributions of this article and those of other survey papers.

C. The Organization of the Article

The remainder of this paper is structured as follows: Section II highlights the key features of O-RAN and provides an overview of open source projects, activities, and contributions, along with the standardization efforts associated

Sec.I

Introductory Remarks
- Literature review gap analysis
- Goals and contributions
- The organization of the article

Sec.II

O-RAN Standards and Open Source Efforts
- Key feates and principles of O-RAN
- Network slicing
- O-RAN standards and activities
- O-RAN-related open ource projects
- Open testbeds

Sec.III

O-RAN Slicing-aware Architecture
- Major components of O-RAN
- The O-gNB and interfaces
- The SMO framework and interfaces
- The underlying infrastructures

Sec.IV

O-RAN Deployment Scenario
- O-RAN NF deployment scenarios
- Network slicing deployment scenarios
- SMO framework deployment options
- 3GPP and NFV-MANO based SMO deployment
- 3GPP and ONAP based SMO deployment

Sec.V

Slicing the Underlying O-RAN Infrastructure
- O-RAN cellular site
- O-RAN cloud platform
- Xhaul transport network
- Xhaul transport network slicing
- Mapping of NSIs to physical infrastructure

Sec.VI

Network Slicing Use Cases in O-RAN
- Slice subnet management & provisioning
- Slice SLA assurance
- Multi-vendor slices
- Advantages of multi-vendor slices
- NSSI resource allocation optimization

Sec.VII

Key Lessons Learned
- Related to O-RAN architecture
- Related to open interfaces
- Related to O-RAN transport network
- Related to O-RAN slicing

Sec.VIII

Major Research Challenges
- Interoperability, performance, and automation
- RIC apps conflict mitigation, and security
- Network slicing and orchestration
- Understanding O-RAN ecosystem

Sec.IX

Conclusions and Future Works
- Background & concluding remarks
- Summary of the content covered in this paper
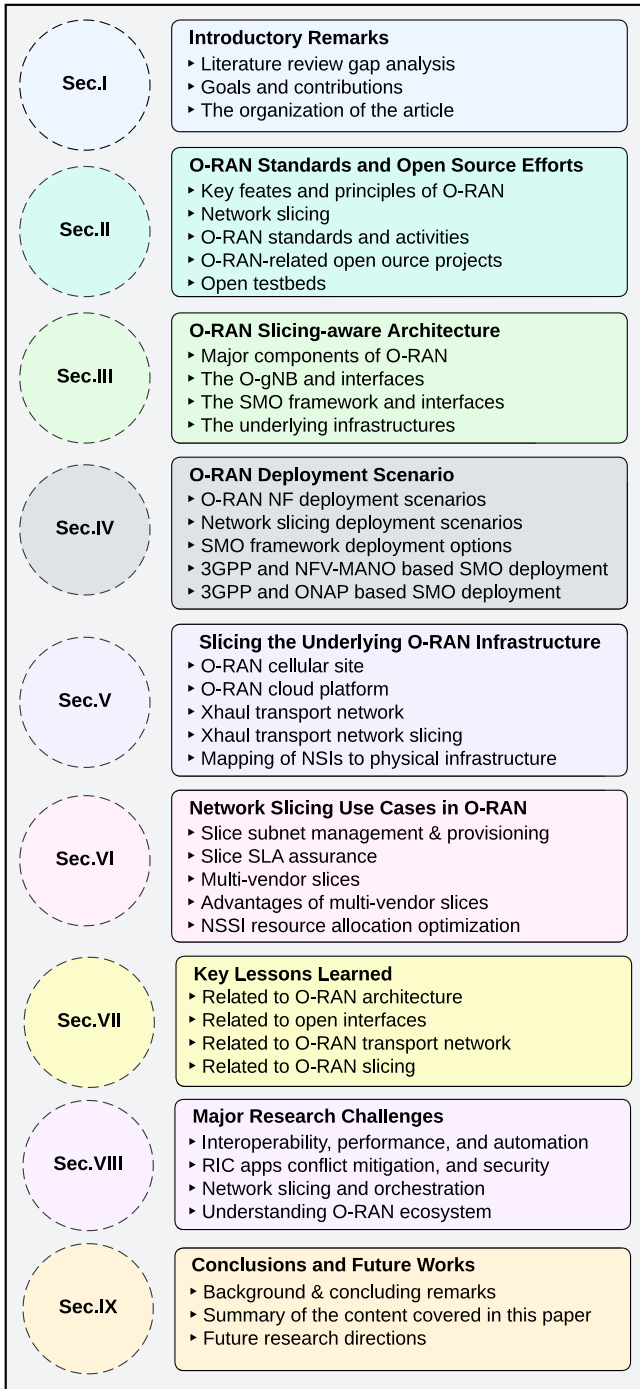- Future research directions

Fig. 1. Overview of the organization and structure of this survey paper. Each box in the figure represents one of the chapters of the paper, encapsulating their respective contributions and themes.

with O-RAN. In Section III, we introduce the architectural components and open interfaces of the O-RAN slicing-aware architecture, highlighting its features and interactions with 3GPP-defined network components, service management, and service orchestration. Section IV covers various deployment scenarios for O-RAN, slicing, and the SMO proposed by the O-RAN Alliance tailored to different use cases. Section V discusses slicing the underlying infrastructure, elaborating O-RAN cloud platform (O-Cloud) slicing, TN slicing, and TN

slice orchestration. In Section VI, we outline high-level use cases expected to be prioritized by the O-RAN community, particularly regarding RAN slicing. Section VII provides an overview of the lessons learned through this survey, while Section VIII identifies key research challenges that require further research and investigation. Finally, Section IX summarizes our work, draws conclusions, and suggests potential directions for future research. An overview of the organization of this survey is illustrated in Figure 1.

II. Ongoing Standardization Efforts and Open Source Contributions to O-RAN

This section provides a detailed overview of the key features of the O-RAN architecture. It also examines ongoing open source initiatives that are actively shaping O-RAN, as well as publicly available experimental platforms that support the development and validation of O-RAN components and interfaces. In addition, we provide an overview of state-of-the-art contributions across various *de facto and de jure* standards development organizations (SDOs) involved in the evolution of the O-RAN. The goal is to provide a comprehensive understanding of the current advancements and collaborative efforts driving innovation within the O-RAN ecosystem.

A. Key Features and Principles of O-RAN

The primary objective behind O-RAN for service providers and network operators is to diversify vendor partnerships and avoid vendor lock-in by enabling the use of non-proprietary software and hardware components source from multiple vendors [58]. Traditionally, RAN has been proprietary and vertically integrated, wherein both hardware and software were tightly coupled and delivered by a single vendor [27]. O-RAN aims to disrupt this conventional model by promoting open interfaces and fostering interoperability among RAN components. The O-RAN architecture emphasizes modularity and flexibility through open standards, enabling operators to integrate hardware and software solutions from different suppliers [51], [59]. This flexibility facilitates enhanced customization of network configuration and encourages broader participation from second- and third-tier equipment manufacturers.

Beyond open interfaces, O-RAN enables full access to NG-RAN through AI-based control mechanisms that support real-time monitoring, proactive resource allocation, and adaptive responsiveness to dynamic radio conditions [60], [61]. It further promotes a disaggregated, virtualized, cloud-native, and interoperable RAN [19], empowering service providers to deploy a fully programmable, intelligent, autonomous, and multi-vendor RAN suited for 5G, 6G, and beyond networks [62], [63].

In the following, we summarize the core principles and defining characteristics of O-RAN, highlighting aspects particularly relevant to service providers, network operators, and other stakeholders in the telecommunications ecosystem.

1) Intelligent and Programmable Network: The O-RAN architecture is inherently intelligent and programmable. This

enables dynamic optimization of network operations and maintenance (OAM) [64]. This programmability allows cellular networks to efficiently adapt to diverse traffic demands and deployment scenarios, supporting the evolving requirements of next-generation wireless communication systems [65].

2) Data Center Economics in the RAN: O-RAN brings the economic principles of DC into the RAN. By leveraging virtualization, COTS hardware, and centralized resource management, it enhances the scalability and economic sustainability of RAN infrastructure by aligning it with the cost-efficient practices of modern DC environments [66]. Through optimized resource utilization, simplified deployment, and improved maintenance processes, O-RAN reduces operational expenditure (OPEX) and significantly lowers the total cost of ownership (TCO) [67].

3) Automation and Manageability: O-RAN places strong emphasis on automation and centralized manageability, aiming to reduce manual intervention in network OAM [65]. It improves operational efficiency while enhancing the reliability, consistency, and scalability of network management processes.

4) Faster Time to Market and Innovation Agility: O-RAN accelerates the deployment of network solutions by enabling service providers and operators to introduce new features and services more rapidly [68]. The modular and open architecture of O-RAN fosters innovation agility by enabling the rapid development, testing, and deployment of new technologies and features. Through standardized interfaces and software-defined components, it facilitates seamless integration of emerging solutions–such as AI/ML-driven xApps and rApps–into the RAN [68]. This agility is essential for maintaining competitiveness in the rapidly evolving telecoms landscape, ensuring that the network keeps pace with technological advancements and shifting industry requirements.

5) Vendor Diversity: O-RAN promotes a competitive, multi-vendor ecosystem by decoupling hardware and software components through standardized interfaces. This architectural openness allows network operators and service providers to select equipment and solutions from a broad range of vendors, promoting flexibility in network design and deployment. Such diversity not only encourages innovation and accelerates technology evolution but also enables operators to tailor their networks and services to specific technical and business requirements [69]. Furthermore, it reduces capital expenditure (CAPEX) by mitigating vendor lock-in and supporting cost-effective integration of interoperable components.

6) Open Source Software: O-RAN leverages open source software to support the development of reference implementations aligned with its open and intelligent RAN architecture [66]. The O-RAN software community (OSC)—a collaboration between the O-RAN Alliance and the Linux Foundation (LF)—plays a central role in implementing O-RAN specifications through open source projects. This approach fosters interoperability, accelerates innovation, and promotes collaboration among diverse industry stakeholders. As a result, it significantly facilitates the evolution, validation, and deployment of disaggregated, vendor-neutral RAN solutions within the telecommunications ecosystem.

7) AI/ML for O-RAN Optimization and Resource Management: One primary enabler of O-RAN is AI/ML. Some AI/ML algorithms are already proposed to optimize performance, enhance autonomous management, and support real-time, data-driven intelligence across O-RAN [70], [71], [72]. A key aspect of this integration is the evolution of AI/ML-driven RICs [72]. This empowers service providers, vendors, and third party developers to deploy intelligent applications (e.g., xApps and rApps) for automated, closed-loop network optimization [64], [72], [73], [74]. In addition, O-RAN enables embedding intelligent control across multiple functional domains, including radio resource management (RRM) and service orchestration [71]. In this context, AI/ML facilitates predictive real-time decision-making to address a broad spectrum of use cases, such as interference mitigation, energy efficiency, dynamic load balancing, QoS-aware scheduling, and self-healing capabilities [70]. Furthermore, AI/ML is integral to O-RAN management in order to offer advanced predictive maintenance, traffic forecasting, and continuous optimization of performance in response to dynamic operating conditions.

Moreover, AI/ML serves as critical enablers for dynamic and intelligent slicing in O-RAN. It can facilitate the creation, adaptation, and real-time management of multiple virtualized, service-specific slices [75]. It empowers O-RAN to intelligently monitor, predict, and respond to fluctuating network demands, thereby optimizing end-to-end (E2E) QoS and resource utilization. Various types of AI/ML algorithms can be beneficial for addressing different aspects of O-RAN slicing. For example, reinforcement learning (RL) and multi-agent RL (MARL) enable adaptive policy learning and slice orchestration, while long short-term memory (LSTM) networks support proactive traffic forecasting and timely reconfiguration of slice parameters [76]. Additionally, federated learning (FL) has emerged as a promising approach for distributed slice management, preserving data privacy while ensuring coordinated learning across decentralized nodes [75]. Lastly, deep learning (DL) models, including convolutional neural networks (CNNs) and Transformers, are increasingly applied for tasks such as anomaly detection and intelligent beamforming.

The O-RAN Alliance plays a crucial role in the development and promotion of O-RAN standards. By embracing the principles set forth by the alliance, the telecommunication industry aims to accelerate innovation, reduce deployment costs, and cultivate a more dynamic and competitive marketplace for RAN solutions towards 5G, 6G, and beyond networks [27].

B. Network Slicing

Network slicing is a transformative architectural paradigm that enables multiple, logically isolated virtual networks—referred to as "slices"—to coexist over a shared physical infrastructure [23]. As a foundational element of 5G, 6G, and future systems, it addresses heterogeneous service demands and unlocks the full potential of next-generation networks. At its core, it partitions the physical network into discrete virtualized slices, each independently configured to specific service requirements and performance criteria—such as bandwidth,

latency, reliability, and security [77], [78]. These slices operate autonomously, allowing concurrent support for diverse use cases—ranging from latency-sensitive industrial automation to bandwidth-intensive mobile broadband—within the same physical network infrastructure [79], [80].

An E2E network slice spans all network domains, encompassing the RAN, TN, and core network (CN) segments [81], [82]. Each network slice is carefully engineered to meet the specific requirements of different services, ensuring logical isolation among slices. This isolation preserves the integrity of individual network slices by preventing faults or malfunctions in one from affecting others, thereby fostering the autonomy and reliability across virtualized networks [83].

To guarantee performance and service quality, operators allocate dedicated resources to each slice—such as computing capacity, bandwidth, QoS provisions, and other critical elements [84]. This resource assurance underscores the commitment to supporting the diverse service types, while ensuring efficiency, scalability and robustness across a cellular network.

Despite the substantial progress in E2E slicing, several challenges persist in the realization of NG-RAN slicing [35], [79]. The complexity arises primarily from the need to balance varying degrees of isolation and resource sharing, while tailoring the user plane (UP) and control plane (CP) functionalities to meet the specific requirements of individual slices [39], [85], [86]. The key challenges include managing the trade-off between resource utilization efficiency and isolation, harmonizing inter-RAN and intra-RAN resource allocation algorithms, and prioritizing slices effectively across different layers of the RAN [19]. In addition, the limited availability of radio resources demands highly efficient resource management strategies to sustain optimal network performance. The introduction of advanced 5G new radio (NR) features–such as bandwidth partitioning (BWP) and physical numerology–further amplifies these challenges by increasing configuration complexity and the need for dynamic adaptation [87], [88].

To address the above challenges, the 3GPP provided guidelines in Release 17 for realizing slicing in NG-RAN. These guidelines encompass various aspects, including support for diverse QoS types, resource segregation, SLA enforcement, among others [89]. The 3GPP specifications further enhance architectural flexibility by presenting multiple implementation options for RAN slicing, such as layer 1 (L1), layer 2 (L2), or the medium access control (MAC)-based approaches [89], [90]. In addition, they specify a M&O framework to support the efficient lifecycle management of RAN slices and their associated resources across the NG-RAN. This framework also ensures interoperability with other standardized architectures for the realization of E2E slicing [19].

C. O-RAN Standards and Activities

As of this writing, numerous de facto and de jure organizations are actively involved—both directly and indirectly—in developing standards for software and hardware components that align with the principles of O-RAN. These efforts

are coordinated through collaborative initiative under various SDOs across the globe. In this subsection, we dive into an exploration of these SDOs and their respective contributions towards the advancement and realization of O-RAN.

1) 3GPP: 3GPP does not directly define standards specific to the O-RAN. However, many architectural components of the 3GPP-defined RAN architecture—including NFs, M&O frameworks, functional split options, and interface specifications—have been adopted and further extended by other SDOs, most notably the O-RAN Alliance, to establish comprehensive standards for the O-RAN architecture. The 3GPP specifications provide a foundational and system-level definition of RAN architecture, distributed across its various Technical Specification Groups (TSGs).

In the course of 5G evolution, 3GPP evaluated eight functional split options and ultimately standardized two NG-RAN split architectures. The first is the high layer split (HLS), corresponding to option 2 from the 3GPP study. It involves dividing the BBU into CU and DU. The second split involves CP and UP separation within the CU, introducing a logical division of signaling and data-handling responsibilities [51]. To support this architecture, 3GPP introduced the F1 interface, which connects CUs to DUs and the E1 interface, which facilitates coordination between CP and UP [91], [92]. The functional split options are analyzed in our earlier work [19].

The introduction of functional split in 3GPP represents a critical step towards disaggregating the standard protocol stack. This involves separating the processing of a specific layer within the protocol stack from the computing entity, thereby promoting architectural openness, intelligent cellular interface, and the feasibility of network slicing [52]. This functional split has served as a pivotal catalyst for the development of subsequent O-RAN-related specifications.

2) O-RAN Alliance: The O-RAN Alliance, established in 2018, is a global industry consortium committed to the ambitious task of modernizing traditional RAN architecture. Its central mission revolves around steering the wireless communication industry towards a future defined by openness, intelligence, automation, cloudification, virtualization, and interoperability within the RAN [93], [94]. This transformative journey is underpinned by a shift towards virtualized and cloud-native network components, the adoption of clearbox hardware, and the implementation of open, standardized interfaces that facilitates seamless communication between various software and hardware components of the O-RAN architecture [93].

To achieve this vision, the O-RAN Alliance follows a systematically organized technical specification governed by its technical steering committee (TSC). The TSC plays a pivotal role in decision-making and provides essential guidance on O-RAN technical matters. It assumes the crucial responsibility for approving specifications prior to their submission for board approval and eventual publication. The current structure of the TSC encompasses eleven technical working groups (WGs), five focus groups (FGs), a dedicated research group, an open source software community, and a minimum viable plan - committee (MVP-C). These specialized divisions collaborate to focus on specific aspects of O-RAN, contributing collectively

to its development, deployment, and evolution of its technical standards. An overview of the specific objectives and focus areas of each division is provided in Table II.

3) Telecom Infra Project: The Telecom Infra Project (TIP) is a global consortium of over 500 companies and organizations working to accelerate the development and adoption of open, disaggregated, intelligent, and standards-based technologies for telecommunications infrastructure. It has contributed significantly to open standards and specifications, with its solutions adopted by major operators worldwide to address specific commercial and operational requirements. TIP's work is organized into multiple project groups (PGs) focused on products, solutions, and software across domains such as RAN, TN, and CN, along with associated M&O layers.

Within TIP's various PGs, the OpenRAN PG plays a central role in enabling an open ecosystem that redefines the 3GPP-based NG-RAN architecture using open components and standardized interfaces. It supports the evolution of 4G, 5G, and future networks by promoting interoperability and vendor diversity. The goals of the OpenRAN PG includes developing and validating interoperable OpenRAN solutions, advancing innovative platforms for network management, and fostering collaboration among operators, vendors, integrators, and global stakeholders [95]. The PG focuses to align requirements for key OpenRAN components, including O-RUs, O-DUs, and O-CUs, which are extensively tested, and validated in TIP Community Labs and PlugFests [96].

The OpenRAN PG is organized into two main workstreams:

a) Component subgroups: They are dedicated on enhancing the performance of individual OpenRAN software and hardware components such as radio unit (RU), DU, CU, radio intelligence automation (RIA), and OpenRAN orchestration and management automation (ROMA).

b) Segment subgroups: They concentrate on developing integrated RAN solutions tailored to diverse deployment scenarios across indoor and outdoor use cases.

In June 2024, the OpenRAN PG released its Release 4 Technical Priorities Document, presenting critical deployment requirements [97]. It addresses radio configurations, hardware/software requirements for each OpenRAN building block, and evolving requirements for the SMO, RIC, and the cloud infrastructure hosting O-RAN components. It emphasizes security, energy efficiency, consolidating related requirements and identifying new priorities across workstreams. These efforts aim to accelerate the development and global adoption of competitive OpenRAN solutions [95], [97].

4) Small Cell Forum: The small cell forum (SCF) is a global organization focused on developing technical specifications and tools to accelerate the adoption of flexible, cost-effective, and scalable cellular network infrastructure. The SCF has played a key role in standardizing essential elements of network technology, including the functional API (FAPI), network FAPI (nFAPI), and enhancement to the X2 interface. These specifications enable an open, multi-vendor platform, thereby reducing the barriers to the densification of stakeholders in the wireless communications industry [98].

The SCF has established its own Open RAN ecosystem with a particular emphasis on small cell deployments. A notable contribution is the introduction of the nFAPI protocol, which pioneered 3GPP's split option 6 [65], dividing the MAC and physical layer (PHY) layers, with the PHY hosted in the small cell RU (S-RU). The nFAPI is pivotal in empowering multi-vendor interoperability, allowing a small cell CU/DU to connect seamlessly with independently deployed S-RUs [99]. It also provides tools and integration support, including the SCF disaggregated RAN transport study (DARTs) suite.

In collaboration with O-RAN Alliance and TIP, SCF has contributed significantly to the advancement of standardized testing processes across the industry, which includes active participation in plugfests [98]. Beyond this, SCF engages with other prominent organizations such as 3GPP, OpenAirInterface (OAI), and a wide range of stakeholders across technical, commercial, and regulatory domains. The forum's goal is to accelerate open RAN adoption across all domains, driving the widespread deployment of virtualized open RAN infrastructure. By promoting the convergence of open systems, open source code, and shared spectrum, SCF aims to enable a broader range of network deployers. This is particularly impactful in areas like enterprise and smart city environments, where small cells are indispensable [98].

D. O-RAN-Related Open Source Projects

The software community plays an essential role in ensuring that software reference implementations are closely aligned with O-RAN and its technical specifications. In this context, the OSC undertakes various responsibilities, including the development and maintenance of open source software, fostering collaboration with other open source initiatives, and promoting related projects and activities that contribute to the advancement of O-RAN. As of this writing, multiple open source platforms compliant with O-RAN principles are publicly available and actively utilized by researchers and academic institutions. These platforms provide a foundational environment for experimentation, prototyping and validation of O-RAN functionalities. The subsequent section examines key contributors and collaborators involved in the implementation and evolution of open source solutions in O-RAN.

1) ONAP: The open network automation platform (ONAP), launched by LF in 2017, is an open source platform for orchestrating, managing, and automating network and edge computing services. It addresses the needs of network operators, cloud providers, and enterprises through real-time, policy-driven orchestration and lifecycle automation of physical and virtual NFs [96]. ONAP leverages SDN and NFV technologies and implements a complete management and orchestration (MANO) layer aligned with ETSI NFV architecture. Beyond supporting fault, configuration, accounting, performance, security (FCAPS) functionalities, ONAP offers a robust framework for network service design.

ONAP collaborates with the OSC, particularly on deploying the SMO and integrating the Non-RT RIC functionalities [72]. This partnership enhances coordination, minimizes duplication of efforts, and streamlines development. Shared priorities are outlined in [100], including a joint study on

TABLE II
SUMMARY OF CONTRIBUTIONS AND FOCUS AREAS ACROSS MULTIPLE WGS SUPERVISED BY THE TSC WITHIN THE O-RAN ALLIANCE

Group	Title	Principal Areas of Focus and Notable Contributions
WG1	Use Cases and Overall Architecture WG	• Exploring a number of use cases, defining system-level requirements, introducing numerous deployment scenarios, and proposing a comprehensive architecture for O-RAN • Investigation into network slicing within O-RAN, including defining several use cases, key requirements, and introducing slicing-aware architecture with interface extensions • Coordination of proof of concepts to demonstrate O-RAN products to the market
WG2	Non-RT RIC and A1 Interface WG	• Defining an architecture for Non-RT RIC and its functionalities, and incorporating the R1 interface to connect the Non-RT RIC framework with rApps • Expanding R1 services within the functionalities of the Non-RT RIC, and enabling interoperability among the interfaces of the various management components of the SMO framework • Discussing the A1 interface, the interface between the Non-RT RIC and the Near-RT RIC, including associated use cases, deployment scenarios, and applications
WG3	Near-RT RIC and E2 Interface WG	• Specifying E2 interface – an interface between the Near-RT RIC and the E2 nodes • Defining the Near-RT RIC architecture and introducing application programming interfaces (APIs) to connect the Near-RT RIC platform and the xApps • Defining several use cases, requirements, and management specifications for the Near-RT RIC, and contributing to service models for E2 interface and E2 nodes
WG4	Open Fronthaul Interfaces WG	• Establishing specifications for an open fronthaul (O-FH) interface between the O-DU and the O-RU within an O-gNB • Setting standards for Control, User, Synchronization, and Management Plane protocols with their corresponding YANG models for the O-FH link • Developing specifications for transport interfaces and conducting O-FH interoperability tests
WG5	Open F1/W1/E1/X2/Xn Interface WG	• Providing interoperable multi-vendor specifications aligned with 3GPP-defined standards for F1, W1, E1, X2, and Xn interfaces, enhancing the overall O-RAN architecture • Defining specifications for O1 interface, covering interaction between the O-CU and O-DU with SMO and discussing the OAM functions • Developing open MH and backhaul (BH) interoperability test specifications
WG6	Cloudification and Orchestration WG	• Specifying cloud-native and virtualized infrastructure for hosting the O-CU and O-DU of an O-gNB, focusing on hardware-software decoupling within the underlying infrastructure • Providing technology and reference designs for leveraging commodity hardware platforms • Identifying use cases, deployment scenarios, and requirements for cloud resource hosting, and defining high-level orchestration architecture for SMO framework and O-Cloud interaction
WG7	White-box Hardware WG	• Specifying standards for comprehensive reference design of high-performance, spectral-efficient, and energy-efficient white box base stations within the O-RAN architecture • Promoting decoupled software and hardware platform for O-RAN components and interfaces • Addressing outdoor and indoor cells with various split options, along with O-FH interface
WG8	Stack Reference Design WG	• Developing a software architecture as well as define a comprehensive design and release plan for O-CU and O-DU of an O-gNB, tailored for NR protocol stack • Providing specifications for interoperability testing of various O-CU and O-DU deployment scenarios with other O-RAN components and interfaces
WG9	Open X-haul Transport WG	• Designing an open TN within O-RAN, meeting FH, MH, and BH service requirements • Concentrating on open transport domain, including transport equipment, physical media, and associated control and/or management protocols within the open TN
WG10	OAM for O-RAN	• Specifying OAM architecture for O-RAN and management services for O1 interface, such as proposing a set of unified operation and notification mechanisms • Developing information models and data models for OAM architecture in O-RAN
WG11	Security WG	• Establishing specifications for O-RAN's security, including its NFs, interfaces, and (r/x)Apps • Defining requirements, use cases, architectures, and protocols to ensure security and privacy of various types of data and stakeholders within the O-RAN architecture
SDFG	Standard Development Focus Group	• Leading in formulating standardization strategies for the O-RAN Alliance and serving as the primary interface between the O-RAN Alliance and other relevant SDOs • Managing coordination of both incoming and outgoing liaison statements
IEFG	Industry Engagement Focus Group	• Engaging with leading industry players and members of the O-RAN Alliance to drive adoption, spread, and ongoing innovation of O-RAN-based technologies and solutions
OSFG	Open Source Focus Group	• Managing O-RAN Alliance's open source activities, including establishing the OSC and developing open source related strategies • Collaborating with other open source communities to drive innovation and adoption of O-RAN
TIFG	Test & Integration Focus Group	• Defining testing and integration approaches, coordinating specifications across various WGs, including E2E test specifications and productization profiles • Planning PlugFests and offering guidelines for third-party open testing and integration centers (OTICs), facilitating integration and verification processes

TABLE II
(Continued.) SUMMARY OF CONTRIBUTIONS AND FOCUS AREAS ACROSS MULTIPLE WGs SUPERVISED BY THE TSC WITHIN THE O-RAN ALLIANCE

Group	Title	Principal Areas of Focus and Notable Contributions
SuFG	Sustainability Focus Group	• Focusing on enhancing energy efficiency and reducing environmental impact in O-RAN • Collaborating with MVP-C to align initiatives across all WGs and FGs within O-RAN Alliance
nGRG	next Generation Research Group	• Researching intelligent O-RAN principles for standardizing 6G and beyond systems • Driving network evolution towards greater intelligence and performance using new technologies
MVP-C	Minimum Viable Plan – Committee	• Providing roadmap for implementing comprehensive O-RAN solutions in commercial networks • Managing O-RAN features, including creation, prioritization, and tracking documents • Approving feature creation and inclusion in relevant releases; collaborating and coordinating with all WGs and FGs within the O-RAN Alliance
OSC	O-RAN Software Community	• Leading the development and possible deployment of open source software for O-RAN architecture in collaboration with the LF • Focusing on aligning with the open architecture and criteria of the O-RAN Alliance to deliver a solution suitable for commercial deployment and O-RAN components and interfaces

the "ONAP/3GPP & O-RAN Alignment-Standards Defined Notifications over VES" use case, which seeks to align ONAP with O-RAN Alliance and 3GPP, fostering interoperability and broader adoption [101]. The Kohn release of ONAP further advances this integration by enhancing cloud-native NF orchestration, supporting intent-driven closed-loop automation, and E2E network slicing across 5G RAN, CN, and TN domains [102], [103].

2) OpenAirInterface: The OpenAirInterface software alliance (OSA) is a nonprofit organization founded by the French research institute EURECOM. It supports a global community of researchers and industry contributors in the development of open source software for the CN and RAN domains of a 3GPP cellular networks. The Alliance supports the advancement of the 3GPP 5G cellular stack, which is maintained within the OAI software packages and are designed to operate on COTS hardware. The OSA is responsible for roadmap development, quality assurance, and community engagement, including the promotion of OAI software packages. These packages are widely used by academic institutions and business communities for a broad range of use-cases. The goal of the OSA is to accelerate the adoption of the OAI.

In the context of 5G, the OAI community and software assets have been expanding rapidly. Current active projects include: OAI 5G RAN, OAI 5G CORE, MOSAIC 5G (M5G), and continuous integration (CI)/continuous deployment (CD). The newly created M5G PG aims to transform both the RAN and CN into agile and open platform for network service delivery. The M5G PG focuses on developing software implementations of the O-RAN E2 protocol, as well as FlexRIC (a flexible RIC), FlexCN (a flexible core controller), and intelligent orchestration tools for RAN and CN domains [104]. Additionally, researchers at Northeastern University have successfully integrated OAI with the OSC RIC, enabling interoperability between open source development and standardized O-RAN implementations. Further details are provided in [105], [106].

3) Open Networking Foundation: Open networking foundation (ONF) is a consortium led by several major network operators, playing a pivotal role in driving the transformation of network infrastructure through open and disaggregated solutions. Among its key initiatives, the software-defined RAN (SD-RAN) project contributes open source components to

the open RAN by developing and testing O-RAN compliant network elements. The project promotes multi-vendor RAN solutions and demonstrates the potential of modular component integration to foster further innovation in RAN [107].

In close collaboration with the O-RAN Alliance and OSC, SD-RAN aims to develop open source components for the O-CU CP, O-CU UP, and O-DU [51]. A cornerstone of the platform is the cloud-native uONOS-RIC (micro-ONOS-RIC), a fully functional Near-RT RIC, which includes an xApp development environment, and a set of reference xApps for managing open RAN elements [107].

Notably, Deutsche Telekom has deployed a fully disaggregated 5G field trial using the SD-RAN platform, integrating components from more than eight vendors via uONOS-RIC. This deployment represents the first comprehensive realization of O-RAN, encompassing O-RU, O-DU, O-CU, RIC, and multiple xApps sourced from various providers—marking a significant milestone in the evolution of open RAN.

4) srsRAN: The srsRAN project, developed by software radio system (SRS), is an open source RAN solution that supports both 4G and 5G technologies. It features an O-RAN-native gNB, providing a comprehensive implementation of the L1/L2/layer 3 (L3) protocol stack with minimal external dependencies [108]. The solution adheres to standards defined by both 3GPP and the O-RAN Alliance. It adopts the 3GPP 5G architecture, implementing functional splits between the DU and CU, with further separation into CU control plane (CU-CP) and CU user plane (CU-UP). The srsRAN platform also supports integration with third-party Near-RT RICs and xApps through FlexRIC, with the ultimate goal of achieving full compliance with the E2 interface [109].

The srsRAN offers deployment flexibility, allowing users to operate a monolithic gNB on a single machine or distribute RAN functions across multiple machines and geographic locations. It supports seamless integration with third-party RICs, PHY solutions, and other O-RAN compliant hardware and software components, making it well-suited for a wide range of use cases, and experimental scenarios.

5) OpenRAN Gym: OpenRAN Gym, led by Northeastern University, is a collaborative open source initiative designed for large-scale, data-driven experimental research within the open RAN ecosystem [45], [110]. Its goal is to unite researchers from academia and industry in a cooperative

environment to accelerate the development of intelligent and AI-driven solutions for open RAN.

OpenRAN Gym builds on frameworks for data collection and RAN control, which enables E2E design and testing of data-driven xApps by providing an O-RAN compliant Near-RT RIC and E2 termination interface [45]. This allows users to collect runtime data, prototype new strategies, and evaluate them in diverse wireless environments before transitioning them into production networks.

The architecture of OpenRAN Gym, as detailed in [45], [110], consists of the following key components:

- Remotely accessible wireless testbeds, like Colosseum, Arena, and PAWR platforms, which support large-scale data collection and validation in real-world scenarios.
- A softwarized RAN using open protocol stacks like srsRAN and OAI to emulate cellular networks.
- A data collection and control framework, such as SCOPE, that extracts key performance indicators (KPIs) and key performance measurements (KPMs) and supports runtime RAN control [45].
- An O-RAN compliant control architecture, such as ColO-RAN, which connects to the RAN through standardized E2 interface, receives runtime KPMs, and coordinates intelligent control through AI/ML-based xApps and rApps.

E. Open Access Testbeds

In addition to the aforementioned open source projects, several experimental testbeds have been developed to support the implementation of softwarized 5G networks by leveraging a variety of open source components. The following subsections provide an overview of selected open-access testbeds.

1) Colosseum: Colosseum is a publicly accessible, large-scale wireless testbed designed to support advanced experimental research through the use of virtualized and softwarized waveforms and protocol stacks, deployed on a fully programmable clear-box platform. Equipped with 256 state-of-the-art software-defined radios (SDRs) and a powerful channel emulator, Colosseum has the capability to simulate almost any scenario. This allows the comprehensive design, development, and validation of solutions at scale across various deployment topologies and channel conditions. It achieves high-fidelity reproduction of radio frequency scenarios through field programmable gate array (FPGA)-based emulation employing finite impulse response (FIR) filters. These filters accurately model channel taps and apply them to SDR-generated signals, enabling realistic replication of real-world radio environments—crucial for evaluating performance under controlled and repeatable conditions [111], [112].

Colosseum also serves as the foundational infrastructure for OpenRAN Gym, which is tightly integrated within the testbed. This integration enables experimentation with E2E O-RAN compliant networks and services, facilitates data collection, and supports the development and validation of AI/ML-driven models, among other essential research activities [45].

2) Powder: Platform for open wireless data-driven experimental research (POWDER) is a city-scale testbed tailored to support a wide range of innovative research experiments. Its infrastructure includes an outdoor area incorporating multiple SDRs-equipped nodes, an indoor laboratory for over-the-air experiments, and a wired attenuator matrix [112], providing an environment for filed tests and controlled lab evaluations.

The primary goal of POWDER is to enable experimental research across heterogeneous wireless technologies, with a particular emphasis on 5G RAN architecture, network orchestration, massive multiple-input multiple-output (MIMO), and many more. The platform includes integrated support for the rapid deployment of the O-RAN architecture, facilitating the configuration and testing of advanced RAN functionalities. Researchers can efficiently prototype and evaluate components such as the Near-RT RIC, xApps, the O-CU subsystem, and open source SMO within POWDER [113], [114].

3) Cosmos: The cloud enhanced open software-defined mobile wireless (COSMOS) is a city-Scale testbed, deployed as a component of the POWDER initiative. The COSMOS is aims to create, develop, and operate an advanced, city-scale wireless testbed that facilitates real-world experimentation with next-generation wireless technologies and applications. It has been certified by the O-RAN Alliance as an open testing and integration center (OTIC) [112], [115]. The COSMOS architecture prioritizes ultra-high bandwidth and low-latency wireless communication, tightly integrated with edge cloud computing. The testbed consists of approximately 40-50 advanced SDR nodes, interconnected via fiber-optic FH and backhaul (BH) TNs, alongside dedicated edge and core cloud computing infrastructure. Researchers can access COSMOS remotely through a Web-based portal, which provides comprehensive tools for experiment orchestration, real-time measurements, and data collection [115].

4) Arena: Arena stands as an innovative open-access wireless testing platform designed to research in sub-6 GHz 5G and beyond spectrum. Located in an indoor office environment, the testbed is anchored by a grid of ceiling-mounted antennas, each connected to programmable SDRs, enabling real-time, scalable, and reproducible experimentation [116]. The platform integrates 12 high-performance computational servers, 24 symbol-level synchronized SDRs, and a total of 64 antennas, providing a unique combination of processing capabilities and spatial diversity. This architecture makes Arena particularly well-suited for exploring technologies in dense spectrum environment [116]. Arena operates on a three-tier physical design: servers and SDRs are housed in a dedicated room, while antennas are strategically distributed across the office ceiling and connected via 100-foot long radio frequency (RF) cables. This layout ensures minimal interference, precise control over experimental parameters, and a realistic representation of indoor wireless propagation characteristics [112], [116].

5) X5G: X5G is a pioneering private 5G network testbed at Northeastern University, Boston. It integrates open source and programmable components across the entire network stack—from the PHY to the CN. Notably, it stands as the first fully programmable multi-vendor and O-RAN compliant testbed of its kind. Developed through a collaborative endeavor involving Northeastern University, NVIDIA, and OAI [117]. The testbed leverage NVIDIA graphics processing units (GPUs) to accelerate L1 (PHY operations), while L2 and L3 are

implemented using the OAI software stack. This integration is based on the SCF FAPI for seamless interaction between the MAC and PHY layers. This integration enables the inline hardware acceleration of computationally demanding PHY tasks on the GPU, fostering scalability and facilitating the integration of AI/ML within the RAN. The NVIDIA aerial RAN CoLab (ARC) platform operates on a specialized multi-vendor infrastructure comprising eight servers for CU and DU, along with four RUs designed for lab installations. It also incorporates O-RAN 7.2x FH, precise timing hardware, and a dedicated 5G CN. By combining the performance advantages of NVIDIA ARC with OAI, X5G provides flexible, high-performance environment for exploring open, programmable, and intelligent wireless systems—paving the way for future advancements in 5G, 6G, and beyond [117].

6) 5GENESIS: The 5GENESIS initiative, funded by the European Union (EU), aims to validate 5G KPIs across a broad spectrum of applications, ranging from controlled laboratory environments to large-scale public events. The project builds upon the collective outcomes of multiple EU projects along with internal research and development contributions from its consortium partners, to establish a unified, E2E 5G infrastructure spanning five test platforms across Europe. Each platform within 5GENESIS is characterized by distinct capabilities and specialized features. However, they are engineered for interoperability within a cohesive architecture, forming a flexible and distributed testing facility. This infrastructure enables comprehensive KPIs validation, supports targeted demonstrations, and facilitates the evaluation of critical 5G and beyond use cases, including diverse deployment scenarios. The trials conducted within the 5GENESIS are primarily focused on assessing and confirming the KPIs defined by 5G Public-Private Partnership (5G-PPP). These evaluations inherently serve to benchmark the performance, scalability and readiness of each individual platform and the collective infrastructure as a whole [118].

7) Insights From Open-Source O-RAN Deployments: Building on the open-source platforms and testbeds introduced earlier, this section synthesizes representative findings from recent evaluations of open 5G deployments.

Our study [119] compares the deployment of commercial and open-source O-gNBs in an industrial environment. The findings indicate that the commercial gNBs offer better coverage, while the open-source gNBs achieve lower and more consistent latency. Despite being less energy-efficient, the open-source gNBs provide greater flexibility, easier configuration, and better maintainability, supported by a vibrant developer community. In contrast, the commercial gNBs operate within a closed system, offering limited scope for customization.

The work in [120] emphasizes how component configurations and interdependency significantly influence E2E performance. Through over-the-air measurements of packet loss and one-way delay under concurrent uplink and downlink transmissions, the authors show that open-source O-RAN-based systems can, under certain conditions, match or surpass conventional solutions in performance, while offering industrial-grade connectivity at lower cost. Interoperability is evaluated in [121], where various open-source RAN and CN combinations are tested, providing critical insights into integration and performance trade-offs.

A performance evaluation presented in [122] highlights the suitability of the srsRAN platform for private 5G network deployments. The study also identifies key limitations, emphasizing that the full potential of 5G remains unrealized–primarily due to constraints associated with general-purpose hardware and the current maturity level of open-source software components. The study [123] employs an OAI-based testbed to evaluate O-RAN performance, with a focus on throughput measurements. The findings expose limitations of the setup and suggest potential underlying causes.

The authors in [124] present a standalone testbed to evaluate downlink performance across two O-RAN implementations: srsRAN and OAI. The results indicate that functional split architectures enhance flexibility and resource efficiency by offloading processing to the O-DU. In [125], the authors propose a cloud-native O-RAN testbed that supports dynamic deployment and O-DU scaling. When central processing unit (CPU) load exceeds 80%, an additional O-DU is instantiated, improving throughput and latency under heavy traffic and demonstrating efficient network resource management.

The demonstration in [126] presents a modular, cloud-native 5G O-RAN, deployed across multiple locations. It enables flexible resource allocation and supports on-demand monitoring applications, facilitating autonomous, and feedback-driven orchestration. The study in [127] evaluates resource scheduling strategies within an O-RAN-compliant 5G network using the ns-3 simulator. By comparing different scheduling mechanisms, the authors demonstrate that the integration of RIC-driven control enhances flexibility and adaptability in scheduling policies.

Further advancing practical implementation, the work in [128] provides a comprehensive guide for developing and deploying O-RAN applications in both simulated and real-world environments. It equips developers with methodologies for architectural evaluation, xApp migration to testbeds, and deployment of key components such as the SMO, Non-RT RIC, and rApps to enable full E2E integration.

III. IN-DEPTH ANALYSIS OF SLICING-AWARE O-RAN ARCHITECTURE

The O-RAN architecture is built on a foundation of open and standardized interfaces, protocols, transport links, NFs, and management functions (MFs). By employing open interfaces and open source software, O-RAN separates the control and user plane, enabling a modular and flexible software stack [129]. Furthermore, the O-RAN Alliance proposes the integration of the AI/ML capabilities to enhance automation and operational efficiency [130]. In this section, we explore the latest O-RAN architecture in a detailed manner, with a particular emphasis on the features that support network slicing. We describe its key components and interfaces, and review network slicing related MFs as defined by 3GPP, ETSI, and ONAP. Moreover, we provide an overview of the underlying infrastructure, including the O-Cloud sites, the open TN, and open cellular network site.
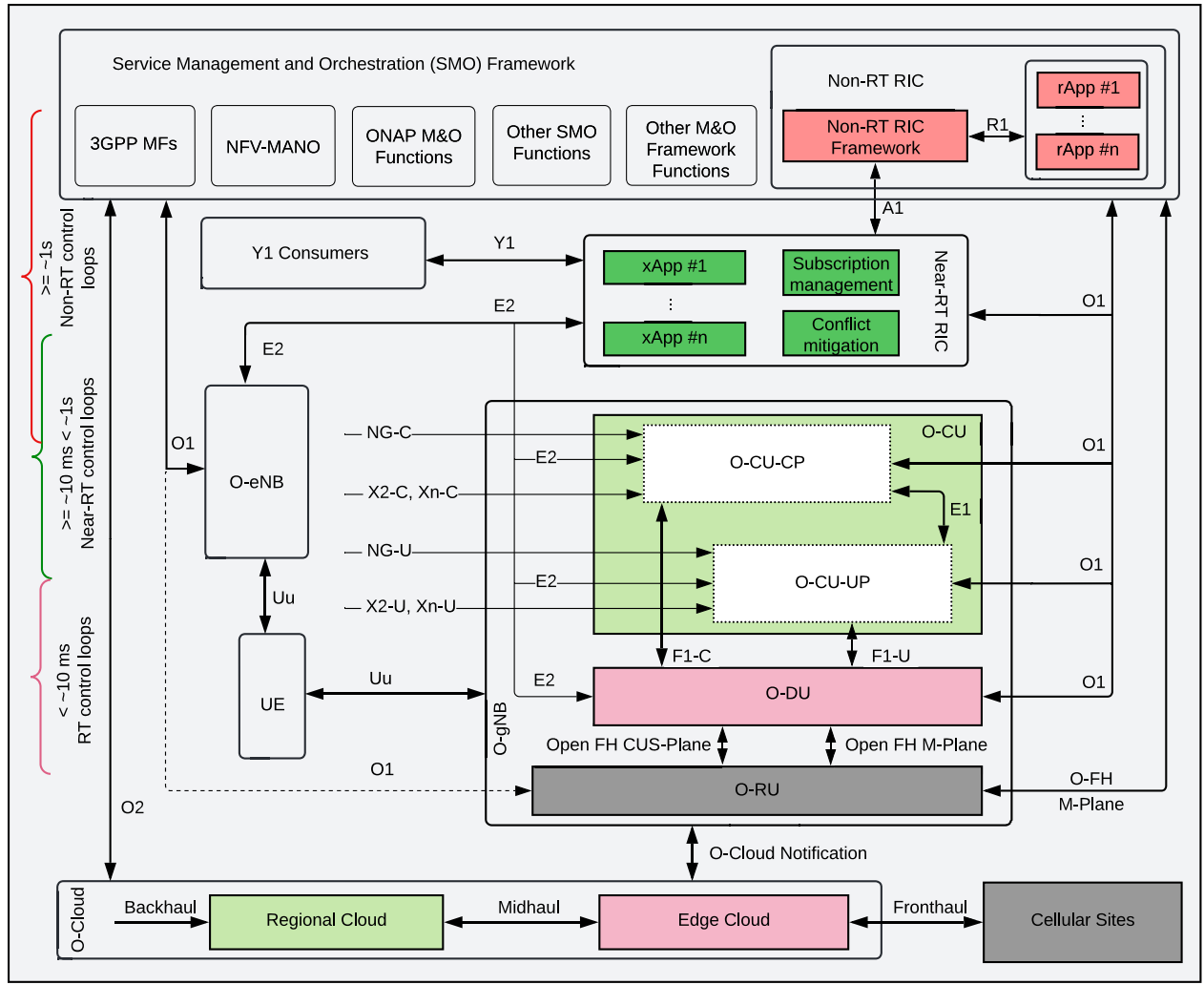
Fig. 2. The latest O-RAN slicing-aware architecture.

A. Major Components of the O-RAN Architecture

The O-RAN architecture follows a disaggregated design paradigm, dividing the cellular base station into multiple logical and physical units, each responsible for specific layers and interfaces of the radio network protocol stack [129]. The O-RAN architecture, as illustrated in Figure 2, consists of four major components: the O-gNB, the RICs, the SMO framework, and the underlying infrastructure.

The O-gNB encompasses the radio functionalities, including tasks such as modulation, coding, resource scheduling, and many others in both uplink and downlink directions. A RIC is a software-defined component within the O-RAN architecture responsible for the control and optimization of O-RAN functions [15], [131]. The SMO framework serves as an automation platform dedicated for the M&O of O-RAN NFs, radio resources and network slices, supporting lifecycle management of O-RAN at scale in an intelligent and autonomous manner [64], [132]. The underlying infrastructure is responsible for hosting the O-RAN components and includes O-Cloud sites, cellular network sites, and transport links.

B. O-gNB (E2 Nodes) and Its Corresponding Interfaces

In the O-RAN architecture, the traditional gNB is systematically disaggregated into distinct logical entities. As illustrated in Figure 2, these nodes include the O-CU, the O-DU, and the O-RU, or a combined O-RAN eNB (O-eNB). Out of these nodes, the O-CU and O-DU are collectively referred to as E2 nodes in O-RAN Alliance terminology [133]. Each E2 node is associated with specific functionalities and standardized open interfaces. This modular decomposition underpins the architectural principles of O-RAN and enables an interoperable next-generation RAN ecosystem. Below, we provide a detailed analysis of each node, together with an in-depth description of its corresponding open interface specifications.

1) O-RAN Centralized Unit: The O-CU is a logical network node responsible for implementing the higher layer protocols of the RAN stack. These include the radio resource control (RRC) layer, which controls the life cycle of radio connections; the service data adaptation protocol (SDAP) layer, which manages the QoS of traffic flows within individual bearers; and the packet data convergence protocol (PDCP) layer, which handles essential functions such as packet reordering,

duplication, and encryption over the air interface [19], [134]. As shown in Figure 2, the O-CU terminates the E2 interface towards the Near-RT RIC and the O1 interface towards the SMO framework [20]. Architecturally, it comprises a single O-CU control plane (O-CU-CP) and potentially multiple O-CU user planes (O-CU-UPs), which communicate over the standardized E1 interface [55]. The O-CU-CP and O-CU-UP interface with the O-DU through the F1-C and F1-U components of the standardized F1 interface, respectively.

According to 3GPP specifications, the O-CU is required to support functionalities associated with network slicing. The O-CU-UP may be deployed either as a dedicated entity for each network slice or shared among multiple slices, depending on the specific requirements and design of each slice [19], [44]. The O-RAN architecture further enhances these capabilities by extending network slicing functionalities beyond those defined in 3GPP. This is achieved through dynamic slice optimization mechanisms facilitated by the Near-RT RIC via the E2 interface [135]. Additionally, the O1 interface supports the configuration of extended slice related parameters to further improve the capabilities of O-RAN architecture. The O-CU is also expected to perform slice-aware resource allocation and enforce isolation mechanisms to ensure compliance with slice specific SLAs. The O-CU is initially configured through the O1 interface based on the requirements of individual slices and is subsequently reconfigured dynamically by the Near-RT RIC through the E2 interface to support evolving slicing use cases. The O-CU may be required to generate and send certain performance metrics (PMs) over both the O1 and E2 interfaces in response to requests from the SMO framework and the Near-RT RIC. These PMs serve as critical inputs for slice performance monitoring and SLA assurance [44], [136].

a) E1 interface: The E1 interface functions as a control interface connecting the O-CU-CP and O-CU-UP entities within an O-gNB [137]. It is standardized by the 3GPP and plays a pivotal role in O-RAN by enabling efficient coordination between control and user plane components of O-CU. The adoption of this standardized interface not only ensures the efficiency of O-RAN but also offers flexibility and scalability for future network evolution and innovation [138].

2) O-RAN Distributed Unit: The O-DU is a logical network node that hosts the lower layer protocols of the RAN stack and serves as a baseband processing unit that handles the high PHY, MAC, and radio link control (RLC) layers [139]. It is typically deployed as a virtualized network function (VNF) that can be hosted within a virtual machine or container at the edge cloud [19]. The O-DU terminates multiple critical interfaces, including the E2 interface (towards the Near-RT RIC), F1 interface (towards the O-CU), and the open fronthaul (O-FH) interface (towards the O-RU). Additionally, it terminates the O1 interface towards the SMO framework to enable the management and orchestration functionalities [20], [44]. It also serves as the aggregation point for multiple O-RUs, terminating the O-FH M-Plane interface to facilitate hierarchical or hybrid management of O-RU within the O-RAN architecture.

In the context of network slicing, the O-DU plays a key role to enable the slice-aware resource management. The MAC layer is responsible to allocate and isolate physical resource blocks (PRBs) per network slice according to the configuration received through the O1 interface, along with O-CU directives over the F1 interface, and dynamic guidance received from the Near-RT RIC through the E2 interface [140], [141]. Similar to the O-CUs, the O-DUs must also generate and report PMs through both the O1 and E2 interfaces, in response to requests from the SMO framework and the Near-RT RIC, respectively. These PM can be used for network slice performance monitoring and SLA assurance [44].

3) O-RAN Radio Unit: The O-RU is a physical node that implements the lower PHY and RF processing functions within an O-gNB, based on the lower layer functional split—specifically, the 7.2x split option [15], [142]. It serves as the termination point for the O-FH interface, as well as the lower PHY functionalities interfacing with UEs. Additionally, the O-RU terminates the O-FH M-Plane interface, which connects to the O-DU and/or the SMO framework, depending on the specific deployment scenario. In O-RAN, a single O-RU is expected to support multiple slice instances to enable radio-level resource sharing [143].

4) O-eNB: The O-RAN architecture also supports the integration of long term evolution (LTE) base stations, referred to as O-eNBs in O-RAN Alliance terminology. An O-eNB may take the form of either a legacy evolved NodeB (eNB) or a next generation eNB (ng-eNB). To ensure compatibility within the O-RAN ecosystem, the associated interfaces and protocols required by these base stations—particulary the E2 and O1 interfaces—must be fully supported [20].

5) E2 Interface: The E2 interface establishes the logical connection between the Near-RT RIC and E2 nodes. It supports two categories of functions: RIC services and E2 support services. RIC services—namely, Report, Insert, Control, Policy, and Query—are enabled through functional procedures, including subscription management, control operations, and information queries. E2 support services facilitate interface and RAN function management through global procedures such as E2 setup, E2 reset, E2 removal, E2 node configuration updates, RIC service updates, and reporting of general error situations [142], [144], [145].

An E2 node comprises a logical E2 agent that terminates the E2 interface and facilitates communication between RIC services and RAN functions. The xApps in Near-RT RIC deliver value-added services leveraging RIC functional procedures over the E2 interface [144], [146]. Furthermore, the E2 interface facilitates the collection of measurements from the RAN to the Near-RT RIC, either periodically or based on predefined trigger events. It supports control and data collection across multiple network dimensions, including individual cells, slices, QoS classes, and specific UEs [144].

Slice-aware xApps utilize the E2 interface to influence the behavior of E2 nodes in a slice specific manner [129]. This includes the configuration of RRM, MAC scheduling policies, and other control parameters across the O-RAN protocol stacks [140]. Moreover, the Near-RT RIC employs the E2 interface to configure and collect slice-specific performance indicators and reports from E2 nodes, supporting real-time

monitoring and closed-loop optimization of network slice performance [44], [147].

6) F1 Interface: The F1 interface connects the O-CU and O-DU within an O-gNB [137]. In O-RAN, it follows the protocol architecture and specifications defined by the 3GPP, complemented by O-RAN-defined interoperability profile specification. Mirroring the O-CU's division into control and user planes, the F1 interface consists of two components: F1-C and F1-U [20].

a) F1-C: The F1-C interface handles control and signaling functions between the O-CU-CP and the O-DU. It facilitates the exchange of control plane information such as connection setup and release, handover management, and radio resource coordination.

b) F1-U: The F1-U interface is responsible for the actual user data transmission between the O-CU-UP and the O-DU. It ensures efficient and reliable transport of user plane data across the O-RAN architecture.

7) X2 Interface: The O-RAN adopts the X2 interface from 3GPP standards to support interoperability profile specifications. It connects the O-CU with other eNBs in an E-UTRAN NR dual connectivity (EN-DC) configuration. The interface is divided into X2-C and X2-U, which handle control plane and user plane information, respectively [20].

8) Xn Interface: The O-RAN adopts the principles and protocol stack of the Xn interface as defined in 3GPP standards, to support interoperability profile specifications. The Xn interface contains two components Xn-C and Xn-U, which connect the O-CU-CP and O-CU-UP, respectively, to other O-gNBs within O-RAN architecture [20].

9) NG Interface: The NG interface, adopted from 3GPP, connects the O-CU to the 5G core (5GC). It comprises two components: NG-C for the control plane and NG-U for the user planes. NG-C connects the O-CU-CP with the access and mobility management function (AMF), while NG-U connects the O-CU-UP to the user plane function (UPF) [20].

10) Uu Interface: The 3GPP defines the interface between the UE and the e/gNB as the Uu interface. The Uu interface encompasses a comprehensive protocol stack spanning from layer 1 to layer 3 and terminates within the NG-RAN architecture. In the decomposed NG-RAN, protocol terminations occur at distinct reference points, none explicitly defined by the O-RAN Alliance. As Uu messages continue to traverse from the UE to the appropriate e/gNB managed function, the O-RAN architecture does not define it as a distinct interface towards a specific managed function [16], [20].

11) Open Fronthaul Interface: The O-RAN FH Specification defines the disaggregation and virtualization of the traditional cellular network site to enhance the efficiency of FH TN in next-generation telecommunication networks. It adopts the 3GPP 7.2x functional split option in the physical layer, partitioning it into high-PHY and low-PHY [148]. The low-PHY functions reside in O-RU, while high-PHY processing is implemented at the O-DU [26], [149]. The intra-PHY lower layer FH split imposes stringent bandwidth and latency requirements, necessitating a dedicated FH service profile for the FH TN [148], [150]. The service profile must be adaptable to diverse deployment scenarios, network topologies, and specific use case requirements. Its framework and latency model align with the reference points outlined in the evolved CPRI (eCPRI) specification. The Open FH interface connects the O-DU and O-RU via the CUS-Plane—comprising control, user, and synchronization—and M-Plane, which supports M&O operations [151], [152]. Furthermore, it supports centralized control of the O-RU by the O-DU and allows, in certain configurations, a single O-DU to manage multiple O-RUs, including those operating across different carrier networks [153].

a) C-Plane: The control plane (C-Plane) in the O-RAN architecture refers to the real-time control signaling exchanged between the O-DU and O-RU. C-Plane messages convey data-associated control information essential for user data processing, including scheduling and beamforming [153] instructions—particularly when such information is not provided via the M-Plane. These messages are transmitted independently for both downlink and uplink directions. To enhance flexibility, C-Plane messages can be transmitted either collectively or individually, depending on the associated channel and specific transmission requirements [152].

b) U-Plane: The user plane (U-Plane) handles the transmission of in-phase and quadrature (IQ) sample data between the O-DU and O-RU [150]. To ensure proper coordination with the C-Plane, the FH interface mandates that C-Plane messages arrive at the O-RU prior to the latest permissible time for the corresponding U-Plane messages. U-Plane data is encapsulated in two-layer header structure: the first layer contains an eCPRI or IEEE 1914.3 header indicating the message type, while the second layer constitutes an application specific header with fields necessary for control and synchronization [149], [152].

c) S-Plane: The synchronization plane (S-Plane) ensures coordination between the O-RU, O-DU and a synchronization controller, typically an IEEE 1588 Grand Master, which may be integrated into the O-DU. The O-RAN supports E2E synchronization of frequency, phase, and time across all relevant network elements—including O-DUs, intermediate switches, and O-RUs—to meet the requirements of both the time division duplexing (TDD) and frequency division duplexing (FDD) operations [152].

d) M-Plane: The management plane (M-Plane) in O-RAN handles non-real-time management operations between the O-DU and the O-RU. Depending on the TN topology, various connectivity models may exist between the O-RU, O-DU, and the SMO framework [151], [153]. The primary requirement of the M-Plane is to ensure E2E connectivity between the O-RU and entities responsible for its management, which may include the O-DU, the SMO, or designated O-RU controllers [151].

The Open FH M-Plane utilizes a NETCONF/YANG-based framework to manage the O-RU, supporting functions such as installation, configuration, software update, performance monitoring, fault, and file management. Two architectural models are defined: First, the hierarchical model, where one or more O-DUs manage the O-RU via a NETCONF interface. Second, the hybrid model, which allows additional direct logical interfaces between the SMO framework and the O-RU,

alongside the existing O-DU–O-RU link [151], [153]. In the hybrid model, the O-RU can establish E2E connectivity with the SMO either directly or through the O-DU. Notably, there is no explicit signaling mechanism to indicate whether a hierarchical or hybrid model is in use. All NETCONF servers compliant with the M-Plane specification must support multiple concurrent sessions, and all the O-RUs expected to be compatible with both architectural models [151].

C. RAN Intelligent Controller

The RIC represents a significant advancement within the O-RAN architecture, introducing a centralized abstraction layer that facilitates greater control and flexibility in RAN operations [154], [155], [156]. As an integral architectural innovation, the RIC empowers MNOs to design and deploy custom control plane functionalities, thereby enhancing the agility, efficiency and programmability of the RAN [155]. It manifests as a software-defined NF designed to manage specific control functionalities, such as mobility management and RRM, which have traditionally been confined to the base stations.

By offering real-time visibility and centralized control over RAN resources, the RIC plays a pivotal role in advancing the O-RAN disaggregation paradigm [25]. It enables critical capabilities including multi-vendor interoperability, intelligent decision making, and dynamic resource allocation—that collectively redefine the operational landscape of O-RAN [131], [154], [155], [156], [157], [158]. Its integration into the O-RAN architecture facilitates intelligent M&O, particularly in the implementation of key concepts such as network slicing [159].

Moreover, the RIC is responsible for configuring network slices, orchestrating NFs, monitoring network performance, and conducting real-time optimizations of RAN resources via open interfaces [158], [160]. As illustrated in Figure 2, the RIC is implemented in two distinct forms—Non-RT RIC and Near-RT RIC—each designed to address specific control loop dynamics and latency requirements [149]. The following sections provide a detailed examination of both variants, elaborating on their functionalities, applications, and strategic importance in the broader context of RAN optimization and automation.

1) Near-Real-Time RAN Intelligent Controller: The Near-RT RIC serves as a logical entity in O-RAN, facilitating precise and close-to-real-time control and optimization of E2 nodes and their resources [161]. It is positioned in close proximity to the E2 nodes and interacts directly with them to enhance performance by leveraging continuous data collection and executing real-time control actions via the E2 interface [159], [162]. Operating within a control loop bounded by latencies between 10 milliseconds and 1 second, it ensures timely responsiveness to network dynamics [146].

The Near-RT RIC serves as a software platform for hosting xApps—modular, microservice-based applications that are intelligent, autonomous, and tailored for specific control functions [163], [164]. These xApps are deployed to the Near-RT RIC as needed to offer targeted functionalities, such as

intelligent RRM [41], [66], [126]. During the onboarding process, xApps define their data dependencies, including the types of data they collect, process, consume, and expose [20], [163]. The integration of xApps within the Near-RT RIC enables dynamic management and optimization of RAN resources, effectively addressing the diverse service requirements of modern cellular networks [27], [126]. By leveraging UE and cell-specific metrics collected through the E2 interface, xApps facilitate real-time optimization of RAN resources and functionalities [164]. This capability ensures efficient resource utilization and contributes to improved user experience in high-demand network environments [133], [165].

Furthermore, the Near-RT RIC gains direct control over E2 nodes and their resources by policies and information delivered through A1 interface from the Non-RT RIC [146]. In specific scenarios, it is authorized to monitor, suspend, override, or control E2 nodes and their resources on rules defined within the E2 service model [20], [145]. Additionally, the Near-RT RIC exposes E2-related application programming interfaces (APIs) that support access to E2 functions, xApp subscription management, and conflict mitigation mechanisms [144], [146].

A critical function of the Near-RT RIC lies in network slicing, where it supports near-real-time optimization of RAN slice subnet by coordinating with O-CU and O-DU components through the E2 interface. To accomplish this, xApps must be slice-aware, allowing them to implement algorithms to meet each slice SLA [44]. Addressing this challenge, a growing body of literature presents various xApps designed for slice management and optimization tasks in O-RAN [165], [166].

The xApps employ AI/ML-based models, guided by A1 policies generated by the Non-RT RIC [31], [70], to make intelligent context-aware decisions. Once a slice is active slice-specific PMs are collected from E2 nodes and integrated with slice configuration data at the Near-RT RIC. It facilitates dynamic and automated slice optimization. The collaborative intelligence between the Non-RT RIC and Near-RT RIC, powered by AI/ML framework, plays a pivotal role in maximizing the overall efficiency and scalability of network slicing within the O-RAN architecture [5], [146].

a) Y1 interface: The Near-RT RIC provides RAN analytics information services to an authorized third party, known as Y1 consumer, via the Y1 service interface [167]. Access to these services is granted following mutual authentication and authorization. Within a public land mobile network (PLMN) trusted domain, Y1 consumers can subscribe to or request analytics data through this interface. Entities outside the PLMN trust domain may also access Y1 services securely via a standardized exposure function. Notably, Y1 consumers are external entities and are not represented as logical O-RAN functions within the architecture, as shown in Figure 2 [20].

2) Non-Real-Time RAN Intelligent Controller: The Non-RT RIC is a core component of O-RAN architecture, responsible for non-real-time management and optimization of the RAN components and underlying resources [133]. As an integral part of the SMO framework, it communicates with the Near-RT RIC over the A1 interface and orchestrates AI/ML workflows [168]. As illustrated in Figure 2, it also facilitates the execution of third-party applications known as rApps.

These modular applications leverage the R1 interface of the Non-RT RIC framework to deliver value-added services such as managing policy-driven suggestions via the A1 interface and enabling control actions for potential implementations through the O1 and O2 interfaces [20], [67].

The Non-RT RIC facilitates the autonomous configuration of O-RAN components, minimizing the need for manual operator intervention. It provides MNOs with insights into network behavior and supports high-level optimization strategies [162]. It performs data analytics and AI/ML model training leveraging SMO provided services such as data collection and provisioning from RAN nodes [147], [168]. Once trained, these models are distributed to the Near-RT RIC for real-time inference and execution.

The Non-RT RIC plays a critical role in network slicing, providing advanced orchestration capabilities. It collects slice-specific PM, configuration parameters, and optional internal metrics from DCs to support AI/ML driven optimization [169]. Trained AI/ML models enable non-real-time optimization of slice-specific parameters over the O1 interface addressing complex tasks such as RRM policy enforcement [41], [133].

The gathered data is also forwarded to the Near-RT RIC, which utilizes it for dynamic slice optimization to mitigate potential SLA violations across network slice instances [36]. While the Near-RT RIC controls the E2 resources via the E2 interface, the Non-RT RIC manages the cloud resources through the O2 interface, with decisions informed by the collected analytics and trained AI/ML models [170].

a) R1 interface: The R1 interface located within the internal structure of the Non-RT RIC, provides access to the framework services that empower rApps retrieve data for initiating intelligent policy generation and RAN optimization [171]. It also supports authorized enrichment data exchange with the Near-RT RIC and allows rApps to share services and analytics within the Non-RT RIC framework [170], [172].

3) A1 Interface: The A1 interface establishes communication between the Non-RT RIC and the Near-RT RIC [20], [171]. It enables the Non-RT RIC to provide policy guidance, known as A1 policies, to the Near-RT RIC [56]. These policies support functions such as provisioning directives for an individual or groups of UEs, monitoring policy states, providing feedback, and exchanging enrichment information required for RIC operation [56], [173]. The A1 interface also facilitates coordination for AI/ML workflows, including model training, distribution, and inference.

In network slicing scenarios, A1 services are essential for SLA assurance. For example, the Non-RT RIC can use A1-based policy management to transmit slice-specific policies that guide the Near-RT RIC in resource allocation and slice-specific control actions, while also receiving feedback on policy compliance and performance [44].

D. The SMO Framework and Its Corresponding Interfaces

O-RAN is designed to deliver flexibility, reliability, scalability, and interoperability across multi-vendor environments. It operates on COTS hardware within a cloud-native, virtualized infrastructure, and relies on automation, and AI/ML to support intelligent and adaptive network management [174], [175]. The SMO framework forms the foundation of autonomous and intelligent M&O within O-RAN [155]. As shown in Figure 2, it integrates a suit of MFs and services tailored to O-RAN. The SMO incorporates management capabilities that are defined by multiple SDOs and ensures interoperability among their MFs through standardized service-based management interfaces [47]. Operating on a services based architecture (SBA), the SMO framework enables seamless provisioning and consumption of key services, including authentication, authorization, service registration and discovery, data management, and trained model distribution sharing [20], [176].

The SMO manages FCAPS operations through the O1 interface, facilitates intelligent RRM via the Non-RT RIC, and orchestrates O-Cloud infrastructure. It should support O-Cloud orchestration by integrating VNF orchestration and federated O-Cloud orchestration and management (FOCOM) via the O2 interface [176]. It also handles workload management and resource provisioning. The SMO communicates with the O-RU for FCAPS-related functions via the O-FH M-Plane interface [44]. The A1 interface connects the Non-RT RIC and Near-RT RIC, enabling the former to collect data, train or select ML models, and transmit them to the latter for real-time execution.

The architecture of SMO—particularly the Non-RT RIC—offers flexible implementation options, enabling the operators to selectively adopt the desired features. The SMO can integrate with an E2E multi-domain service orchestrator, coordinating domain-specific modules across the RAN, TN, and CN. It facilitates on-demand creation and management of network slice instances (NSIs) over distributed 5G infrastructure [176].

To support network slicing, the SMO must adhere to the architectural requirements defined by 3GPP, ETSI, and ONAP, utilizing standardized MFs that align with their respective specifications. These MFs of perform tasks such as slice creation, operation, modification, termination as well as scaling the underlying resources. The O-RAN Alliance maintains alignment with network slicing concepts and architectural principles established by the aforementioned SDOs [74] while extending them with general guidelines tailored to O-RAN's modular and open architecture. For example, O-RAN mandates compatibility with 3GPP interface specifications [156], standardized management service interfaces for slicing operations, support for multi-vendor interoperability, flexibility in deployment options, and support for multi-operator slice subnet management [44], [171].

1) O1 Interface: The O-RAN managed elements and the management entities within the SMO framework are logically connected through the O1 interface, as shown in Figure 2. The O1 interface facilitates essential operations and management tasks for O-RAN components, including FCAPS, software and file management, among many others. The key components such as the O-CU, O-DU, and Near-RT RIC are managed through this interface, enabling the SMO to access and control relevant O-RAN NFs [177].

In the context of O-RAN slicing, the O1 interface supports the configuration of O-RAN nodes with slice-specific parameters tailored to the service requirements of individual network slices. The 3GPP defines a slice-specific information model, which includes RRM policy attributes—for example, the distribution of PRBs across slices [40]. These models can be extended to include slice profiles and additional configuration parameters to support O-RAN slicing use cases over the O1 interface. Moreover, the interface enables the collection of slice-specific performance metrics and fault reports from O-RAN nodes , supporting SLA assurance and closed-loop automation within the slicing framework [44].

2) O2 Interface: The O2 interface is an open, logical interface that facilitates secure communication between the SMO framework and distributed O-Cloud sites [178]. It supports the lifecycle management of VNFs that operates on the O-Cloud infrastructure. Within the O-RAN, the O-Cloud hosts essential NFs, while the O2 interface allows the SMO to coordinate O-Cloud infrastructure management and deployment activities.

In addition to lifecycle management of O-Cloud infrastructure, the O2 interface supports the orchestration of resource management such as inventory, monitoring, provisioning, and software management. It provides logical services that govern the deployment of O-RAN NFs on cloud resources. The O2 interface is inherently extensible, allowing new features to be incorporated without altering the underlying protocol or management processes. It is vendor-neutral and is unaffected by particular SMO framework and O-Cloud implementations. Through the O1 and/or O2 interfaces, operators can dynamically manage, reconfigure, and upgrade network components hosted within the O-Cloud environment [179].

3) 3GPP Network Slicing MFs Within the SMO Framework: The 3GPP-defined MFs for the M&O of network slicing include the communication service management function (CSMF), network slice management function (NSMF), network slice subnet management function (NSSMF), and network function management function (NFMF) [38]. These MFs can be integrated within the SMO framework in accordance with the requirements outlined in the 3GPP specifications. Additionally, the O-RAN Alliance has defined both functional and non-functional requirements for network slicing architecture, as detailed in [44]. The requirements specified by both 3GPP and the O-RAN Alliance are critical for the successful realization of network slicing within the O-RAN architecture and for ensuring the effective operation of the MFs. The provision of management services (MnSs) for mobile networks—including network slicing—can be achieved through a set of functional blocks, as illustrated in Figure 3.

a) Communication service management function (CSMF): It is responsible for translating communication service requirements—received from third parties such as operations support system (OSS)/business support system (BSS), network slice as a service (NSaaS) tenants, and other external entities—into network slicing requirements for an E2E network slice [180], [181].

b) Network slice management function (NSMF): The NSMF manages an E2E network slice based on the
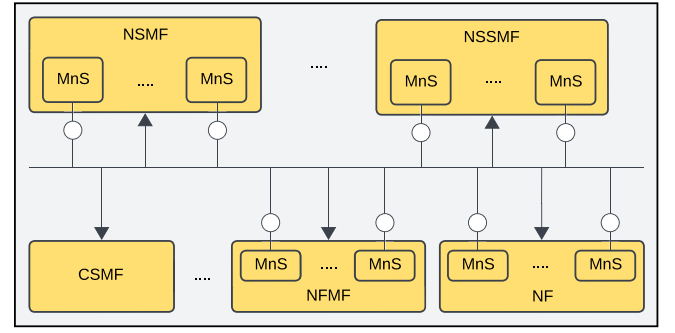

Fig. 3. 3GPP management architecture within the SMO framework.

requirements determined by the CSMF. It is responsible for coordinating and managing the necessary resources to support the associated communication services and interfaces with the NSSMFs responsible to manage individual network slice subnets within that domain (e.g., RAN NSSMF) [180], [181].

c) Network slice subnet management function (NSSMF): The NSSMF operates within individual network domains (i.e., RAN, TN, and CN) and is responsible to instantiate the required resources based on the instructions provided by the NSMF. Each NSSMF orchestrates the domain-specific resources within its respective subnet to fulfill the service requirements assigned to that domain. For example, the RAN NSSMF orchestrates the RAN segment of a network slice by performing life-cycle management, configuration management, performance monitoring, and fault management. Additionally, the RAN NSSMF interfaces with the RIC to execute control plane operations related to O-RAN slices within the O-RAN architecture [180], [181].

d) Network function management function (NFMF): The NFMF offers NF management services within the NF management framework. It is responsible for managing multiple NFs, including application-level management of both VNF and physical network function (PNF) [47]. Additionally, the NFMF produces NF provisioning services, which includes configuration, fault, and performance management. At the same time, it also acts as a consumer of the NF provisioning service exposed by individual VNFs and PNFs [180].

To provide comprehensive M&O solutions, the SMO framework can be customized to include either all or a subset of the aforementioned 3GPP-defined MFs. The selection and integration of these MFs depend on deployment-specific considerations, which are further discuss in Section IV.

In addition, the SMO framework can be extended to incorporate management functions from both the network functions virtualization management and orchestrierung (NFV-MANO) and ONAP [32]. Within this context, the NFV-MANO is responsible for the management and orchestration of VNFs as well as the virtualized resources associated with an O-RAN network slice. Readers seeking detailed architectural information are encouraged to refer to the [182], [183].

Furthermore, ONAP is currently being integrated by the OSC, with selected SMO functionalities leveraging and extending components of ONAP's existing infrastructure [131]. This integration enables faster automation of

new services and comprehensive lifecycle management—capabilities that are essential for 5G and next-generation network—through real-time, policy-driven orchestration and automation of both PNFs and VNFs [184]. Further details regarding the roles of NFV-MANO and ONAP, along with their deployment scenarios within the SMO framework, are provided in Section IV.

E. The Underlying O-Cloud and O-Transport Infrastructure

The underlying infrastructure of O-RAN comprises the O-Cloud sites (which include the regional cloud and edge cloud sites), the cellular network sites, and the TNs. The regional and edge cloud sites along with cellular network sites, provide necessary hosting environment for essential O-RAN NFs. The TN ensures connectivity between various virtual or physical NF deployed across cellular and/or cloud sites. In this section, we provide a concise overview of these key infrastructure components within the O-RAN architecture.

1) Cellular Network Site: A cellular network site refers to the physical locations where O-RUs are deployed, which may be mounted on the same structure as the antenna or situated at the base of the installation. Typically, a cellular site is engineered to support multiple sectors, enabling the deployment of several O-RUs within a single site. They facilitate the exchange of user data, control plane signaling, and synchronization information between the O-RU and the O-DU. The distribution of cellular sites may follow uniform or non-uniform pattern, depending on factors such as user density, geographical terrain, and network topology. Based on their coverage area and transmit power, cellular sites are typically categorized as Macro, Micro, Pico, and Nano types.

2) Cloud Site: A cloud site refers to a physical location equipped with cloud infrastructure resources, suitable for O-Clouds, and possibly accommodating other non-O-Cloud resources. Within the O-RAN architecture, O-Clouds are strategically deployed at both regional and edge cloud sites. These sites serve as centralized platforms for hosting VNFs, SDN controllers, and other cloud-native applications. The regional cloud sites typically provide broader geographical coverage and greater computational capacity, while the edge cloud sites bring compute and storage resources closer to the network edge. It enables low-latency and high-bandwidth support for a wide range of services and applications.

a) Edge cloud: Edge cloud refers to a site that hosts virtualized RAN functions, particularly those required to support multiple nearby cellular network sites. It provides centralized processing capabilities for these sites, enabling efficient coordination and management. The physical coverage of an edge cloud can vary depending on the operator's deployment strategy and use case—it may serve either broad region or a more localized area. Regardless of scale, the edge cloud must maintain sufficient proximity to the associated O-RUs to satisfy the stringent latency requirements of O-DU functions. This proximity ensures low-latency communication between O-RUs and O-DUs, thereby enabling efficient network operations and timely service delivery.

b) Regional cloud: A regional cloud refers to a cloud site that supports virtualized RAN functions for a large number of cellular network sites, typically spanning multiple edge clouds. It enables a higher degree of centralization by hosting the key functions, including O-CU and the Near-RT RIC. To fulfill the latency requirements of hosted functions, the regional cloud site must be located sufficiently close to the associated O-DUs. This proximity ensures that communication between the O-CU, O-DU, and Near-RT RIC remains within acceptable latency bounds. It enables effective coordination and optimization of RAN resources across a wide geographical area, while maintaining high level of performance, scalability, and responsiveness.

3) O-RAN Cloud Platform: The O-RAN cloud platform (O-Cloud) refers to a cloud computing environment composed of physical network infrastructures. It supports the deployment of critical O-RAN components such as the SMO, Near-RT RIC, and E2 nodes, along with the associated software and required M&O services. Each O-Cloud is composed of a group of CPUs, random access memory (RAM), storage, network interface cards (NICs), basic input and output system (BIOS), baseboard management controllers (BMCs), and hardware accelerators. These elements work together to handle computationally intensive tasks across the platform [20].

Depending on the deployment scenario, the O-Cloud can virtualize a variety of NFs and take on the execution of RAN functionalities within the overall O-RAN architecture. A more detailed discussion on these aspects can be found in Section V.

4) O-Cloud Notification API: The O-Cloud notification API facilitates event subscription for consumers such as the O-DU, which operates within the O-Cloud environment. This API allows consumers to subscribe to receive event notifications and status updates from the O-Cloud. Additionally, the O-Cloud exposes event producers, making it possible for cloud workloads to access relevant notifications and statuses that would otherwise remain internal to the platform [20], [172].

5) Transport Network: In disaggregated O-RAN deployments, components such as the O-CU and O-DU may be deployed on separate, geographically distributed O-Cloud sites. To support seamless communication among O-CU, O-DU and O-RU, a robust networking infrastructure must interconnect these elements across the cellular network site and distributed O-Cloud sites through open and highly reliable TNs [185].

The TN encompasses multiple segments, including FH, MH, and BH. It supports both the NR and legacy technologies such as LTE and universal mobile telecommunications system (UMTS). The transport services span the C-Plane, U-Plane, S-Plane, and M-Plane, and are designed to support the operational requirements of diverse operators and support various E2E services, including URLLC and eMBB.

The TN must be highly flexible to support various use cases, applications, and heterogeneous RAN architecture. Each segment of the physical TN may need to concurrently support multiple network slice instances, distinct 5G services, and various 3GPP interfaces, tailored to specific deployment scenarios and performance requirements.

a) Fronthaul: In O-RAN, FH refers to the communication link between the O-DU and O-RU within the RAN infrastructure. It encompasses control, user, synchronization, and management planes. To meet the stringent latency requirements associated with FH, O-RUs and their corresponding O-DUs are deployed in close physical proximity [186].

b) Midhaul: The MH network is a logical segment of the TN that facilitates communication between the O-DU and O-CU, enabling the transport of 3GPP F1/W1/E1 interfaces [20]. When the O-DU and O-CU are deployed as a unified entity, these interfaces are internal and not exposed, effectively eliminating the need for a distinct MH segment. It also provides inter-O-CU communication, specifically supporting the transport of the 3GPP Xn interface. In deployments where MNOs have not adopted a split architecture between the O-DU and O-CU, these interface functions are instead handled within the BH network.

c) Backhaul: In O-RAN architecture, the BH network connects the O-CU to the 5G CN. It comprises both CP and UP components to ensure a clear separation between user data and control signaling defined by 3GPP. The CP includes multipoint interfaces such as N1, N2, N4 and Xn-c, which facilitates communication between the O-CU-CP, the UPF and other 5G CN components. The UP includes interfaces such as N3 (between O-CU-UP and UPF), N9 (between UPF instances), and Xn-u (between O-CU-UP nodes), supporting efficient transmission of user data.

IV. O-RAN NFs, Network Slicing, and SMO Deployment Options

The O-RAN NFs can be implemented as VNFs or PNFs. They can be hosted by the underlying O-Clouds infrastructure and cellular sites, respectively. Regardless of whether the NFs are virtualized or physical, they must be strategically allocated to suitable hosts within the O-RAN infrastructure. This mapping process plays a key role in implementing the logical network functionalities of an O-RAN slice, particularly in cloud computing environments, where resource optimization and performance considerations are critical.

Deployments can vary from fully distributed to highly centralized configurations depending on the placement of NFs on edge and regional O-Cloud sites—referred to as point-of-presence (PoP) in ETSI terminology. The decision entails determining the optimal execution location for each logical function, with potential impacts on performance, scalability, cost, and other crucial factors [187]. In this regard, the O-RAN Alliance has introduced the O-Cloud architecture and outlined several deployment scenarios for O-RAN NFs within the cloud-native framework [178]. In addition, the document highlights numerous considerations essential for the effective deployment of logical NFs across multiple O-Clouds environment. The diverse slicing and NF deployment options within O-RAN require a range of M&O solutions, leading to the multiple deployment options for the SMO framework.

In the following subsections, we explore various deployment alternatives for O-RAN NFs, aligning them with the underlying infrastructure. Additionally, we analyzed different network slicing deployment strategies within O-RAN. Furthermore, we highlight multiple deployment options for the SMO framework, emphasizing the significance of network slice MFs.

A. O-RAN NF Deployment Scenario

The O-RAN Alliance has explored various approaches for virtualizing the O-RAN NFs in regional and edge clouds proposing different deployment scenarios aligned with O-RAN specifications. These scenarios are characterized by the specific grouping of functionalities across key locations such as cellular sites, edge clouds, and regional clouds [188]. Additionally, they are distinguished by whether the functionality at a particular location is implemented by an O-RAN PNF-based solution—where software and hardware are tightly integrated and share a unified identity—or through cloud-based services.

Figure 4 illustrates several NF deployment scenarios as presented in [178], [189]. At the top, it provides an overview of the NFs, while each scenario exhibits how these NFs are deployed—either as cloudified NFs within O-Cloud or as PNFs at the cellular site. A detailed explanation of each deployment scenario is provided in the following.

1) Scenario A: In this scenario, the Near-RT RIC, O-CU, and O-DU are deployed at the edge cloud as VNFs, whereas the O-RUs are deployed on cellular network sites. This scenario is ideal for dense urban deployments with ample FH capacity, enabling the pooling of BBU functionalities at a central location. It reduces latency but comes with potentially higher deployment costs compared to other scenarios.

2) Scenario B: In this deployment scenario, the O-CU and O-DU are deployed at the edge cloud site in order to reduce latency, while the Near-RT RIC is deployed at the regional cloud site in order to gain a wider network perspective for performance optimization and improvement.

3) Scenario C: In this deployment scenario, the O-CU is co-located with the Near-RT RIC in the regional cloud site, and the O-DU is positioned at the edge cloud site. This scenario is tailored to support deployments in areas with limited remote Open FH capacity, imposing restrictions on the number of O-RUs. Two additional variations, C.1 and C.2, have been introduced to address the specific requirements of certain network slice instances [178], [188].

4) Scenario D: This deployment scenario is akin to Scenario C (see above). However, the O-DU is deployed as PNF at the edge O-Cloud site in this scenario.

5) Scenario E: This deployment scenario mirrors Scenario D, with the key distinction that all components, including both O-DU and O-RU, are fully virtualized within the same edge cloud site. This approach is being considered for future use, acknowledging that the virtualized versions of the low-PHY layer and other O-RU aspects are not currently available.

6) Scenario F: This deployment scenario involves the virtualization of both O-DU and O-RU, but they are hosted on separate O-Cloud sites. Like Scenario E, this scenario is also considered for future use for the similar reason.

Within the context of O-RAN deployment, as discussed in the above scenarios, the Open FH plays a pivotal role in the definition of the interface between VNFs deployed within the
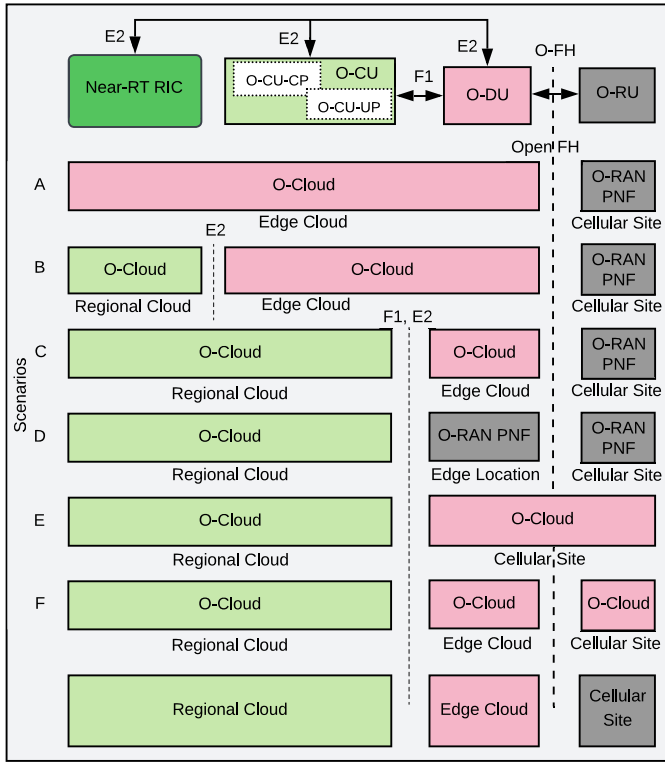
Fig. 4.   O-RAN NFs deployment scenarios onto the underlying O-Cloud sites and cellular network sites.

O-Cloud and the cellular sites. O-RUs are always located at the cellular site, while O-DU can reside at the edge cloud site. To meet strict latency requirements, O-DU placement can be adjusted closer to the cellular site; however, extending it farther from the cellular sites may violate RAN internal or RAN service-specific timing constraints [189].

A common deployment scenario involves moving O-DU instances toward or even directly to the cellulari site alongside O-RU, particularly when the edge cloud site must be closer to the cellular site due to fiber availability or other constraints. However, such adjustments may compromise the benefits of centralization and resource pooling [149].

The placement of O-CU and its associated UPF is determined by the latency requirements of the F1 interface or specific service constraints. For example, O-CU-UP and UPF for URLLC services must remain at the edge cloud site. whereas for eMBB, deployment at regional cloud site is feasible. Additionally, for services without specific latency requirements, O-CU-UP and UPF can be placed in the core cloud site [178]. Centralizing O-DU is particularly beneficial in densely populated networks where multiple cellular sites remain within the latency limits between O-RU and O-DU. Conversely, in sparsely populated areas, centralizing only the O-CU is often more practical in O-RAN deployment.

B. Network Slicing Deployment Options in O-RAN

Network slicing is a foundational concept in next-generation network architectures, centered around the creation of logical E2E virtual connections between end users or vertical customers and their target applications and services [80], [82].

This is achieved through the strategic allocation of network resources to ensure that each service or application receives the necessary support to satisfy its specific QoS requirements and meet predefined SLA, thereby enabling reliable and differentiated service delivery [83], [190], [191], [192].

The architecture of network slicing is structured into three distinct layers: the infrastructure layer (IL), network function layer (NFL), and service layer (SL) [79], [193].

1) Infrastructure Layer: The IL encompasses the entire physical network infrastructure, comprising RAN, CN, and TN components. It is responsible for the deployment, control, and management of network infrastructure, as well as allocation of computing, storage, network, and radio resources to network slices. Additionally, it manages how these resources are exposed to upper architectural layers.

2) Network Function Layer: The NFL manages the lifecycle of NFs including both physical and virtual. These functions are deployed over virtualized infrastructure and interconnected to deliver E2E service that adheres to specific constraints defined in the service design of a network slice.

3) Service Layer: The SL focuses on service definition and orchestration. It maps service descriptions onto the underlying network infrastructure and encompasses the functional design of slice management and orchestration entities. The SL plays a pivotal role in translating service demands into actionable configurations for lower layers, to ensure efficient and scalable network slicing operations [12].

The deployment of slicing within O-RAN is facilitated by decoupling software and hardware components integrated with NFV [192]. Determining the allocation of specific logical functions to appropriate O-Cloud platforms, and identifying which functions should be co-located, is essential for the implementation of slicing in O-RAN architecture [147]. Certain O-RAN components such as the Near-RT RIC, O-CU-CP, O-DU, and O-RU, are designed to be shared across multiple network slices. In contrast the O-CU-UP is typically dedicated to individual slices to ensure isolation and performance.

Moreover, the strategy for mapping NF to either shared or separate cloud platforms must align with the specific service requirements and deployment constraints of each use case [36]. One of the potential deployment models proposed by the O-RAN Alliance for slicing is illustrated in Figure 5, where the O-RU is deployed as a PNF at a cellular site. The Near-RT RIC is virtualized at a regional cloud site, while both the O-CU and O-DU are virtualized on a location-independent edge cloud platform. The O-CU/O-DU are connected to the Near-RT RIC with the E2 interface, and the O-CU communicates with the O-DU through the F1 interface [44].

The deployment scenario depicted in Figure 5 can be realized through multiple configurations by virtualizing the shared components such as Near-RT RIC, O-CU, and O-DU across the regional and edge cloud platform as shown in Figure 4. For example, instead of sharing a common O-DU for all slice instances, a dedicated O-DU may be created for each slice instance in the scenario illustrated in Figure 5.

It is important to recognize that while the requirements for PNFs, cloudified network services, and O-Cloud platforms may vary. However, the logical network function requirements
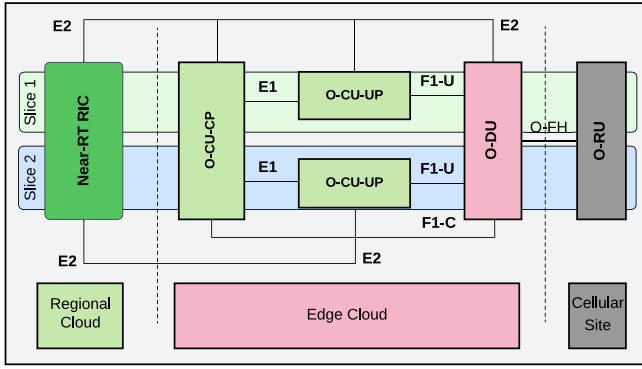
Fig. 5.   O-RAN reference slicing deployment option.

Fig. 6.   The 3GPP and NFV-MANO-based SMO deployment option with a particular emphasis on O-RAN slicing.

always remain the same [44], [178]. For Example, in the scenario illustrated, a single O-CU-CP instance manages the control plane operations for both network slices, whereas each slice is assigned a dedicated O-CU-UP instance. If the UE is connected to both slices, only one RRC connection is established, handling handover and cell assignments through the shared O-CU-CP. However, each service associated with a different NSI can benefits personalized QoS management and independent flow control through an individual SDAP/PDCP stacks within their respective O-CU-UP [44].

For the management and orchestration of slicing, the O-RAN slicing-aware architecture leverages the SMO framework, which incorporates a dedicated slice MF block. This block contains 3GPP-defined NSMF, NSSMF, and NFMF. Additionally, it may integrate additional MFs specified by the ETSI Industry Specification Group (ISG) NFV or derived from the ONAP. The following section provides a comprehensive analysis of different deployment options for the SMO framework in the context of network slicing management.

C. SMO Framework Deployment Options

As discussed in Section III, the SMO framework is responsible for the M&O of O-RAN components and resources. The SMO framework may comprise management components and systems developed by various SDOs. To date, several SMO solutions emerged in the market, claiming compliance with the latest specifications of the O-RAN Alliance. However, these frameworks often lack transparency, as their internal architectural and operational mechanisms are not publicly disclosed, thereby constraining comprehensive insight into their functional capabilities [21].

To promote openness and interoperability, the O-RAN Alliance has proposed two open source solutions: the ONAP and the NFV-MANO framework. ONAP serves as a comprehensive platform for the M&O of virtualized and software-defined elements within O-RAN architecture. Its affiliation with the LF enables integration with other key projects such as Kubernetes, Akraino, Acumos, and OpenDaylight [184]. ONAP is also the preferred SMO platform used by OSC in their open source O-RAN code releases.

Alternatively, the open source MANO (OSM), developed under the NFV-MANO framework, by the ETSI offers comparable SMO functionalities in a more lightweight design
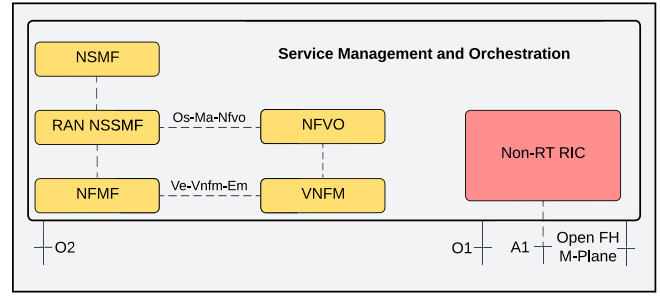
compared to ONAP. Notably, in May 2021, ETSI entered into a cooperation agreement with the O-RAN Alliance, marking the beginning of collaborative efforts to integrate the OSM framework within the O-RAN architecture [194].

In the remainder of this subsection, we examine the deployment scenarios of NFV-MANO and ONAP, along with their potential impacts on the network slicing architecture, as elaborated in [44].

1) 3GPP and NFV-MANO-Based SMO Deployment: The deployment options of the SMO framework—aligned with both the 3GPP management system and the NFV-MANO framework—emphasize the core principles and essential requirements of network slicing. These include the virtualization and softwarization of RAN resources and components, as well as the seamless integration of AI/ML capabilities and programmable control within the SMO framework [169], [195].

This deployment option integrates the slice MFs and network MFs defined by 3GPP with the functional blocks defined by the ETSI ISG NFV. The NFV-MANO framework is responsible for the M&O of VNFs, including processes such as the automation, monitoring, and operation of virtualized functions deployed over a multi-tenant and virtualized infrastructure. Figure 6 illustrates this deployment scenario, where 3GPP-defined slice MFs (such as the NSMF, NSSMF, and NFMF) are integrated with the functional blocks defined within the NFV-MANO. The detailed functionalities of the 3GPP-defined MFs are discussed in Section III.

Moreover, the study group under O-RAN WG 1 has identified four different options for the deployment of SMO framework. These options vary exclusively in the placement of the 3GPP-defined slice MFs and are designed to support different network slice management topologies. Each deployment option explore the potential implications for the O-RAN slicing-aware architecture. The four possible deployment options for the SMO framework are explained as follows:

- Deployment Option 1: In this option, the network slice MFs—namely the NSMF and RAN NSSMF—are deployed within the SMO framework, as shown in Figure 6.
- Deployment Option 2: In this option, both the NSMF and RAN NSSMF are decoupled from and deployed externally to the SMO framework.
- Deployment Option 3: This deployment option integrates the NSMF within the SMO framework, while the RAN NSSMF is placed outside the SMO framework.

- Deployment Option 4: In this deployment option, the NSMF is deployed outside the SMO framework, whereas the RAN NSSMF is integrated within the SMO.

Within O-RAN, the RAN NSSMF, including its interactions with the SMO framework, is the primary area of focus for the O-RAN Alliance [36]. During the creation and provisioning of network slice subnet instance (NSSI), the RAN NSSMF, in coordination with the SMO framework, triggers the instantiation of essential NFs, such as the Near-RT RIC, O-CU-CP, O-CU-UP, and O-DU, according to specific slice requirements. Following the establishment of RAN NSSI, the RAN NSSMF in coordination with SMO framework, may execute procedures for NSSI modification and termination [44].

Each RAN NSSI is identified using the network slice selection assistance information (NSSAI). The NSSAI includes one or a list of single NSSAIs (S-NSSAIs) each serving as unique identifier for a RAN slice [196]. An S-NSSAI is a combination of two values. The first value is the mandatory slice/service type (SST) field, which defines the type of network slice. The SST is an 8-bit value ranging from 0 to 255 and may represent a standardized service type such as eMBB, URLLC, or a network-specific slice type. The second value is the optional slice differentiator (SD) field, a 24-bit value used to distinguish among slices with the same SST. According to 3GPP specifications [196], the NSSAI may contain up to eight S-NSSAIs, which means a single UE can be connected with a maximum of eight RAN NSSIs simultaneously.

The architectural components of the NFV-MANO framework integrated within the SMO framework may consists of following core functional blocks.

a) Network function virtualization orchestrator (NFVO): The NFVO has two responsibilities within the NFV-MANO framework: Firstly, it executes resource orchestration by coordinating the allocation and management of NFV infrastructure (NFVI) resources across multiple virtualized infrastructure managers (VIMs). Secondly, it performs network service orchestration, managing the lifecycle of network services by coordinating groups of VNFs that collectively deliver complex service functionalities. The NFVO enables joint instantiation and configuration of VNFs, ensuring inter-VNF connectivity, and manages dynamic service adaptation. The network service orchestration function relies on collaborative interactions with both the virtual network functions manager (VNFM) and the resource orchestration function that enables abstracted access to the NFVI resources irrespective of the underlying VIMs implementations. Furthermore, it manages VNFs that shares resources within the underlying NFVI [180], [183].

b) Virtual network functions manager (VNFM): The VNFM is responsible for the lifecycle management of one or more VNF instances within a network slice [197]. These VNFs may belong to the same type or different functional categories. In addition, the VNFM is responsible for the FCAPS management of the VNFs, and it also facilitates elastic scalability, that enables VNFs to be dynamically scaled up or down in its designated service region [183].

c) NFV infrastructure (NFVI): The NFV recognizes both software and hardware accelerators as auxiliary resources capable of virtualization that can be exposed as virtual accelerators within the VNF layer [182]. The NFVI includes all the underlying components of the infrastructure—hardware and software—necessary to host VNFs. It abstracts and presents these resources in virtualized forms to be utilized by VNFs and network services, including virtualized compute, storage, and networking capabilities [183].

However, it is important to highlight that current NFV-MANO specifications do not comprehensively address NFVI management aspects, particularly regarding physical infrastructure within cellular networks. As a result, full support of complete infrastructure management service (IMS) functionality is not achievable under the current specifications [182].

d) Virtualized infrastructure manager (VIM): The VIM is responsible for managing and controlling the compute, storage, and network resources of the NFVI within the underlying telecommunication infrastructure [197]. While the deployment and maintenance of the VIM fall outside the formal scope of the NFV-MANO framework, the interfaces it exposes are explicitly included within its scope [183]. The NFV-MANO framework utilizes these interfaces to influence the decisions made regarding the three types of resources (compute, storage, and networking) within the underlying infrastructure.

e) Element management (EM): The EM is equivalent to the NFMF within the 3GPP management system. It is responsible for the FCAPS management of a VNF, encompassing both functional and application layer perspectives. Notably, this functional block also manages the FCAPS of a VNF, but exclusively from a virtualization standpoint [183].

In addition to these functional blocks, the ETSI ISG NFV introduced five new MFs as part of Release 4. These additions aim to enhance the NFV-MANO's capability to support containerized network functions and manage the transport aspects of virtualized infrastructures [183], [197]. Comprehensive details regarding these newly introduced MFs can be found in the Release 4 specifications published by ETSI.

2) 3GPP and ONAP-Based SMO Deployment: As discussed in Section II, the ONAP framework provides the necessary management, orchestration, and automation capabilities for E2E network architecture. Within the OSC, the SMO leverages ONAP along with other components to enable standardized, modular orchestration functions. Particularly noteworthy is the Non-RT RIC, which complements the SMO and utilizes ONAP for efficient A1 policy management [72]. ONAP encompasses predefined workflows and user interfaces (UIs) for 3GPP-defined network slice orchestration functions—namely, the CSMF and the NSMF—along with an additional interface to external NSSMFs for managing RAN, CN, and TN domains. These slice MFs empower the ONAP framework to orchestrate and allocate an E2E NSI, comprising suitable NSSIs across RAN, CN, and TN to meet specific service and use case requirements [103].

ONAP proposes two deployment options for the SMO framework, that emphasizes enhanced integration with the O-RAN architecture, improved cloud-native NF orchestration, and progressing towards intent-driven, closed-loop automation.
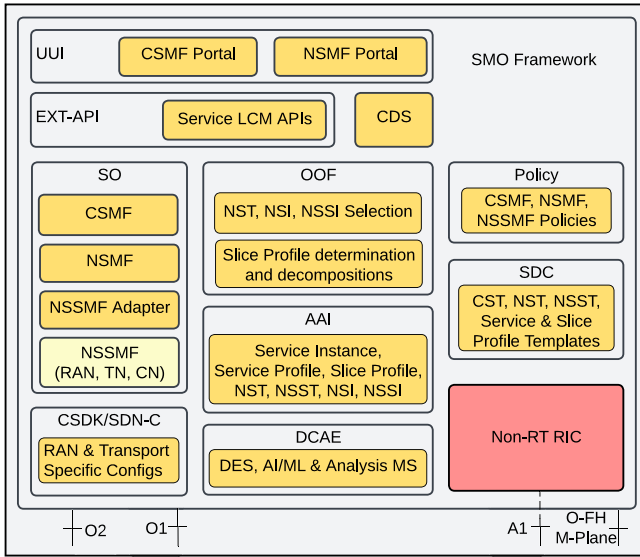
Fig. 7. 3GPP and ONAP-based SMO deployment option with a particular emphasis on O-RAN slicing.

In the first option, the RAN NSSMF is deployed within SMO and is responsible for the M&O of the RAN network slice subnet, including O-RAN NFs and the associated TN components. The RAN NSSMF determines the slice-specific configuration of O-RAN NFs based on slice profile received from the NSMF and identifies the corresponding requirements for the FH and MH interface. It then communicated to the TN management domain (MD), which executes the actual configuration using ETSI zero touch network and service management (ZSM) based MD approach [44], [103].

The second option assigns more comprehensive role to the NSMF, which not only determine the slice profile of RAN NFs, FH, and MH segments but also stitches together E2E network slice instances across domains. This centralization enables the NSMF for consistent orchestration across RAN and TN subnets. In both options, separate RAN network slice subnet templates (NSSTs) are designed to support customized RAN and TN configurations [103].

Figure 7 illustrates the ONAP-based SMO framework incorporating 3GPP-defined slicing management functions. The realization of 5G E2E network slicing depends on the coordination and integration of multiple ONAP functional modules, each described in detail in the following section.

a) Use case user interface (UUI): The UUI operations support an extensive range of lifecycle management actions through an simple point-and-click interface, thereby enabling network operators and service providers to execute tasks more easily [184]. Within the CSMF portal, users can create service request forms to establish network services using specific NSIs. These services can be viewed in a list and managed performing operations like activation, deactivation, or termination. In parallel, the NSMF portal offers functionality to manage slicing-related tasks triggered by customers. Operators can monitor task status, take appropriate actions, and refine slice configurations as suggested by the ONAP optimization framework (OOF). Additionally, the NSMF

portal includes a comprehensive network slicing resource management interface, which allows users to visualize and manipulate existing network slices, NSIs, and NSSIs [198].

b) External API (EXT-API): The EXT-API offers north-bound interoperability for the ONAP platform, serving as an access point for third-party frameworks. Upon receiving a service request, EXT-API responds with a *Service Order ID*, which can be used to track the status of the service order. Subsequently, the EXT-API activates the service orchestrator (SO) API to initiate the actual service creation process. This action represents progress in establishing uniform external interfaces for automated network slice orchestration [103], [198].

c) Controller design studio (CDS): The CDS framework provides blueprint definitions and archives for configuration management processes. It comprises a graphical user interface (GUI) and run time components. The GUI manages user input and displays both the design time and the run time activities. At run time, it allows users to direct the system to resolve the dynamic parameters in blueprint and generate final configuration, which are subsequently downloaded to VNF. The major role of the CDS is to generate and populate a controller blueprint, create a configuration file, and download it to VNF/PNF [184].

d) Service orchestrator (SO): The SO automates sequences of activities, tasks, rules, and policies to execute specified processes required for the on-demand creation, modification, or removal of network, application, or infrastructure services and resources [103]. Within the SO, distinct business process management notation (BPMN) workflows are established for the CSMF and NSMF. The CSMF workflow manages service requests originating from the CSMF portal and stores order information in a communication service instance within the active and available inventory (AAI). It then interacts with the NSMF workflow to initiate slice requests.

The NSMF workflow is responsible for generating service profiles, NSI, and NSSI, all of which can be reused or shared across multiple services [103]. Furthermore, the SO incorporates an NSSMF adapter that interacts with internal or external NSSMFs for NSSI orchestration. The NSSMF functionality includes a common part for subnet capability queries from SO, invoking domain-specific NSSMF functions for the RAN, CN, and TN domains. The specialized workflows of the domain-specific NSSMF handle the essential tasks involved in creating or updating the NSSI according to the guidance provided by the OOF, particularly for new NSSI creation or reuse [103].

e) ONAP optimization framework (OOF): The OOF offers a declarative and policy-driven method for developing and executing optimization applications such as placement and change management, scheduling optimization [103]. The SO interacts with the OOF to select the network slice template (NST) and NSI/NSSI. The OOF may recommend either creating new instances or reusing the existing ones. In the case of NSI/NSSI selection, the OOF could return an existing NSI if it is shareable and suitable, an existing NSSI if shareable and no suitable NSI exists, or a slice profile if the service

request is non-shareable or no suitable NSI or NSSI exists. The recalibration of NSI and NSSI selection is managed by the orchestration task, which allows network operators to intervene manually through the NSMF portal in UUI [198].

f) Service design and creation (SDC): The SDC offers tools, methods, and repositories for defining, simulating, and certifying system assets along with their corresponding processes and policies. These assets are categorized into four groups: resources, services, products, or offers. The SDC environment serves a diverse range of users through shared services and utilities. Within the design studio, product and service designers can onboard, extend, or retire resources, services, and products [184].

g) Active and available inventory (AAI): The AAI provides real-time and historical views of a system's resources, services, products, and their interrelationships. It serves as a dynamic registry, continuously updated by controllers in real-time to support the flexibility of SDN/NFV. The AAI module introduces three additional nodes: Communication-service-profile, Service-profile, and Slice-profile, along with modifications to the service-instance nodes. Furthermore, three new nodes have been incorporated as attributes of the service-instance node. To align with SDC templates such as communication service template (CST), Service Profile Template, Slice Profile Template, NST, and NSST, the run-time instances include communication service instance (CSI), Service Profile Instance, Slice Profile Instance, NSI, and NSSI. The Slice Profile Instance for the all three subnets—RAN, CN, and TN—are distinct [102], [103].

The AAI offers query APIs to CSMF and NSMF, enabling them to retrieve various information such as communication service instances, service profile instances, NSI, and NSSI. Additionally, AAI provides creation of APIs to SO, allowing the creation of communication service profiles, service profiles, slice profiles, and the establishment of relationships between service instances [103].

h) Common controller software development kit (CCSDK)/SDN controller (SDN-C): The CCSDK/SDN-C components manage specific configurations for both the RAN and TN subnets of a network slice. When requested by the SO from the TN NSSMF, they set up and configure a new TN NSSI, including updating the TN during NSI reuse, as well as during activation, deactivation, and termination phases. Similarly, when invoked by the SO from the RAN NSSMF, they (re)configure existing RAN NFs for RAN NSSI or NSI reuse. Additionally, when policy triggers closed loop actions within the RAN for RAN NSSIs, they send relevant configuration updates to the Near-RT RICs [103].

i) Data collection, analysis and event (DCAE): In collaboration with other ONAP runtime components, DCAE provides closed loop automation. It introduces two new micro-services [198]. The first is **data exposure service (DES)**, which offers a simplified interface for network operators, slice tenants, or other ONAP component to query both current and historical PM/KPI data. The second is **Slice Analysis MS**, which analyzes PM data received from the RAN through the PM-Mapper micro-service to detect any updates. When it receives configuration updates, it initiates a control loop by

transmitting a suitable data movement as a platform (DMaaP) message to policy framework.

j) Policy framework: At a granular level, policies consists of machine-readable rules that define actions triggered by specific events or requests, based on given conditions. This approach enables the rapid policy adjustments by updating rules, allowing the adjustment of technical behaviors without rewriting code. The Policy framework simplifies the management of complex mechanisms using abstraction [184].

V. Slicing the Underlying Infrastructure Within the O-RAN Architecture

The transition from distributed to centralized architectures marks a significant shift in RAN designs. The O-RAN adopts the centralized model where major NFs, such as the O-CU, are located in DCs at O-Cloud sites. The O-DU optionally resides either in DC or at the cellular network site. The TN enables data paths across various NFs between and within the RAN and CN domains, thereby delineating distinct TN segments, such as the FH, MH, and BH [199]. Network slicing integrated into the virtualized O-RAN infrastructure boosts efficiency and unlocks unprecedented opportunities for innovation and service differentiation [46].

In the rest of this section, we delve into the underlying infrastructure within the O-RAN architecture, including the components of the cellular network site, the O-Cloud platform, and the Xhaul TN domain, examining the network slicing aspects in these critical elements and domains.

A. O-RAN Cellular Network Site

In wireless networks, a cellular network site serves as a fixed hub for transmitting and receiving radio signals and ensures consistent coverage over a specified area. A cellular network site encompasses two primary components: first, one or more antennas that transmit and receive radio signals, and second, a supply unit that houses essential switching and control elements critical for managing antenna operations.

In standard design, cellular network sites are structured to support multiple sectors, thus inherently associated with several O-RUs. The O-RU serves as a fundamental component in establishing seamless PHY layer connections with UEs [20], [178]. It integrates antenna elements with important RF components such as transceivers and amplifiers to ensure efficient signal processing. Additionally, the O-RU manages lower-level PHY tasks such as digital beamforming and fast fourier transform (FFT) operations [68]. The Open FH interface is crucial for connecting the O-RU with the O-DU to ensure seamless communication in O-RAN architecture.

As explained in Section IV, a balance between FH latency and cost considerations is necessary to optimize the deployment of O-RAN components. Consolidating all elements of the O-gNB at cellular network site minimizes latency but is the most expensive option. Conversely, relocating control and connection anchors towards a centralized edge cloud facilitates resource management across multiple sites while preserving low-latency data processing. Strategically relocating processing functions to the edge cloud while retaining only the
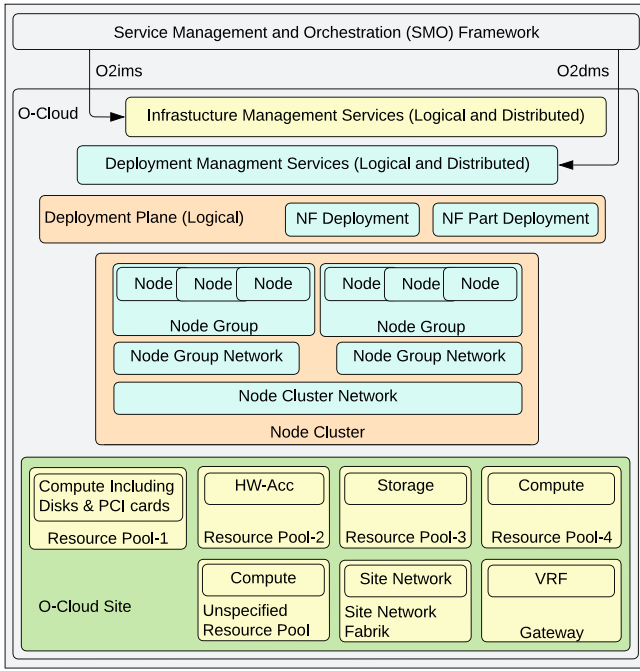
Fig. 8.   Key components in an O-Cloud within the O-RAN architecture.

O-RU at the cellular network site achieves an optimal balance between optimized latency and cost efficiency.

For services with less stringent time requirements, moving the Near-RT RIC and O-CU to a regional cloud may increase latency to around 50 milliseconds. However, this approach optimizes resource processing across various cellular network sites. By centralizing these functions, a single Near-RT RIC can efficiently manage resource allocation while keeping critical processing units closer to end users.

B. O-RAN Cloud Platform

One of the objectives of the O-RAN Alliance is to enhance the flexibility and deployment speed of the RAN architecture while lowering both capital and operational expenses through O-Cloud-based implementations. The O-Cloud platform comprises hardware and software components that deliver O-Cloud capabilities and services to host O-RAN NFs. The logical architecture of the O-RAN combined with the O-Cloud platform and technologies provide a fully open and cloud-native solution where software is decoupled from hardware.

Hardware and software decoupling follows a three-tiered approach: a hardware layer, an intermediary layer with the cloud stack and acceleration abstraction functions, and a top layer dedicated to virtual RAN functions (i.e., O-CU and O-DU). These layers can be sourced from different vendors and the decoupling ensures interoperability between a cloud stack and numerous hardware suppliers as well as accommodate RAN VNFs from various RAN software providers [178].

An O-Cloud platform can automate and autonomously manage tasks with a certain level of complexity such as placing NF deployment workloads on suitable O-Cloud nodes, executing self-repair, and auto-scaling based on deployment artifacts, and policies, without SMO intervention. As illustrated in

Figure 8, an O-Cloud includes O-Cloud resources, Resource pools, and O-Cloud services across multiple sites including the software that manages resource provisioning, nodes, clusters, and deployments on them. It includes the functionality to support the deployment and management services. The O-Cloud provides a unified reference point for all the elements and services within its boundary and scope.

An O-Cloud site refers to a collection of O-Cloud resources at a specific geographical location. The resources are interconnected through O-Cloud site network fabrics. Multiple O-Cloud sites can be interconnected to form a distributed O-Cloud, that requires bridging, routing, or stitching at networking layer in between each O-Cloud site and its respective external transport network attachment point [178].

The O2 interface facilitates connectivity to various O-Cloud services offered by the O-Cloud platform in conjunction with the SMO framework. These services are tailored to address specific functionalities and requirements within the O-Cloud ecosystem [160], [178]. The following sections provide a detailed explanation of each O-Cloud component within the O-RAN architecture, as shown in Figure 8.

1) Infrastructure Management Services: Within the intricate framework of O-Cloud site operations, infrastructure management service (IMS) is a crucial subset of O2 functions, entrusted with the deployment and management of O-Cloud infrastructure. The IMS assumes an important role in provisioning by efficiently allocating and configuring resources for O-Cloud node clusters [200].

In addition, the IMS provides fault and performance management identifying issues and providing measurements to SMO through the O2ims interface. It also provides O-Cloud inventory reporting through O2ims containing details of O-Cloud sites, deployment management services, node clusters, and resources [200]. The O2ims inventory services enables the SMO framework to understand the requested allocation and available O-Cloud capabilities and capacities. The O-Cloud lifecycle management involves registering, structuring, and configuring infrastructure services and resources. Furthermore, this component performs maintenance operations, such as switching O-Cloud nodes to maintenance mode autonomously or on demand to ensure a seamless communication with various components within the SMO framework.

In the following paragraphs, we explore the concepts and perspectives related to O-Cloud IMS and the O-Cloud infrastructure as detailed in reference [178]:

a) O-Cloud resource: This is a defined unit comprising capabilities such as compute, hardware acceleration, storage, and gateway within an O-Cloud site. These resources are provisioned and utilized for the O-Cloud deployment plane, enabling efficient allocation and management of computing resources in cloud-based network infrastructure.

b) O-Cloud resource pool: The O-Cloud resource pool consists of a grouping of O-Cloud resources possessing similar capabilities and traits within an O-Cloud environment. It comprises one or more such resources, each equipped with network connections and, optionally, internal hardware accelerators and storage devices. Additionally, it may include standalone

servers lacking an associated O-Cloud site network fabric, like infrastructure deployed at a cellular network site.

c) Unspecified O-Cloud resource pool: It refers to a collection of O-Cloud resources listed in the O-Cloud IMS inventory but not yet categorized or allocated to any specific O-Cloud resource pool within an O-Cloud site.

d) O-Cloud site network fabric: It serves as an inter-connecting resource within an O-Cloud site, linking various resources within a site to enable seamless communication and data exchange between them. This enhances the overall functionality, interactions, and resource utilization.

e) O-Cloud site network: It represents a meticulously provisioned network resource, which showcases its defined capabilities and characteristics derived from an intricately configured O-Cloud site network fabric.

2) Deployment Management Services: The deployment management service (DMS) efficiently handles various tasks by leveraging information received over O2dms. These tasks encompass the strategically placing O-RAN NF deployment workloads within O-Cloud node clusters. Additionally, DMS manages the entire lifecycle of these workloads, including resource allocation, configuration adjustments, and the executing essential lifecycle management operations such as autonomous scaling, self-healing, and workload relocation within the same O-Cloud node cluster to meet service level expectation (SLE). It also supervises the cessation of NF deployments based on directives from the SMO framework. Furthermore, the DMS ensures the O-Cloud inventory is regularly updated with the latest status information on resources dedicated (or allocated) to NF deployment workloads.

Below are the conceptual insights related to O-Cloud DMS and its interaction with O-Cloud resources generated or modified via IMS provisioning as demonstrated in [178].

a) O-Cloud deployment plane: It refers to a conceptual framework comprising O-Cloud nodes, O-Cloud networks, and O-Cloud node clusters, which are pivotal components for NF deployments. This framework is established by leveraging O-Cloud resources provisioned via IMS, derived from O-Cloud resource pools and O-Cloud site network fabrics.

b) O-Cloud NF deployment: NF deployment refers to deploying software on O-Cloud resources to implement cloud-ified NFs, either fully or partially. This enables the deployment of NFV within cloud environments.

c) O-Cloud node: It is a network connected computer or function, that can be provisioned into O-Cloud node clusters by IMS. The nodes comprises physical or logical components, and expose IMS-assigned resources to form O-Cloud deployment plane constructs. Additionally, an O-Cloud node may operate independently as a standalone entity.

d) O-Cloud node cluster: It consists of a set of O-Cloud nodes operating together via interconnected O-Cloud node cluster networks. The operating system and cluster software of these nodes identify their capabilities and characteristics managed by IMS.

e) O-Cloud node cluster network: It denotes a dedicated network infrastructure tailored for an O-Cloud site network allocated to an O-Cloud node cluster.

f) O-Cloud node group: It refers to a subset of O-Cloud nodes in an O-Cloud node cluster treated equally, particularly by the O-Cloud node cluster scheduler. These nodes are interconnected through O-Cloud node cluster networks and optionally through O-Cloud node group networks.

g) O-Cloud node group network: It refers to the O-Cloud site network designated for a specific grouping of O-Cloud nodes within an O-Cloud node cluster.

C. Xhaul Transport Network

The Xhaul TN serves as a unified TN providing seamless connectivity within and between RAN and CN components. In the O-RAN architecture, it integrates various TN segments across RAN and CN functions such as FH, MH, and BH. The TN, particularly the access segments like FH and MH, can simultaneously manage diverse transport flows, which becomes especially important when operators integrate mixed-use cases into their RAN deployments.

An efficient network resource management strategy is essential due to the varying nature of these transport flows, each with distinct requirements for latency, throughput, and transmission reliability [24]. This is crucial to mitigate complexity and maintain optimal performance across the TN [201]. A strategic approach involves categorizing transport flows into transport slices according to shared service requirements, which enables more structured and efficient TN management. These slices can be further subdivided into sub-slices as needed enabling tailored support for diverse E2E user applications [202]. As network slicing integration advances, discussions within O-RAN focus on incorporating network slicing into the existing TN infrastructure [203]. This includes determining which mobile interfaces–FH, MH, BH, and N6– require network slicing, the structure of these slices, and the optimal number of slices needed at the TN level.

In the subsequent sections, we explore the architecture of the TN, focusing on the complexities of network slicing. The aim is to provide insights into the fundamental principles and practical considerations crucial for the successful deployment and operation of Xhaul TNs within the O-RAN architecture.

1) Xhaul TN Architecture: In the O-RAN Xhaul TN architecture, the FH network connects the O-DU and O-RU with latency models based on eCPRI reference points [48]. The MH network enables communication between O-DU and O-CU with 3GPP-defined F1/E1 interfaces, while the BH network connects the O-CU to the CN [202].

The Xhaul TN architecture must be highly adaptable to accommodate varying requirements based on the specific use case and RAN designs. It involves accommodating numerous next-generation network services, multiple network slices [33], and diverse 3GPP interfaces [202] across different segments of the physical transport network. The O-RAN Alliance in its WG9 transport requirements document [204] has meticulously outlined key prerequisites for the O-RAN Xhaul TN, encompassing bandwidth and latency expectations within the 5G network, as well as logical transport connectivity needs across FH, MH, and BH and even the N6 interface.
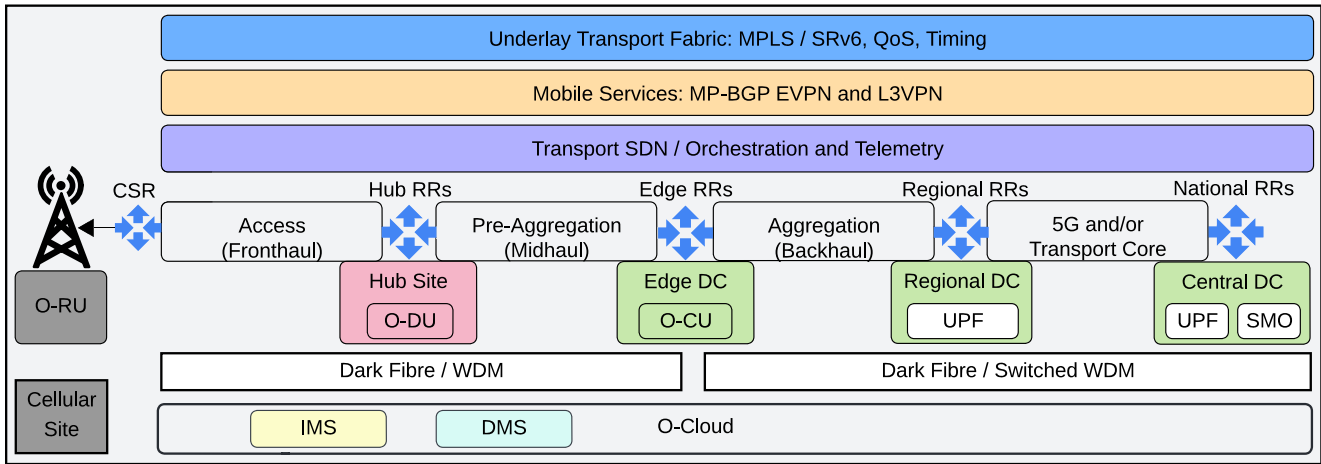
Fig. 9. Packet switched Xhaul TN architecture with a common underlay transport fabric overlaid with a mobile service layer.

The O-RAN introduces new requirements for FH networks, particularly regarding FH latency and data rate [204]. To overcome these obstacles, wavelength division multiplexing (WDM) has emerged as a promising solution, offering various architectural approaches such as passive, active, and semi-active WDM. For more details on WDM, please refer to [205].

The deployment of an E2E Xhaul TN relies on packet switched transport solutions, which is influenced by various key factors. These encompass the extent of packet switching components spanning from cell sites to the transport core and the potential integration of other technologies within the FH network [206]. Additional considerations include the nature of the underlying Layer 0/Layer 1 transport, the choice of network protocols implemented at the packet switching layer, and the framework for implementing the overlay services on the Xhaul TN infrastructure.

Figure 9 illustrates a unified E2E packet switched TN infrastructure. The architecture consists of a common underlay infrastructure overlaid with a service layer that utilizes the shared transport fabric to support mobile services. The DCs are strategically distributed and integrated into the TN architecture, which enables VNFs/PNFs essential for mobile and fixed communication services.

The underlay infrastructure is designed to be scalable which ensures that it meets the diverse service requirements of an O-RAN TN. In contrast, the service infrastructure, or overlay operates above the underlay supporting FH, MH, and BH segments of the O-RAN TN [42], [202]. However, the logical architecture shown in Figure 9 may vary in physical implementation. For instance, some operators may adopt packet switched technology in the MH and BH, while using simpler physical networking for the FH [202], [207].

a) Xhaul TN underlay/fabric technologies: Underlay networks form the physical infrastructure of a TN, comprising Ethernet switches, routers, dense WDM (DWDM) equipments, and the fiber optic cabling that interconnects these components into a coherent topology. To support an Xhaul TN environment, the packet switched network must handle both L2 and L3 services. Currently, L2 underlay networks predominantly rely on Ethernet, often utilizing virtual LANs (VLANs) for segmentation. Within O-RAN, following two prevalent packet switched underlay technologies are discussed in [202].

- *Multi-protocol label switching (MPLS):* MPLS employs label switching in the data plane with multiple control plane technologies including segment routing (SR) an extension to interior gateway protocol (IGP) and border gateway protocol (BGP). Regardless of the MPLS control plane used, the service layer is independent and supports native Ethernet and L3 services [208].
- *SR over IPv6 (SRv6):* SRv6 is built on the SR architecture and operates using an Internet protocol (IP)v6 data plane, where segments are identified by segment IDs (SIDs) embedded in the IPv6 header [209]. While it shares some similarities with SR-MPLS, key differences exist, particularly in the requirements for scaling the underlay infrastructure to support 5G environment.

The underlay is anticipated to offer a comprehensive set of tools necessary to deliver essential network services, encompassing functionalities such as universal connectivity, prioritization, isolation, scalability, rapid convergence, shortest path routing, traffic engineering, packet-based QoS, and precise timing mechanisms [202].

b) Xhaul TN overlay/services infrastructure: Overlay networks utilize network virtualization principles to create virtualized networks composed of overlay nodes such as routers. They leverage technologies like ethernet VPN (EVPN) and multi-protocol border gateway protocol (MP-BGP)-based layer 3 VPNs (L3VPNs) for tunneling encapsulation within the overlay service layer. This encapsulation enables data packets transmission over the underlying physical network while maintaining logical separation and isolation between different virtual networks or network segments [202].

Both MPLS and SRv6 packet switched underlays utilize EVPN for L2 support and MP-BGP for L3VPNs. In the MP-BGP architecture, the protocol is configured with suitable address-family support for both EVPN and L3VPN, facilitating the transmission of service connectivity information among provider edge (PE) devices [208].
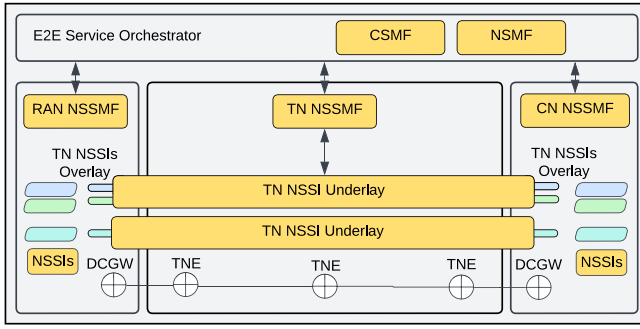
Fig. 10.   Functional architecture of TN slicing.

Ethernet services are provided by EVPN, where EVPN virtual private wire service (VPWS) acts as a transport service for open FH and radio over ethernet (RoE) to ensure redundancy for open FH interface. BGP L3VPNs support IPv4 and IPv6. This approach presents flexible connectivity models with default shortest path routing, along with the option for automatic steering into SR policy [210].

Mobile IP services are facilitated by MP-BGP-based L3VPNs, which establish L3 connectivity among various mobile components. With BGP L3VPN support, network operators can deploy both IPv4 and IPv6 virtual private networks (VPNs), thereby enabling adaptable connectivity models to meet diverse service and network requirements.

2) Xhaul TN Slicing: The packet switched TNs provide a robust framework to support the network slicing. The Xhaul TN infrastructure is designed to support the diverse transport demands. They require tailored solutions for control, management, and user plane interfaces. This includes 5G TN segments, where FH, MH, and BH interfaces require tailored resource allocation [77], [211]. Each interface has specific latency, bandwidth, and traffic demands, which slicing helps to address by enabling precise resource customization [212].

In TN, slicing is categorized into **hard slicing** and **soft slicing**, which determine the level of isolation between network slices [213]. Hard slicing allocates resources exclusively to a particular NSI, ensuring strict assignment with limited resource sharing. Conversely, soft slicing preserves the characteristics of a transport slice while allowing shared and reusable resources across different NSIs [77], [214]. This approach enhances flexibility and improves resource utilization. While hard slicing prioritizes exclusive allocation, soft slicing fosters efficient resource sharing, enhancing resource management.

Figure 10 illustrates Xhaul TN slicing, incorporating orchestration infrastructure along with the RAN, CN, and Xhaul TN. It integrates NSSIs at overlay with TN underlays, demonstrating a comprehensive system design.

The seamless mapping of NSIs to physical or logical TN instances is essential for maintaining coherence between the TN architecture and the unique requirements of each network slice. This mapping process heavily depends upon the available deployment options within the Xhaul TN [215]. The following sections elucidate some key concepts as outlined in O-RAN Alliance specification documents.

a) Transport plane: Within the O-RAN TN infrastructure, both L2 EVPN and L3VPNs leverage MP-BGP to establish individual NSIs. These VPNs support numerous instances and endpoints while offering diverse connectivity models. Furthermore, four distinct approaches are outlined for constructing the underlay transport plane(s), each designed to optimize network performance and address specific network slice operational requirements [202].

- *Single transport plane for all slices:* In this configuration, a single transport plane serves as the backbone for all network slices and ensures a uniform distribution of traffic paths. As a result, each slice follows identical routes between network endpoints, fostering consistency and cohesion across the TN infrastructure. This approach represents the softest form of slicing within the underlay transport plane, as it prioritizes resources sharing and harmonious coexistence among slices.
- *Transport plane per 5G service type:* The underlay network configuration to support various 5G service types involves constructing dedicated transport planes, each customized to accommodate the distinct forwarding behaviors of different services. These transport planes are accessible to multiple customers and enables the deployment of VPNs and traffic steering mechanism. They can adopt distinct topologies and optimizations depending on various criteria, such as service requirements and network conditions. For example, URLLC service types emphasize reliability by selecting the most dependable links and optimizing paths based on link delay metrics. In contrast, eMBB service types prioritize cost-effective, high-bandwidth links, with paths selection determined by IGP metrics that are correlated with link capacity. For narrowband IoT (NB-IoT) service types, which do not require low latency or high capacity, a separate transport plane could be designated, with paths established based on traffic engineering (TE) metrics that favor links specifically suited to these services.
- *Transport plane per slice customer:* In this approach, rather than allocating a transport plane for each 5G service type, a separate transport plane is assigned to individual customers. While similar techniques are employed, scalability now depends on the number of customers utilizing the network, rather than the diversity of 5G service types. To tackle scaling challenges, a hybrid approach can be implemented combining the mappings per customer and per 5G service type. For example, the primary approach might entail mapping according to 5G service types with a subset of premium customers receiving dedicated mappings to individual transport planes.
- *Transport plane per 5QI group:* In this configuration, the TN supports the integration of traffic streams with different 5G QoS identifier (5QI) values into specific network slices. It enables the efficient allocation of numerous 5QIs values to a limited pool of TN resources, such as queues within transport network equipments (TNEs). As a result, the network can effectively manage and prioritize various types of traffic within these designated NSIs.

b) Quality of service: In TN, QoS is essential to ensure that various types of traffic receive appropriate levels of service, such as bandwidth, and latency, to meet specific

performance requirements. The importance of QoS becomes even more pronounced when slicing the transport infrastructure to accommodate diverse traffic types or services. Similar to the transport plane, various strategies can be used to provide varying levels of isolation between slices. These strategies may include traffic prioritization, bandwidth allocation, traffic shaping, and congestion management techniques [215]. When addressing QoS within the context of TN slicing, it's crucial to consider both edge QoS and core QoS solutions.

- *Edge QoS:* Edge interfaces within packet transport networks often experience delays and congestion. In the context of network slicing, it's essential to implement traffic conditioning at the network's edge upon entry and scheduling upon exit to ensure that each slice retains its allocated bandwidth. Moreover, when mobile clients present traffic via VLANs, the PE router must posses hierarchical QoS capabilities to effectively manage both the overall allocated bandwidth at the VLAN level and the designated class bandwidth within each VLAN.
- *Core QoS:* In the CN, various QoS strategies are employed to manage bandwidth allocation and queue sharing across different slices to ensure optimal performance and efficient resource utilization.

c) Service models: In addition to the physical infrastructure, it is essential to establish dedicated networks for managing and controlling various aspects of the 3GPP and O-RAN frameworks. These management and control plane networks are crucial for regulating the functionality and performance of the overall O-RAN architecture [202]. Within the TN slice architectures, management networks and frameworks can function as independent VPN entities, or in some cases, multiple MFs may be consolidated into a single VPN. This consolidation improves efficiency by centralizing tasks and control within a unified network structure, thereby simplifying M&O and enhancing overall coordination.

The TN management network utilizes MP-BGP L3VPN technology to interconnect TNEs through an any-to-any or hierarchical topology, with optional out-of-band networks for redundancy [216]. DC management networks provide centralized oversight of DCs, particularly those supporting 5G and O-RAN services, by integrating logical Ethernet interfaces with TNEs via MP-BGP L3VPN. The O-RAN FH management network operates as a distinct VPN separate from the control VPN, ensuring effective management of mobile elements. Similarly, the O-RAN control and management network consolidates A1, E2, and O1 interfaces into a unified VPN that spans the TN, while the 3GPP control plane network manages all control plane traffic and ensures seamless connectivity and operations for mobile components across different network slice instances and protocols [214].

VI. Exploring Use Cases Related to Network Slicing in O-RAN Architecture

A use case is a concept that describes how a system can be utilized to achieve specific goals or tasks. It outlines the interactions between users or actors and the system to accomplish a particular outcome. The exploration of network slicing within O-RAN encompasses deploying network slices for various use cases while also highlighting the diverse requirements of business customers (also known as tenants) seeking to realize their specific needs. These requirements may encompass ultra-reliable services, high-bandwidth communication, massive machine-type communication, and low latency, among many others. The O-RAN Alliance has identified specific use cases for network slicing to showcase its potential in meeting the demands of business customers. In this section, we delve into several use cases outlined in O-RAN Alliance specifications, expected to be supported within the context of the slicing-aware O-RAN architecture. The requirements derived from these use cases will be integrated into O-RAN architecture as network slicing requirements. Prioritizing and specifying support for these use cases by the O-RAN community are essential, as not all of them have been realized by the specifications of the O-RAN Alliance yet.

A. Slice Subnet Management and Provisioning Use Cases

The aspects related to the M&O of network slicing, including NSI and NSSI, are provided by the 3GPP TSG service and system aspects 5 (SA5). The NSI refers to an instance of an E2E network slice, while the NSSI represents a part of an NSI, such as NSSI for the RAN domain or NSSI for the CN domain. For a comprehensive detailed discussion on the lifecycle management and provisioning of both the NSI and NSSI, interested readers may refer to [38], [217]. In this subsection, we outline the most essential procedures for O-RAN slice subnet management and Provisioning, ensuring alignment with the 3GPP slice management framework and requirements [218]. The use case further outlines the steps involved in its various phases, which include *Creation, Activation, Modification, Deactivation, Termination, Configuration, and Feasibility Check* within O-RAN architecture. It encompasses a diverse array of actors at every phase with their defined roles [44]. The actors include all O-gNB components and MF while their roles are designated as NFMS provider (NFMS_P), NSSMS provider (NSSMS_P), NSSMS cunsumer (NSSMS_C), and O-RAN NFs. Table III outlines the actors, their roles, and the corresponding phases of involvement.

1) Creation: The objective of this phase is to establish the O-RAN NSSI (O-NSSI) or initialize the existing one to meet the RAN slice subnet requirements. The phase assumes that the NSSMS_P is already aware of O-Cloud M&O and begins when the request for an NSSI is received by the NSSMS_P. The NSSMS_P evaluates the feasibility of the request by analyzing the network slice subnet requirements. It then decides whether to modify an existing O-NSSI or create a new slice subnet. The VNFs within O-RAN will then be instantiated by a service request from NSSMS_P to O-Cloud M&O. The response will then be forwarded to O-NSSI, which configures its constituents of O-NSSI using O-RAN NF provisioning service [219]. After that, the NSSMS_P activates TN Manager to establish necessary links such as A1, E2, as well as FH and MH connectivity. The network slice subnet requirements are forwarded to the Non-RT RIC, and NSSMS_C will be informed about the resulting status of this

TABLE III
ACTORS AND THEIR ROLES IN DIFFERENT PHASES OF O-RAN SLICE
SUBNET MANAGEMENT AND PROVISIONING USE CASES

Actors	Creation	Activation	Modification	Deactivation	Termination	Configuration	Feasibility Check
NSMF	NSSMS_C						
NSSMF	NSSMS_P						
NFMF	NFMS_P					NFMS_P	
SMO OAM	NFMS_P					NFMS_P	
O-gNB	NF						
Near-RT RIC	NF						
Non-RT RIC	NF		NF		NF		NF
O-Cloud M&O	MOP		MOP		MOP		MOP
MOP→ O-Cloud M&O provider within SMO							
NF→ O-RAN Network Functions							

TABLE IV
THE O-RAN NFs, ALONG WITH THEIR CURRENT STATE, ARE
REQUIRED TO SATISFY THE PREREQUISITES FOR ACTIVATION

NF	Installed	Activated
Near-RT RIC	Yes	Yes
O-CU-CP	Yes	No
O-CU-UP	Yes	No
O-DU	Yes	No
O-RU	Yes (as PNF)	No

process. Upon successful completion of all steps, the phase concludes with the establishment of the necessary O-RAN NFs and O-NSSI, along with the configuration of the Non-RT RIC [44]. The failure exception may occur due to a full or partial failure of any of the above identified steps.

2) Activation: The goal of this phase is to activate the O-NSSI. It requires that an O-NSSI has already been created but is in an inactive state [217]. This means that several O-RAN NFs may be contained in the O-NSSI but not yet have been activated. To begin the procedure, NSSMS_C sends a request to NSSMS_P to activate the O-NSSI. The NSSMS_P then identifies and decides to activate the parts that are inactive. For example, consider the elements listed in Table IV, where all the NFs are inactive since they are not shared with other O-NSSI, but Near-RT RIC is only activated for other services. NFMS_P makes sure that all the constituents of NSSI are installed and activated on request of NSSMS_P. When all inactive constituents of O-NSSI are activated, NSSMS_P receives a notification from NFMS_P and notifies NSSMS_C about the activation of the O-NSSI. It also changes the administrative state of the O-NSSI to unlocked. This phase ends when all steps are successfully completed, activating the O-NSSI; otherwise, a failure exception is triggered [44].

3) Modification: The objective of this phase is to ensure compliance with O-RAN slice subnet requirements by refining the existing O-NSSI. The only prerequisite of this phase is that the VNF packages for virtualized O-RAN NFs intended for the O-NSSI have been previously incorporated [44]. The process initiates upon receiving a request to modify an existing O-NSSI along with new requirements by the NSSMS_P. Subsequently, feasibility is assessed, leading to two potential outcomes. Should the requirements prove unattainable, the NSSMS_P informs the NSSMS_C of the status along with O-NSSI details. Conversely, the provided information is segmented into modification requests for each constituent of the O-NSSI. If there are additional O-NSSIs managed by other NSSMS_Ps, their respective NSSMS_Ps are notified of the modification via the primary NSSMS_P, thereby activating their O-NSSIs. Additionally, the NSSMS_P can sequentially initiate various required aspects as below:
- A service modification request to O-Cloud M&O, if the O-NSSI contains virtualized parts.
- NF provisioning service to reconfigure the O-NSSI constituents, if the O-NSSI contains NF instances.
- O-RAN TN Manager coordination process, if the NSSI contains TN part, to set up or modify necessary connectivity such as A1, E2, FH, and MH.

Upon successful completion of the above steps, the NSSMS_P informs the Non-RT RIC of the revised network slice subnet requirements and O-NSSI details. Subsequently, it notifies the NSSMS_C regarding the process status, along with relevant O-NSSI information. This phase concludes with modification to the O-NSSI and associated O-RAN NFs, as well as the configuration of the Non-RT RIC to align with the updated slice requirements and O-NSSI specifics.

4) Deactivation: This phase deactivates a currently active O-NSSI. The prerequisite is that O-NSSI exist, is active, and its constituent O-RAN NFs are not shared with other O-NSSIs. NSSMS_C decides to deactivate the O-NSSI upon request of its authorized consumer and sends a deactivation request to the NSSMS_P to start the process. The NSSMS_P identifies the active constituents of the O-NSSI and proceeds to deactivate them. For example, NSSMS_P identifies following active O-RAN NFs that are not shared with other O-NSSI.
- The O-CU-CP NF constituent: It calls the NF provisioning service to request NFMS_P to deactivate the O-CU-CP. The O-CU-CP terminates the E2 interface connecting to the Near-RT RIC and releases the E1 interface between the O-CU-CP and O-CU-UP.
- The O-CU-UP NF constituent: It invokes the NF provisioning service to request NFMS_P to deactivate the O-CU-UP and the O-CU-UP terminates the E2 interface connection with the Near-RT RIC.
- The O-DU NF constituent: It invokes the NF provisioning service to request NFMS_P to deactivate the O-DU. The

O-DU terminates the F1 interface connection with O-CU and the E2 interface connection with Near-RT RIC.

- The O-RU constituent: It invokes the NF provisioning service to request NFMS_P to deactivate the O-RU and the O-RU initializes to terminate the M-Plane interface connecting to the O-DU.

Once the the NFMS_P deactivates the requested constituents, it set the *administrativeState* of that constituent to *locked* and notify the NSSMS_P that the constituent is successfully deactivated. The NSSMS_P sets the *administrativeState* of the O-NSSI to *locked*. Lastly, if all the steps above completes without triggering any exception, this phase is concluded with the deactivation of the O-NSSI.

5) Termination: This phase involves disassociating an existing but inactive O-NSSI when it is no longer required. Upon receiving the termination request, the NSSMS_P takes one of two actions. If the O-NSSI is shared, it is disassociated using the previously described modification phase. If the O-NSSI is non-shared, it is terminated. If there are constituent NSSIs within the O-NSSI that are not directly managed by the NSSMS_P, it requests the respective NSSMS_Ps to release them. It also requests the O-Cloud M&O to terminate the non-shared virtual O-RAN NFs that are no longer required [219] and starts the TN manager coordination process. If all the above steps succeed, the O-NSSI is terminated and NSSMS_P notifies the Non-RT RIC and the NSSMS_C of the final status.

6) Configuration: This phase involves (re-)configuring an existing O-NSSI. It assumes that NSSMS_P is serving authorized customers and know the respective NSSMS_Ps and NFMS_Ps responsible for the management of O-NSSI constituents and NFs. The NSSMS_C initiate (re-)configuration of the O-NSSI and its constituents by sending the slice subnet (re-)configuration information to the NSSMS_P. The NSSMS_P breaks down the received (re-)configuration information to prepare the configuration management for each constituent. The constituents managed directly by the NSSMS_P is configured accordingly. If the constituents are managed by other NSSMS_Ps, the respective NSSMS_P is requested to configure them. For the O-NSSI with constituents O-RAN NFs managed by NFMS_P, the NSSMS_P sends configuration requests through the respective NFMS_Ps.

If any step fails partially or fully, an exception is triggered. Otherwise, the required (re-)configuration is successfully completed for the relevant constituent. The NSSMS_P then sends the configuration results to the NSSMS_C.

7) Feasibility Check: This phase assesses the possibility of provisioning an O-NSSI and confirms whether its requirements are attainable. The precondition is that the NSSMS_C has acquired or received the necessary requirements for the network slice subnet. To start the feasibility check, if an O-NSSI meets the network slice subnet requirements, NSSMS_C sends a request to the NSSMS_P. The NSSMS_P then identifies the involved constituents and may consult the SMO and the Non-RT RIC regarding the fulfillment of requirements. It then checks the availability of network constituents by submitting reservation requests to the O-Cloud M&O.

In addition, the NSSMS_P may request the TN manager to collect information regarding the feasibility of the TN links.

If all steps completed successfully without an exception, the NSSMS_P provides the feasibility check results, including details of reserved resources to the NSSMS_C. Subsequently, the feasibility check phase is concluded.

B. RAN Slice SLA Assurance

The 3GPP standards provide a flexible 5G infrastructure that enables the creation and management of customized networks to meet diverse requirements across various applications and business verticals. These standardized requirements define key performance metrics such as throughput, energy efficiency, latency, and reliability [143]. Network slicing that extends across the CN, TN, and RAN, ensures the strict adherence to performance criteria throughout the entire lifecycle of a network slice, with a particular emphasis placed on the RAN architecture [192]. However, the dynamic nature of the RAN architecture makes it challenging to maintain consistent service quality for each RAN slice within the complex multi-vendor O-RAN environment [46], [162]. Addressing this challenge requires further research and standardization efforts to establish the mechanisms and parameters for the RAN slice SLA assurance [136]. The SLA is a contract between the network service provider and the customer, defining responsibilities, performance standards, and service expectations [220].

O-RAN with its open interfaces and AI/ML-assisted architecture offers a promising approach for implementing the mechanisms that enable operators to fully leverage the opportunities of slicing [218]. For instance, the O-RAN architecture and interfaces empower operators to optimize spectrum resource utilization by dynamically allocating resources across slices based on changing usage patterns. The use case progresses through the following phases:

1) Creation and Deployment of RAN Slice SLA Assurance Models and Control Apps: In this phase, the training and deployment of the model begin with the activation of an O-RAN slice. The prerequisites include an established A1 interface between the Near-RT RIC and the Non-RT RIC, as well as an O1 interface between the SMO and the Near-RT RIC. The phase starts with RAN slice activation. The Non-RT RIC retrieves a RAN slice SLA from the SMO framework, specifically the NSSMF, then collects performance measurements (e.g., CSI, latency) via the O1 interface and enrichment information (e.g., public safety apps, location-based information) from external applications. The Non-RT RIC then analyzes collected performance measurements and/or enrichment information over an extended monitoring period, which contributes in the model training process [192].

The Non-RT RIC performs model training and obtains RAN slice SLA assurance models using either an AI/ML model or a control app [129]. If an AI/ML model is used, it can be deployed internally for slow loop optimization or sent to the Near-RT RIC via the O2 interface for fast loop optimization. Conversely, if a control app is chosen, the SMO deploy it to the Non-RT RIC for slow loop optimization or transfer it to the Near-RT RIC via O2 interface for fast loop optimization. The Non-RT RIC updates the RAN slice SLA assurance

model and control app based on received feedback, either internally or from the Near-RT RIC via the A1 interface. The phase terminates with the deactivation of the RAN slice.

2) Slow Loop RAN Slice SLA Optimization: This phase achieves slow loop RAN slice SLA optimization. The pre-conditions for this phase mirror those of the Creation and Deployment phase, with the addition that the RAN slice SLA assurance model or control apps are already deployed. The Non-RT RIC has two options for slow loop optimization. It can adjust the RAN configuration in accordance with long-term trends, using data from the O1 interface or develop A1 policies tailored to the requirements of the RAN slice SLA. The A1 policies incorporate inputs such as A1 feedback, O1 long-term trends, and operator-defined RAN intents.

In the second option the SMO framework updates the slice configuration of the Near-RT RIC or RAN nodes based on instructions from the AI/ML model or control app. After the update, two outcomes are possible: either the Near-RT RIC and the RAN nodes implement the updated configuration, or the Near-RT RIC receives the updated A1 policies, take control of the RAN nodes and provide feedback to the Non-RT RIC.

3) Fast Loop RAN Slice SLA Optimization: In this phase, the Non-RT RIC evaluates the necessity to generate a policy to ensure slice SLA assurance for the Near-RT RIC. The evaluation is based on the RAN slice SLA requirements and operator-defined RAN intents. It also considers feedback from the Near-RT RIC via the A1 interface or long-term trends observed through the O1 interface, as well as enrichment information from external application servers.

Afterwards, the Near-RT RIC is furnished with slice-specific O1 configurations from SMO and A1 policies from the Non-RT RIC. It proceeds to collect performance measurements via the E2 interface. The collected performance measurements, combined with the A1 policies from the Non-RT RIC, and analyzed by to guide the RAN nodes to meet the slice SLA. The phase concludes with the deactivation of the RAN slice.

C. Managing Multi-Vendor Network Slices

This use case involves managing multiple network slices, each incorporating the RAN components from different vendors. For example, network slice 1 uses O-DU and O-CU from vendor A, while network slice 2 employs components from vendor B, with O-RU from vendor C being shared between both slices [221]. This enables the use of different slices for specific application scenarios, as each component offers unique specifications. While the implementation may vary, they all involve a single O-RU connected to one or more O-DUs. To support multiple slices, the schedulers of the virtualized O-DU (vO-DU) and virtualize O-CU (vO-CU) must manage each NSI separately [162].

The vendor providing vO-DU and vO-CU functionalities must have a robust service-specific customized scheduler.

Moreover, effective coordination between the vO-DU and vO-CU is essential for seamless allocation of radio resources in multi-vendor slices, preventing conflicts. The coordination is evaluated based on service objectives and their impacts on the O-RAN architecture [143]. For instance, the following three potential coordination approach could be explored:

Case 1: The resource allocation between the vO-DU and vO-CU is managed with loose coordination through the O1/A1/E2 interface. Each vO-DU and vO-CU pair is responsible for allocating radio resources to individual business customers within the radio resources allocated by both the Near-RT RIC or the Non-RT RIC.

Case 2: A moderate level of coordination where the resource allocation can be negotiated between slices or between the vO-DU/vO-CUs via the X2 and F1 interfaces, after managed through the O1/E2/A1 interface. The negotiation period is extended to several seconds, influenced by the periodic exchange of the X2 and F1 messages between the vO-CUs.

Case 3: A tight coordination through a new interface between the vO-DUs for adaptive resource allocation, which needs a more frequent negotiation.

The utilization of multi-vendor network slices is applicable in scenarios involving RAN sharing. In such cases, two network operators possess their respective vO-DU and vO-CU components from distinct vendors while jointly utilizing the O-RU component. However, the scenario with O-DU and O-CU components from different vendors within a single slice requires further examination [221].

Adopting a multi-vendor approach cultivates a resilient and adaptable network ecosystem, benefiting operators and end-users alike. Upon the successful implementation of multi-vendor scenarios, the anticipated benefits include:

1) Flexibility and Time-to-Market Deployment: Numerous vendors offer virtualized RAN components like the vO-DU, vO-CU, and schedulers for different network slices. Network operators can thus select the most suitable components for each network slice, whether they prioritize high data rates or low latencies. This flexibility also enables network operators to introduce new services effortlessly, with the option to implement additional functions from different vendors without changing their existing setups and configurations [51].

2) Flexible Deployment for RAN Equipment Sharing: In scenarios where multiple vendors aim to share RAN equipment and resources, challenges may arise concerning vendor selection and the placement of RAN functions. However, by addressing these challenges through collaborative use cases, network operators can reach agreements on shared RAN equipment and resources, thereby optimizing CAPEX and OPEX [71], [222] and potentially opening doors to further business investment opportunities.

3) Supply Chain Risks Reduction: In scenarios where a vendor discontinues support for certain vO-DU and vO-CU functions due to business circumstances, network operators retain the ability to implement substitute vO-DU and vO-CU functions from different vendors within a multi-vendor

framework. This proactive approach serves to alleviate potential risks to network operators' ongoing business operations, bolstering their resilience amidst market dynamics [221].

D. NSSI Resource Allocation Optimization

The increasing complexity of the existing 5G and emerging 6G networks, marked by the proliferation of millimeter-wave small cells and diverse services like eMBB, URLLC, and mMTC, poses significant challenges in dynamically and efficiently allocating resources among network nodes [15]. These services, realized as NSIs, exhibit varying characteristics such as high-speed data, ultra-low latency, and sporadic traffic patterns influenced by factors such as time, location, UE distribution, application types, and others.

To tackle the aforementioned challenges, the optimization of resources allocated to NSSI is crucial. Various scenarios, such as Internet of Things (IoT) applications running during off-peak hours or weekends and large events causing a surge in data flow, are considered. The data collected from the O-RAN nodes serves as input to train an AI/ML model embedded within the NSSI, enabling proactive determination of traffic demand patterns for different times and locations across network slices. This approach facilitates the automatic and intelligent reallocation of resources ahead of network issues, optimizing resource utilization, and ensuring flexibility in responding to diverse service requirements [162].

Implementing resource quota policies within NFs, notably E2 nodes within their respective NSSIs, facilitates efficient management of resource allocation across diverse slices [143]. This flexibility enables the prioritization of resource distribution based on service importance, fostering effective resource sharing during periods of both abundance and scarcity. Premium service slices within an NSSI may receive a more substantial allocation of resources compared to standard or best-effort service slices, while emergency services also benefit from additional resource allocation during critical situations [84]. Acting as constraints for resource allocation, these policies aim to optimize resource utilization across slices. They are adaptable and can be tailored to specific requirements, such as analyzing past resource allocation failures evident in RAN node measurements. This ensures optimal utilization, mitigates historical trends, and minimizes resource inefficiencies.

The O-RAN components involved in this use case are the SMO framework, the Non-RT RIC, and the O-RAN nodes. The SMO establishes the default NSSI resource quota policy, which acts as a parameter for optimizing resource allocation. Meanwhile, the Non-RT RIC gathers performance metrics from the O-RAN nodes, employs the AI/ML models to analyze historical data, predicts traffic demand patterns, and determines appropriate resource adjustments for each NSSI [44], [143]. Subsequently, the Non-RT RIC optimizes the NSSI resource allocation by adjusting attributes and updating cloud resources through the O1 and O2 interfaces, respectively. The O-RAN nodes facilitate performance data collection and configuration updates regarding the NSSI resource allocation via the O1 interface. They also facilities management data collection.

The process of the NSSI Resource Allocation Optimization on the Non-RT RIC may encompass the following steps:

1) Monitoring: The Non-RT RIC monitors the RAN to collect data through the O1 interface and gathers RAN performance measurements from the RAN nodes.

2) Analysis & Decision: The Non-RT RIC leverages an appropriate AI/ML models to analyze measured data and forecast future traffic demand for each NSSI within a specified time interval and geographical location. Based on this analysis, the AI/ML model determines the necessary actions to adjust resources such as the VNF resources and slice subnet attributes for the RAN NFs specifically the E2 Nodes within their respective NSSI at the designated time and location.

3) Execution: The Non-RT RIC executes operations through two sequential steps guided by model inference. Firstly, it adjusts slice subnet attributes via the OAM functions in SMO framework, utilizing O1 interface to configure E2 nodes [223]. Secondly, it triggers a request to the O-Cloud M&O to update resource allocation via the O2 interface. The SMO framework coordinates these operations following recommendations from the Non-RT RIC.

VII. Key Lessons Learned

In this section, we summarize a number of key insights and lessons learned from our study on the ongoing research, development, and deployment efforts of slicing-aware O-RAN architecture. Drawing on a comprehensive review of the literature within the research community, as well as documents from various SDOs (mainly the O-RAN Alliance), we identify several critical observations and valuable lessons. Below, we present a list of these major lessons learned.

A. Lessons Learned Related to the Architecture of O-RAN

We discussed on several occasions that the O-RAN architecture is composed of key components and interfaces that support its open, intelligent, and modular design. We gained several valuable insights from our study on these components and interfaces. The key takeaways are summarized below.

Fully Disaggregated Cloud-based RAN: The adoption of O-RAN marks a pivotal shift towards fully disaggregated RAN architectures, where key O-gNB functions are modularized into distinct components such as the O-CU, O-DU, and O-RU. The key contribution of O-RAN Alliance is the definition of O-FH interface that splits the DU into O-DU and O-RU. **We learned that such a disaggregation can play a significant role in the isolation of O-RAN slices. To further enhance isolation, resource efficiency, and support for cloud-native solutions, it may be beneficial for network operators and vendors to further split the O-CU and O-DU into micro VNFs, such as the virtual network function components (VNFCs) defined by the ETSI ISG NFV.**

RAN Intelligent Controller: Within the context of the O-RAN architecture, the RIC is pivotal in managing and optimizing RAN functions through various control loops. These control loops, are designed to operate at various time scales to enhance network performance. For example, near-real-time control loops operating in milliseconds to seconds

allow the RIC to manage dynamic tasks such as load balancing, interference management, and resource allocation. In contrast, non-real-time control loops operating above second support broader network optimization goals, such as policy-based configurations and performance tuning. Depending on the control loops, the RIC is classified as Near-RT RIC and Non-RT RIC. The RICs are equipped with specialized software tools, known as xApps and rApps, to enhance RAN automation and intelligence. **The key takeaway from these intelligent controllers is that by enabling their distinct control loops, the RIC facilitates continuous, adaptive decision-making that enhances the efficiency and responsiveness of the RAN. Therefore, this approach contributes to a more intelligent and automated environment in O-RAN**.

Service Management & Orchestration: The SMO centralizes the M&O of the resources in O-RAN. It plays a vital role in optimizing and automating the network by monitoring the health, performance, and QoS within the O-RAN ecosystem. The integration of rApps within the SMO enables a more granular level of control and adaptability in the RAN by supporting various time-sensitive and critical functionalities. This setup not only enhances real-time decision-making capabilities but also fosters E2E automation and self-optimization across the O-RAN architecture. The SMO could further evolve by integrating additional management functions, such as slicing management from other SDOs, including 3GPP and NFV-MANO, for the orchestration of O-RAN slicing. **Throughout our study, we learned that this integration would support a more robust, flexible, and standardized approach to network management within O-RAN. Combined with xApps and rApps, the SMO offers network operators a comprehensive management tool, which enables seamless orchestration, automation, and enhanced adaptability across the O-RAN architecture**.

B. Lessons Learned Related to Standardization and Interfaces

In O-RAN, a major advancement lies in the development and promotion of open standards. One major obstacle has been the absence of standardized interfaces and protocols. This makes it difficult for hardware and software from different vendors to work seamlessly together. The adoption of open interfaces, which include both the interfaces inherited from 3GPP and additional new open interfaces defined by the O-RAN Alliance [224]. This approach facilitates a broader ecosystem of interoperability, moving beyond traditional proprietary interfaces to enable a more diverse integration of hardware and software solutions from multiple vendors. **One key lesson is that open interfaces in cellular networks reduce vendor lock-in, which enables flexible component selection and fostering vendor diversity. However, this interoperability supports customized configurations, which drives innovation and efficiency by allowing seamless multi-vendor integration. This ultimately makes the O-RAN architecture more adaptable to the changing demands in 5G, 6G, and beyond cellular networks**.

C. Lessons Learned Related to Vendor Diversity and Collaboration Across the O-RAN Ecosystem

Alongside open interfaces, a key advantage of O-RAN is vendor diversity. By decoupling hardware from software, operators are no longer tied to a single vendor. This approach allows network operators to choose the best of breed solutions [225]. This fosters a competitive multi-vendor ecosystem, driving innovation, reducing costs, and accelerating the deployment of new technologies. **Based on this, another key lesson is the role of O-RAN in defining the O-Cloud reference architecture, which emphasize the cloudification and automation of RAN functions**. This transformation has attracted prominent information technology (IT) companies, historically focused on infrastructure and software, into the telecommunication sector. These players are not only contributing essential infrastructure like servers and container-as-a-service (CaaS) platforms but are also entering the realm of telecommunication by developing critical RAN components such as the RICs and O-CU/O-DU units [226]. **Their involvement is accelerating innovation, strengthening standards development, and expanding global testing capabilities. This trend underscores a major shift towards software-driven RAN functions, with cloud and automation technologies shaping the next generation of mobile networks, setting a foundation for more adaptive, efficient, and scalable wireless communication infrastructures**.

D. Lessons Learned Related to Network Slicing

The integration of network slicing with O-RAN allows network operators to deliver more tailored network services while maximizing commercial benefits [82]. In O-RAN, network slicing leverages its disaggregated architecture and the separation of user and control plane traffic, enhancing flexibility and efficiency. This approach enables RAN NFs to be deployed across edge, regional, or central clouds depending on the specific requirements of the use case and application. This flexibility allows network resources to be dynamically adapted to diverse needs. The TN slicing in O-RAN is implemented through VPNs, categorizing transport flows into distinct transport slices represented by NSIs. **A key lesson learned is that this approach makes the mapping of NSIs to specific physical or logical transport networks a critical aspect of TN slicing, as precise mapping ensures that each slice receives the required network resources and isolation for the respective use case.** Currently, as of O-RAN slicing phase-3, O-RAN supports slicing over the MH and BH segments, while FH slicing remains unsupported but is anticipated to be introduced in later releases of O-RAN specifications.

E. Lessons Learned Related to Transport Network in O-RAN

In O-RAN, the TN is divided into distinct segments—FH, MH, and BH—and is primarily supported by a packet-switched architecture. This architecture relies on an underlay fabric, typically based on MPLS and SRv6, to ensure reliable and efficient data transport across each network segment in both upstream and downstream directions. Overlay services,

provided by L2/L3 VPNs such as EVPN and L3VPN, offer flexible, E2E connectivity and enable service isolation. **One key lesson learned from our study with respect to TN is that this segmented, packet-switched approach enhances data flow and scalability across O-RAN, enabling efficient and adaptable connections throughout a cellular network**.

VIII. Existing Major Research Challenges

The O-RAN architecture and technologies hold substantial potential to drive the evolution of mobile networks toward next-generation solutions. They offer a flexible, disaggregated, and multi-vendor architecture for cellular networks. Although the key principles and specifications for O-RAN are established, the technology remains in its early development stages. As with any emerging technology, the widespread adoption of O-RAN presents significant challenges. This section outlines some key research and engineering challenges, highlighting several obstacles as identified throughout this research work that must be addressed to ensure the successful deployment and integration of O-RAN in cellular networks.

A. Challenges in Multi-Vendor Interoperability

The O-RAN architecture strives to foster seamless multi-vendor interoperability. This approach empowers cellular network operators to evade vendor lock-in and stimulate a more competitive market. However, ensuring seamless integration of components (both software and hardware) from diverse vendors remains one of the most significant research challenges. **For full multi-vendor interoperability, the E2 interface must undergo more extensive testing, with detailed test definitions and profiling**. This testing involves ensuring that components from different suppliers can work harmoniously without compromising performance, security, manageability, or other KPIs. **Experts also doubt whether the industry will unite around a single set of standards**, as vendor-specific interpretations, rapid technological changes, and integration with legacy cellular systems complicate the goal of true "plug-and-play" interoperability [227]. Therefore, **some experts see full interoperability within the context of O-RAN as more aspirational than realistic**.

B. Optimizing Performance and Resource Management

Transitioning from traditional, vertically integrated RAN to a cloud-based RAN architectures, multi-vendor O-RAN presents a complex environment where network resources must be allocated and managed effectively to ensure seamless, reliable, and real-time communication across a wide range of applications and use cases. **One of the primary obstacles in achieving optimal network performance in O-RAN is the allocation and utilization of network resources**. While current cellular networks already struggle with managing diverse traffic flows against network capacity, the challenge is magnified in an O-RAN architecture, which must support a variety of services and applications with unique demands [228]. In addition, the O-RAN architecture should be adaptable to meet the demands of dynamic resource management, enabling support for applications and services like network slicing that cater to distinct use cases or user groups.

The RICs, with their xApps and rApps, play a crucial role in enabling intelligent, targeted resource allocation. This approach allows the O-RAN systems and technologies to adapt dynamically to changing network conditions. In [228] numerous studies have been presented that highlight xApp-based solutions for optimized resource allocation, congestion management, and enhanced network performance. This demonstrates a strong potential for maintaining high service quality in the practical implementations of O-RAN.

QoS is another pivotal aspect for O-RAN performance, especially for latency-sensitive applications and services like autonomous driving and telemedicine that require URLLC [229]. Achieving ultra-low latency is a significant research and engineering challenge in O-RAN, especially in defining Open FH interface requirements between the O-DU and O-RU of an O-gNB [230]. Additionally, **determining the most effective placement of functionalities across the network and ensuring scalability are crucial to meeting performance objectives of O-RAN**. While high performance often demands a more complex architecture, simpler designs may constrain system capabilities [225]. Balancing these factors is key to optimizing both efficiency and adaptability within the context of O-RAN architecture.

To achieve the aforementioned objective, **critical factors like dynamic service chaining, virtualized operating systems, deployment strategies, and functional grouping require thorough analysis** [228]. For example deploying network functions closer to the edge of a cellular network can help reduce congestion at key interfaces, and improving data flow. Additionally, numerous studies suggest using RIC applications to optimize network performance metrics, implementing power regulation to enhance throughput, and continuously monitoring AI/ML models to sustain high-quality network services and prevent performance declines.

C. Leveraging Automation and AI for O-RAN Management

Network and service automation in O-RAN presents both opportunities and challenges, as it goes beyond merely introducing open interfaces to drive the cloudification and automation of RAN operations. **To fully harness the potential of O-RAN and address the increasing complexity, network operators can heavily invest in cloud computing, edge computing, automation, and orchestration technologies**. As networking controllers and orchestrator have evolved from hardware-based solutions to software-defined controllers and are now moving towards AI-driven networks and services, this shift introduces new layers of intricacy.

The RIC and the integration of rApps and xApps will pave the way for a highly automated OAM of O-RAN, achieving the vision of a ZSM, defined within the ETSI framework [187]. However, **integrating these standards into an automated O-RAN system while balancing operational efficiency with fault tolerance requires substantial technological advancements and collaboration across the ecosystem**. The increasing reliance on AI/ML in future networking systems

highlights the inevitable need for robust automation solutions that can handle the complexities of O-RAN effectively.

D. Conflict Mitigation Between the Applications of RICs

The applications of RIC, including xApps and rApps, are designed to optimize and manage O-RAN operations and maintenance by using advanced AI/ML algorithms. These include training models on live data, maintaining low-latency performance, and ensuring that AI/ML models operate transparently and without bias. However, **conflicts between these applications can arise when multiple applications are working to optimize the same network parameters simultaneously**, leading to potential performance degradation [56]. Effective mitigation strategies for conflict detection, resolution, and avoidance are essential to ensuring seamless operation among the applications of both types of RICs.

To address conflicts that arise between xApps and rApps within the Near-RT RIC and Non-RT RIC, a conflict mitigation module is introduced to both RICs. These conflicts may relate to specific users, bearers, or cells, and often stem from RIC configurations or actions. According to the WG3 specifications of the O-RAN Alliance, conflicts fall into three categories: direct, indirect, and implicit.

Direct conflicts, which are easily identifiable by the conflict mitigation module, occur when multiple xApps apply conflicting configurations to the same control target or request more resources than are available. In such cases, the conflict mitigation module resolves the issue by determining which xApp takes precedence and limiting the control action accordingly. On the other hand, indirect and implicit conflicts are less obvious and harder to detect, as they do not present a direct relationship between the conflicting xApps. For instance, one app might optimize the network for certain user groups while unintentionally degrading performance for others. **These types of conflicts are identified and managed through ongoing verification and system monitoring** after control policies are implemented within the Near-RT RIC and Non-RT RIC.

E. Challenges Related to Network Slicing and Orchestration

For O-RAN slicing, different split configurations are essential to accommodate various slice types, fulfilling their unique performance and resource requirements effectively. Additionally, **in the TN, slicing is not yet fully supported on FH components due to the lack of slicing capabilities in O-RUs and O-DUs**. This functionality is anticipated to be introduced in phase-5. A key use case is the "shared O-RU scenario", where a single O-RU is expected to serve multiple slices and multiple O-DUs. In this scenario, the system should be capable of mapping PLMN ID information to the corresponding VLAN and optional IP pair on the control and user planes of the Open FH interface.

Developing effective mechanisms for the M&O of various types of slices within O-RAN is another pressing challenge. The intricacies of assigning resources to different network slice instances, guaranteeing SLAs which are tailored to the requirements of diverse use cases. Achieving this while balancing resource efficiency and avoiding over-provisioning is a critical research challenge for optimizing network slice performance and therefore requires further research and study.

Moreover, **managing network slice instances across an E2E cellular network–from the RAN, TN, to the CN–adds another layer of complexity**. Each network segment must be able to adapt dynamically to fluctuating traffic demands and maintain seamless coordination with the SMO framework to enable real-time control of resources. The O-RAN architecture, with its disaggregated and multi-vendor environment, **further complicates the M&O process as it requires precise synchronization among multiple components and vendors**.

Additionally, the integration of AI/ML into the orchestration process holds promise for automating and optimizing network slice management. Intelligent applications, such as rApps and xApps running on the RICs, can predict traffic demands, optimize resource allocation in real-time, and enforce SLA policies. However, **implementing these AI/ML-driven optimizations across a multi-vendor O-RAN architecture remains an open research area** due to the need for standardized interfaces and seamless data sharing across different vendors.

Network slicing within the O-RAN architecture **involves a multi-dimensional challenge that spans resource management, SLA enforcement, orchestration, real-time optimization, security, and multi-vendor coordination**. Addressing these major research challenges requires not only advanced algorithms and AI/ML integration but also the development of standardized frameworks that support interoperability, unification, and automation across the diverse O-RAN components and interfaces in next-generation of wireless networks.

F. Technical and Standardization Gaps

While O-RAN offers clear benefits, and build upon 3GPP architecture and protocol, also collaborate with ETSI, there are gaps in standardization and technical implementations, especially related to the full integration of network slicing capabilities and automation. **Harmonizing O-RAN standards with existing 3GPP, ETSI, and other SDOs is an ongoing standardization and research challenge**. There is a need for unified frameworks that can accommodate the diverse requirements of O-RAN and traditional network infrastructures. To advance the growth of open cellular networks, these organizations need to work together on establishing common standards, conducting interoperability tests, supporting open source development, and advocating for effective policies.

G. Security Challenges and Risk Mitigation

The O-RAN architecture introduces several security challenges due to its openness and disaggregation principles, which expand potential threats to network and user data [231]. **These challenges include vulnerabilities in the global supply chain and increased attack surfaces**. Network operators need advanced monitoring systems to detect and prevent threats,

while also leveraging automation and distributed security analytics. Virtualized environments enable quick deployment of security patches, but compliance with relevant security standards and certifications, e.g., 3GPP, ETSI is essential. Additionally, **the use of open source protocols, third-party interfaces, and cloud services requires careful security management to ensure network resilience and reliability**.

H. Challenges in Managing the O-RAN Ecosystem

The decentralized nature of O-RAN presents this significant research challenge, **as no single vendor is responsible for the entire E2E RAN implementation**. Although the OSC has established standard interfaces to facilitate interoperability among vendors, challenges persist regarding operations, administration, and maintenance. For example, alarm handling, system commissioning, fault resolution, and performance monitoring in a multi-vendor environment. In the event of unexpected alarms or KPI degradation, effective troubleshooting relies on collaborative efforts among all involved vendors to pinpoint the root cause and implement corrective measures. **Without well-defined processes for engineering support and escalation, such situations can lead to operational delays or conflicts between network operators and suppliers** within the O-RAN architecture.

In such a multi-vendor O-RAN environment, **accurately diagnosing issues and tracing them to their root cause is essential to avoid miscommunication and delays in fault resolution**. This requires advanced systems capable of real-time diagnostics, alongside the integration of AI/ML applications that can predict faults and failures before they happen. Such predictive capabilities enable operators and vendors to take timely, preemptive action, ensuring network stability and minimizing downtime. However, **developing AI models that can adapt to the complex and dynamic nature of O-RAN presents a significant challenge**.

Furthermore, O-RAN must remain adaptable to the continually evolving cellular network landscape. This necessitates regular updates from third-party providers to maintain compatibility. Managing these adjustments within a multi-vendor ecosystem adds complexity, with additional operational costs. While O-RAN seeks to lower overall RAN expenses through enhanced interoperability, the ongoing setup and maintenance across multiple vendors may lead to higher expenditures over time. **This highlights the need for robust cost-monitoring practices to effectively manage expenses and mitigate potential financial strain after deployment**.

IX. Concluding Remarks and Future Outlook

In conclusion, the exploration of O-RAN illuminates its transformative potential within the context of cellular communications systems. As the wireless industry evolves to meet the demands of 5G, 6G, and beyond, O-RAN emerges as a promising paradigm shift, offering flexibility, interoperability, and cost-efficiency in telecommunications networks deployment and management. Through our comprehensive analysis in this paper, it becomes evident that O-RAN's disaggregated approach to wireless network elements, enabled by open interfaces, automation, intelligence, and SDN principles, fosters innovation and competition among vendors while reducing vendor lock-in. This approach not only spurs the development of diverse and specialized NFs but also empowers operators to tailor their networks to specific use cases and environments with greater agility and granularity through the deployment of network slicing at both network and management domains. In essence, while the journey towards realizing the full potential of network slicing in O-RAN may be fraught with several challenges, the destination promises a network architecture that is more open, agile, and responsive to the evolving needs of wireless communication in the next decade.

To explore the topic of network slicing within O-RAN in a detailed manner, we presented its several aspects in this paper, including the architectural framework, network slice deployment options, M&O procedures, and underlying infrastructure, among many others. We began by exploring the ongoing standardization activities within various SDOs and the efforts of the OSC with respect to the realization of O-RAN. Then, we discussed the O-RAN architecture with a particular emphasis on network slicing, covering its SMO framework, O-gNB functionalities, and underlying infrastructure. Next, we studied a number of deployment options for O-gNBs and various types of network slice instances, as well as several deployment options for the MFs and management systems within the SMO framework. We then surveyed network slicing associated with the underlying infrastructure within O-RAN, covering slicing in the cellular network sites, O-Cloud sites, and transport networks. Finally, we addressed several use cases related to the deployment of O-RAN slicing.

Looking ahead, future research endeavors could extend the current work by exploring the potential of xApps and rApps in O-RAN, delving into their capabilities for enhancing network intelligence, service orchestration, and resource optimization. The xApps and rApps may employ advanced ML algorithms to dynamically allocate resources, predict traffic patterns, and optimize performance for each network slice. Additionally, integrating AI/ML models into various optimization functions within O-RAN presents a promising avenue for improving network efficiency, performance, and user experience. By harnessing the power of advanced analytics and automation, future research initiatives can further unlock the transformative potential of O-RAN, propelling the evolution of wireless network infrastructure into a new era of connectivity and innovation. We hope that the insights, together with the deep dive into the O-RAN slicing specifications, architecture, and interfaces, will provide more flexibility for O-RAN slicing deployment by using advanced AI/ML models, as well as various types of xApps and rApps.

In addition to the above research directions, the exploration of a unified SMO architecture that integrates NFV-MANO and ONAP, alongside the decoupled SMO use case defined by the O-RAN Alliance, presents a promising avenue for improving network management efficiency. Tackling the research challenges outlined in Section VIII will be crucial for enhancing the capabilities and reliability of O-RAN deployments, which ultimately foster the development of more resilient and adaptable cellular network infrastructures.

ACKNOWLEDGMENT

Khurshid Alam and Dennis Krummacker are with the Department of Intelligent Networks, German Research Center for Artificial Intelligence, 67663 Kaiserslautern, Germany (e-mail: khurshid.alam@dfki.de; dennis.krummacker@dfki.de).

Mohammad Asif Habibi and Matthias Tammen are with the Division of Wireless Communications and Radio Navigation, Department of Electrical and Computer Engineering, University of Kaiserslautern (RPTU), 67663 Kaiserslautern, Germany (e-mail: asif@eit.uni-kl.de; matthias.tammen@rptu.de).

Walid Saad is with the Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA 22203 USA (e-mail: walids@vt.edu).

Marco Di Renzo is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91192 Gif-sur-Yvette, France, and also with the Department of Engineering, Centre for Telecommunications Research, King's College London, WC2R 2LS London, U.K. (e-mail: marco.di-renzo@universite-paris-saclay.fr; marco.di_renzo@kcl.ac.uk).

Tommaso Melodia is with the Institute for the Wireless Internet of Things, Northeastern University, Boston, MA 02115 USA (e-mail: melodia@northeastern.edu).

Xavier Costa-Pérez is with the 6G Networks R&D Department, NEC Laboratories Europe, 69115 Heidelberg, Germany, also with the Department of AI-Driven Systems, i2CAT Research Center, 08034 Barcelona, Spain, and also with the Department of Engineering Sciences, Catalan Institution for Research and Advanced Studies (ICREA), 08034 Barcelona, Spain (e-mail: xavier.costa@neclab.eu).

Mérouane Debbah is with the 6G Research Center, Department of Computer and Information Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE (e-mail: merouane.debbah@ku.ac.ae).

Ashutosh Dutta is with the Applied Physics Laboratory, Johns Hopkins University, Laurel, MD 20723 USA (e-mail: ashutosh.dutta@jhuapl.edu).

Hans D. Schotten is with the Department of Intelligent Networks, German Research Center for Artificial Intelligence, 67663 Kaiserslautern, Germany, and also with the Division of Wireless Communications and Radio Navigation, Department of Electrical and Computer Engineering, University of Kaiserslautern (RPTU), 67663 Kaiserslautern, Germany (e-mail: hans.schotten@dfki.de).

REFERENCES

[1] V. S. Pana, O. P. Babalola, and V. Balyan, "5G radio access networks: A survey," *Array*, vol. 14, pp. 1–10, Jul. 2022.

[2] M. A. Habibi et al., "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.

[3] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-RAN: Challenges and opportunities," in *Proc. IEEE Wir. Commun. Net. Conf. (WCNC)*, 2020, pp. 1–6.

[4] C. Chaccour et al., "Less data, more knowledge: Building next-generation semantic communication networks," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 1, pp. 37–76, 1st Quart., 2025.

[5] W. Azariah, F. A. Bimo, C. W. Lin, R. G. Cheng, N. Nikaein, and R. Jana, "A survey on open radio access networks: Challenges, research directions, and open source approaches," *Sensors*, vol. 24, no. 3, p. 1038, 2024.

[6] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[7] N. P. Kuruvatti, M. A. Habibi, S. Partani, B. Han, A. Fellan, and H. D. Schotten, "Empowering 6G communication systems with digital twin technology: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112158–112186, 2022.

[8] B. Han et al., "Digital twins for industry 4.0 in the 6G era," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 820–835, 2023.

[9] S. Kerboeuf et al., "Design methodology for 6G end-to-end system: Hexa-X-II perspective," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 3368–3394, 2024.

[10] M. Kassi and S. Hamouda, "RAN virtualization: How hard is it to fully achieve?" *IEEE Access*, vol. 12, pp. 38030–38047, 2024.

[11] M. A. Hasabelnaby, M. Obeed, M. Saif, A. Chaaban, and M. J. Hossain, "From Centralized RAN to open RAN: A survey on the evolution of distributed antenna systems," Nov. 2024, *arXiv:2411.12166*.

[12] P. K. Thiruvasagam et al., "Open RAN: Evolution of architecture, deployment aspects, and future directions," Jan. 2023, *arXiv:2301.06713*.

[13] J. L. Frauendorf and É. Almeida de Souza, *The Evolution of RAN (Radio Access Network), D-RAN, C-RAN, V-RAN, and O-RAN*. Cham, Switzerland: Springer Int. Publ., 2023, pp. 139–154.

[14] F. Mungari, C. Puligheddu, A. Garcia-Saavedra, and C. F. Chiasserini,, "O-RAN intelligence orchestration framework for quality-driven xApp deployment and sharing," *IEEE Trans. Mobile Comput.*, vol. 24, no. 6, pp. 4811–4828, Jun. 2025.

[15] L. Bonati et al., "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 21–27, Oct. 2021.

[16] *5G; NG-RAN; Architecture Description; Technical Specification; Release 17 Version 17.1.1*, 3GPP Standard TS 38.401, Aug. 2022.

[17] W. Rouwet, *Open Radio Access Network (O-RAN) Systems Architecture and Design*. Amsterdam, The Netherlands: Elsevier, 2022. [Online]. Available: https://books.google.de/books?id=mzFJEAAAQBAJ

[18] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Comput. Netw.*, vol. 182, pp. 1–28, Aug. 2020.

[19] M. A. Habibi, F. Z. Yousaf, and H. D. Schotten, "Mapping the VNFs and VLs of a RAN slice onto intelligent PoPs in beyond 5G mobile networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 670–704, 2022.

[20] (O-RAN Alliance, Alfter, Germany). *Working Group 1; O-RAN Architecture Description; Technical Specification; Release 004 Version 13.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[21] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1376–1411, 2nd Quart., 2023.

[22] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What O-RAN can and cannot do!" *IEEE Netw.*, vol. 36, no. 6, pp. 206–213, Nov./Dec. 2022.

[23] F. Rezazadeh et al., "On the specialization of FDRL agents for scalable and distributed 6G RAN slicing orchestration," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3473–3487, Mar. 2023.

[24] *The Essential Guide for Understanding O-RAN*, Keysight, Santa Rosa, CA, USA, Mar. 2023.

[25] S. Maxenti et al., "AutoRAN: Automated and zero-touch open RAN systems," Apr. 2025, *arXiv:2504.11233*.

[26] "Driving open RAN forward; an improved open fronthaul interface bringing performance to open RAN," Ericsson, Stockholm, Sweden, White Paper, Nov. 2023.

[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 J. Commun. Soc.*, vol. 4, pp. 1615–1658, 2023.

[28] S. Niknam et al., "Intelligent O-RAN for beyond 5G and 6G wireless networks," in *Proc. IEEE Globecom Workshops*, 2022, pp. 215–220.

[29] C. K. Thomas, C. Chaccour, W. Saad, M. Debbah, and C. S. Hong, "Causal reasoning: Charting a revolutionary course for next-generation AI-native wireless networks," *IEEE Veh. Technol. Mag.*, vol. 19, no. 1, pp. 16–31, Mar. 2024.

[30] M. Chen et al., "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart., 2019.

[31] A. Giannopoulos et al., "Supporting intelligence in disaggregated open radio access networks: Architectural principles, AI/ML workflow, and use cases," *IEEE Access*, vol. 10, pp. 39580–39595, 2022.

[32] A. U. Rehman, R. L. Aguiar, and J. P. Barraca, "Network functions virtualization: The long road to commercial deployments," *IEEE Access*, vol. 7, pp. 60439–60464, 2019.

[33] M. A. Habibi, B. Han, F. Z. Yousaf, and H. D. Schotten, "How should network slice instances be provided to multiple use cases of a single vertical industry?" *IEEE Commun. Stand. Mag.*, vol. 4, no. 3, pp. 53–61, Sep. 2020.

[34] "Critical 5G components—Radio access networks," Ericsson, Stockholm, Sweden, White Paper, Mar. 2024.

[35] C.-Y. Chang and N. Nikaein, "RAN runtime slicing system for flexible and dynamic service execution environment," *IEEE Access*, vol. 6, pp. 34018–34042, 2018.

[36] (O-RAN Alliance, Alfter, Germany). *Working Group 1; Study on O-RAN Slicing; Technical Report; Version 02.00*. Apr. 2020. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[37] M. A. Habibi et al., "The architectural design of service management and orchestration in 6G communication systems," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2023, pp. 1–2.

[38] "Telecommunication management; study on management and orchestration of network slicing for next generation network," 3GPP, Sophia Antipolis, France, Rep. 28.801, Jan. 2018.

[39] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "On radio access network slicing from a radio resource management perspective," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 166–174, Oct. 2017.

[40] B. Khodapanah, A. Awada, I. Viering, A. N. Barreto, M. Simsek, and G. Fettweis, "Framework for slice-aware radio resource management utilizing artificial neural networks," *IEEE Access*, vol. 8, pp. 174972–174987, 2020.

[41] C.-L. I, S. Kuklinskí, and T. Chen, "A perspective of O-RAN integration with MEC, SON, and network slicing in the 5G era," *IEEE Netw.*, vol. 34, no. 6, pp. 3–4, Nov./Dec. 2020.

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

[43] N. Sen and A. Antony Franklin, "Intelligent admission and placement of O-RAN slices using deep reinforcement learning," in *Proc. IEEE 8th Int. Conf. Netw. Softw. (NetSoft)*, 2022, pp. 307–311.

[44] (O-RAN Alliance, Alfter, Germany). *Working Group 1; Slicing Architecture; Technical Specification; Release 004 Version 14.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[45] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "OpenRAN gym: An open toolbox for data collection and experimentation with AI in O-RAN," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2022, pp. 518–523.

[46] W. Wu et al., "Dynamic RAN slicing for service-oriented vehicular networks via constrained learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2076–2089, Jul. 2021.

[47] M. A. Habibi, G. M. Yilma, X. Costa-Pérez, and H. D. Schotten, "Unifying 3GPP, ETSI, and O-RAN SMO interfaces: Enabling slice subnets interoperability," in *Proc. IEEE FNWF*, Baltimore, MD, USA, 2023, pp. 1–8.

[48] D. Wypiór, M. Klinkowski, and I. Michalski, "Open RAN—Radio access network evolution, benefits and market trends," *Appl. Sci.*, vol. 12, no. 1, p. 408, 2022.

[49] A. Arnaz, J. Lipman, M. Abolhasan, and M. Hiltunen, "Toward integrating intelligence and programmability in open radio access networks: A comprehensive survey," *IEEE Access*, vol. 10, pp. 67747–67770, 2022.

[50] B. Brik, K. Boutiba, and A. Ksentini, "Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 228–250, 2022.

[51] M. Q. Hamdan et al., "Recent advances in machine learning for network automation in the O-RAN," *Sensors*, vol. 23, no. 21, 2023.

[52] M. Alavirad et al., "O-RAN architecture, interfaces, and standardization: Study and application to user intelligent admission control," *Front. Commun. Netw.*, vol. 4, pp. 1–17, Mar. 2023.

[53] M. Liyanage, A. Braeken, S. Shahabuddin, and P. Ranaweera, "Open RAN security: Challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 214, May 2023, Art. no. 103621.

[54] Y. Huang et al., "Validation of current O-RAN technologies and insights on the future evolution," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 487–505, Feb. 2024.

[55] M. Polese et al., "Empowering the 6G cellular architecture with open RAN," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 245–262, Feb. 2024.

[56] S. Marinova and A. Leon-Garcia, "Intelligent O-RAN beyond 5G: Architecture, use cases, challenges, and opportunities," *IEEE Access*, vol. 12, pp. 27088–27114, 2024.

[57] M. Wani, M. Kretschmer, B. Schröder, A. Grebe, and M. Rademacher, "Open RAN: A concise overview," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 13–28, 2025.

[58] M. A. Habibi, B. Han, M. Nasimi, N. P. Kuruvatti, A. Fellan, and H. D. Schotten, *Towards a Fully Virtualized, Cloudified, and Slicing-Aware RAN for 6G Mobile Networks*. Cham, Switzerland: Springer Int. Publ., 2021, pp. 327–358.

[59] (Deutsche Telekom AG, Bonn, Germany). *Easy and Simple: Network Disaggregation*. Feb. 2020. [Online]. Available: https://www.telekom.com/en/company/details/easy-simple-network-disaggregation-593822

[60] C.-X. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 16–23, Feb. 2020.

[61] D. Kataria et al., "Artificial intelligence and machine learning," in *Proc. IEEE FNWF*, 2022, pp. 1–70.

[62] J. Groen et al., "Securing O-RAN open interfaces," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 11265–11277, Dec. 2024.

[63] T. Nolle, "Transformation and 5G O-RAN," VMware, Palo Alto, CA, USA, White Paper, May 2021. [Online]. Available: https://telco.vmware.com/resources.html

[64] M. A. Habibi, B. Han, M. Saimler, I. L. Pavón, and H. D. Schotten, "Towards an AI/ML-driven SMO framework in O-RAN: Scenarios, solutions, and challenges," in *Proc. IEEE FNWF*, 2024, pp. 1–8.

[65] "The evolution of open RAN," 5G Americas, Bellevue, WA, USA, White Paper, Feb. 2023. [Online]. Available: https://tinyurl.com/eoran5ga

[66] L. Gavrilovska, V. Rakovic, and D. Denkovski, "From cloud RAN to open RAN," *Wireless Pers. Commun.*, vol. 113, no. 3, pp. 1523–1539, Aug. 2020.

[67] A. Hörnes et al., "Deutsche Telekom and partners demonstrate non-real time RAN optimization in a multi-vendor environment," TinyURL, Sioux Falls, SD, USA, White Paper, Sep. 2023. [Online]. Available: https://tinyurl.com/dtnrtric

[68] "Everything you need to know about open RAN," Parallel Wireless, Nashua, NH, USA, White Paper, 2020. [Online]. Available: https://tinyurl.com/pworan

[69] "White paper; O-RAN: An open ecosystem to power 5G applications," White Paper, vIAVI Solut., Scottsdale, AZ, USA, 2021.

[70] Q. Sun, N. Li, C.-L. I, J. Huang, X. Xu, and Y. Xie, "Intelligent RAN automation for 5G and beyond," *IEEE Wireless Commun.*, vol. 31, no. 1, pp. 94–102, Feb. 2024.

[71] P. H. Masur, J. H. Reed, and N. K. Tripathi, "Artificial intelligence in open-radio access network," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 9, pp. 6–15, Sep. 2022.

[72] H. Lee, J. Cha, D. Kwon, M. Jeong, and I. Park, "Hosting AI/ML workflows on O-RAN RIC platform," in *Proc. IEEE Globecom Workshops*, 2020, pp. 1–6.

[73] J. Zhang et al., "Intent-driven closed-loop control and management framework for 6G open RAN," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 6314–6327, Feb. 2024.

[74] M. A. Habibi, G. M. Yilma, U. Fattore, X. Costa-Pérez, and H. D. Schotten, "Unlocking O-RAN potential: How management data analytics enhances SMO capabilities?" *IEEE Open J. Commun. Soc.*, vol. 5, pp. 4710–4730, 2024.

[75] Y. Sanjalawe, S. Fraihat, S. Al-E'Mari, M. Abualhaj, S. Makhadmeh, and E. Alzubi, "A review of 6G and AI convergence: Enhancing communication networks with artificial intelligence," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 2308–2355, 2025.

[76] F. Lotfi and F. Afghah, "Open RAN LSTM traffic prediction and slice management using deep reinforcement learning," in *Proc. 57th Asilomar Conf. Signals Syst. Comput.*, 2023, pp. 646–650.

[77] S. Bhattacharjee et al., "Network slicing for TSN-based transport networks," *IEEE Access*, vol. 9, pp. 62788–62809, 2021.

[78] A. S. Abdalla and V. Marojevic, "End-to-end O-RAN security architecture, threat surface, coverage, and the case of the open fronthaul," *IEEE Commun. Stand. Mag.*, vol. 8, no. 1, pp. 36–43, Mar. 2024.

[79] X. Li, R. Ni, J. Chen, Y. Lyu, Z. Rong, and R. Du, "End-to-end network slicing in radio access network, transport network and core network domains," *IEEE Access*, vol. 8, pp. 29525–29537, 2020.

[80] P. Muñoz, Ñ. Adamuz-Hinojosa, J. Navarro-Ortiz, O. Sallent, and J. Pérez-Romero, "Radio access network slicing strategies at spectrum planning level in 5G and beyond," *IEEE Access*, vol. 8, pp. 79604–79618, 2020.

[81] *Official Document NG.135; Technical Requirements for E2E Network Slicing*, v4.0, GSM Association, London, U.K., Jul. 2024.

[82] H. Cheng et al., "ORANSlice: An open source 5G network slicing platform for O-RAN," in *Proc. 30th. Annu. Int. Conf. Mob. Comput. Netw.*, 2024, pp. 2297–2302.

[83] P. Wyszkowski, J. Kienig, K. Zielinski, L. Czekierda, and M. Zawadzki, "Comprehensive tutorial on the organization of a standards-aligned network slice/Subnet design process and opportunities for its automation," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 2, pp. 1386–1445, 2nd Quart., 2024.

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

[85] R. Schmidt and N. Nikaein, *Radio Access Network Slicing System*. Hoboken, NJ, USA: Wiley, May 2020, pp. 1–32.

[86] J. Thaliath et al., "Predictive closed-loop service automation in O-RAN based network slicing," *IEEE Commun. Stand. Mag.*, vol. 6, no. 3, pp. 8–14, Sep. 2022.

[87] K. Boutiba, A. Ksentini, B. Brik, Y. Challal, and A. Balla, "NRflex: Enforcing network slicing in 5G new radio," *Comput. Commun.*, vol. 181, pp. 284–292, Jan. 2022.

[88] W. Jiang et al., "Terahertz communications and sensing for 6G and beyond: A comprehensive review," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2326–2381, 4th Quart., 2024.

[89] "NR; study on enhancement of radio access network (RAN) slicing; technical report; release 17," 3GPP, Sophia Antipolis, France, Rep. 38.832, Jun. 2021.

[90] S. Kukliński, L. Tomaszewski, and R. Kolakowski, "On O-RAN, MEC, SON and network slicing integration," in *Proc. IEEE Globecom Workshops*, Dec. 2020, pp. 1–6.

[91] C. Harper and S. Sirotkin, *NG-RAN Architecture*. Hoboken, NJ, USA: Wiley, 2021, pp. 123–234.

[92] R. MacKenzie et al. (NGMN Alliance, Düsseldorf, Germany). *NGMN Overview on 5G RAN Functional Decomposition*. Feb. 2018. [Online]. Available: https://www.ngmn.org/publications/ngmn-overview-on-5g-ran-functional-decomposition.html

[93] "O-RAN alliance overview presentation," O-RAN Alliance, Alfter, Germany, White Paper, Aug. 2024.

[94] A. Lacava et al., "Programmable and customized intelligence for traffic steering in 5G networks using open RAN architectures," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2882–2897, Apr. 2024.

[95] D. Mavrakis, "The road to open RAN productization," aBI Res., New York, NY, USA, White Paper, Sep. 2021.

[96] A. Kliks et al., "Towards autonomous open radio access networks," *ITU J. Future Evol. Technol.*, vol. 4, no. 2, pp. 251–268, May 2023.

[97] (Deutsche Telekom AG, Bonn, Germany). *Open RAN Technical Priority Document; Release 4*. Jun. 2024. [Online]. Available: https://telecominfraproject.com/openran-mou-group/

[98] P. Chitrapu, *Small Cell Open RAN: A Catalyst for New 5G Business Models*, Small Cell Forum, Dursley, Gloucestershire, Apr. 2023.

[99] *5G nFAPI Specifications Document*, Small Cell Forum, Dursley, Gloucestershire, Jul. 2022.

[100] M. Skorupski et al. "Common ONAP/O-RAN-SC areas." Accessed: Mar. 17, 2025. [Online]. Available: https://tinyurl.com/osconap

[101] V. Begwani. "3GPP O-RAN ONAP alignment." Nov. 2019. [Online]. Available: https://tinyurl.com/vbegwani

[102] *Kohn Release Document*, ONAP, San Francisco, CA, USA, Dec. 2022.

[103] L. Meng et al., "ONAP E2E network slicing technical overview," White paper, Jun. 2020. [Online]. Available: https://tinyurl.com/onape2ensto

[104] "MOSAIC5G." Accessed: Mar. 17, 2025. [Online]. Available: https://mosaic5g.io

[105] F. Kaltenberger et al., "Driving innovation in 6G wireless technologies: The OpenAirInterface approach," *Comput. Netw.*, vol. 269, Sep. 2025, Art. no. 111410.

[106] M. Polese et al., "Colosseum: The open RAN digital twin," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 5452–5466, 2024.

[107] O. Sunay et al., "SD-RAN: ONF's software-defined RAN platform consistent with the O-RAN architecture," oNF, Alfort, France, White Paper, Aug. 2020.

[108] P. S. Upadhyaya, A. S. Abdalla, V. Marojevic, J. H. Reed, and V. K. Shah, "Prototyping next-generation O-RAN research Testbeds with SDRs," May 2022, *arXiv:2205.13178*.

[109] "srsRAN documentation." Accessed: Mar. 17, 2025. [Online]. Available: https://docs.srsran.com

[110] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "OpenRAN gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms," *Comput. Netw.*, vol. 220, pp. 1–11, Jan. 2023.

[111] L. Bonati et al., "Colosseum: Large-scale wireless experimentation through hardware-in-the-loop network emulation," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, Dec. 2021, pp. 105–113.

[112] *Profiles of U.S. Open RAN Research and Testing Facilities*, Nat. Spectrum Consort., Open RAN Advisory Group, Washington, DC, USA, Feb. 2023.

[113] "ORAN on POWDER." POWDER. Accessed: Mar. 17, 2025. [Online]. Available: https://powderwireless.net/oran

[114] D. Johnson, D. Maas, and J. V. der Merwe, "NexRAN: Closed-loop RAN slicing in POWDER—A top-to-bottom open-source open-RAN use case," in *Proc. WiNTECH*, 2022, pp. 17–23.

[115] D. Raychaudhuri et al., "Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, vol. 14, Apr. 2020, pp. 177–189.

[116] L. Bertizzolo et al., "Arena: A 64-antenna SDR-based ceiling grid testing platform for sub-6 GHz 5G-and-beyond radio spectrum research," *Comput. Netw.*, vol. 181, Nov. 2020, Art. no. 107436.

[117] D. Villa et al., "An open, programmable, multi-vendor 5G O-RAN testbed with NVIDIA ARC and OpenAirInterface," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2024, pp. 1–6.

[118] H. Koumaras et al., "5GENESIS: The genesis of a flexible 5G facility," in *Proc. IEEE 23rd Int. Workshop Comput. Aided Model. Des. Commun. Links Netw.*, Sep. 2018, pp. 1–6.

[119] H. Djuitcheu et al., "Performance evaluation of industrial O-RAN: A comparative analysis of gNB providers," Techrxiv, Apr. 2025.

[120] A. Ingo Grohmann, M. Seidel, S. A. W. Itting, R.-G. Cheng, M. Reisslein, and F. H. P. Fitzek, "Multi-UE 5G RAN measurements: A gamut of architectural options," *IEEE Access*, vol. 13, pp. 1846–1866, 2025.

[121] M. Amini and C. Rosenberg, "A comparative analysis of open-source software in an E2E 5G standalone platform," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2024, pp. 1–6.

[122] J. E. Håkegård, H. Lundkvist, A. Rauniyar, and P. Morris, "Performance evaluation of an open source implementation of a 5G standalone platform," *IEEE Access*, vol. 12, pp. 25809–25819, 2024.

[123] J. Shin, "Performance evaluation of OpenAirInterface-based 5G standalone testbed," in *Proc. Int. Conf. ICT Converg.*, 2024, pp. 1944–1945.

[124] M. Guemdani, P. Gu, C.-D. Phung, and S. Secci, "Comparative E2E performance analysis of O-RAN designs in a 5G standalone testbed," in *Proc. IEEE Int. Conf. High Perform. Switch. Routing*, 2025, pp. 1–6.

[125] D. Kim, E. Lee, and S. Pack, "Design and implementation of cloud-native open RAN testbed," in *Proc. Int. Conf. ICT Converg.*, 2024, pp. 2167–2168.

[126] J. Baranda, A. Bel, S. Barrachina-Muñoz, M. Payaró, and J. Mangues-Bafalluy, "Distributed sequential cloud-native deployment of an end-to-end 5G network with O-RAN functions," in *Proc. 15th Int. Conf. Netw. Future*, 2024, pp. 142–144.

[127] A. K. Subudhi et al., "Performance evaluation of scheduling scheme in O-RAN 5G network using NS-3," in *Proc. IEEE FNWF*, 2024, pp. 590–595.

[128] J. L. Herrera, S. Montebugnoli, D. Scotece, L. Foschini, and P. Bellavista, "A tutorial on O-RAN deployment solutions for 5G: From simulation to emulated and real testbeds," *IEEE Commun. Surveys Tuts.*, early access, Jun. 9, 2025, doi: 10.1109/COMST.2025.3577687.

[129] N. D. Tripathi and V. K. Shah, *Fundamentals of O-RAN*. Hoboken, NJ, USA: Wiley, 2025.

[130] "Open RAN: An interoperable, innovative, and sovereign architecture," CST SA, Riyadh, Saudi Arabia, White Paper, Jan. 2023. [Online]. Available: https://www.cst.gov.sa/en/research-innovation/Documents/Open_RAN.pdf

[131] M. V. Ngo et al., "RAN intelligent controller (RIC): From open-source implementation to real-world validation," *ICT Exp.*, vol. 10, no. 3, pp. 680–691, 2024.

[132] R. Niklasson et al., "An intelligent platform: The use of O-RAN's SMO as the enabler for openness and innovation in the RAN domain," Ericsson, Stockholm, Sweden, White Paper, Nov. 2021.

[133] M. Dryjański, L. Kulacz, and A. Kliks, "Toward modular and flexible open RAN implementations in 6G networks: Traffic steering use case and O-RAN xApps," *Sensors*, vol. 21, no. 24, p. 8173, Dec. 2021.

[134] M. K. Motalleb, V. Shah-Mansouri, and S. N. Naghadeh, "Joint power allocation and network slicing in an open RAN system," Nov. 2019, *arXiv:1911.01904*.

[135] R. Wiebusch, N. A. Wagner, D. Overbeck, F. Kurtz, and C. Wietfeld, "Towards open 6G: Experimental O-RAN framework for predictive uplink slicing," in *Proc. IEEE Int. Conf. Commun.*, 2023, pp. 4834–4839.

[136] E. Moro, M. Polese, A. Capone, and T. Melodia, "An open RAN framework for the dynamic control of 5G service level agreements," in *Proc. IEEE Conf. Netw. Funct. Virtualiz. Softw. Defin. Netw.*, 2023, pp. 141–146.

[137] L. M. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "The evolution of mobile network operations: A comprehensive analysis of open RAN adoption," *Comput. Netw.*, vol. 243, Apr. 2024, Art. no. 110292.

[138] *5G; NG-RAN; E1 General Aspects and Principles; Technical Specification; Release 16 Version 16.5.0*, 3GPP Standard TS 38.460, May 2022.

[139] A. Ndao, X. Lagrange, N. Huin, G. Texier, and L. Nuaymi, "Optimal placement of virtualized DUs in O-RAN architecture," in *Proc. IEEE 97th Veh. Technol. Conf.*, 2023, pp. 1–6.

[140] S.-P. Yeh, S. Bhattacharya, R. Sharma, and H. Moustafa, "Deep learning for intelligent and automated network slicing in 5G open RAN (ORAN) deployment," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 64–70, 2024.

[141] M. Silva et al., "O-RAN and RIC compliant solutions for next generation networks," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2023, pp. 1–7.

[142] W. Jie and R. Jain. "An overview of open RAN: Basics, recent advances and future research." Dec. 2022. [Online]. Available: https://tinyurl.com/jiewustl

[143] (O-RAN Alliance, Alfter, Germany). *Working Group 1; Use Cases Detailed Specification; Technical Specification; Release 004 Version 16.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[144] (O-RAN Alliance, Alfter, Germany). *Working Group 3; E2 General Aspects and Principles; Technical Specification; Release 004 Version 07.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[145] (O-RAN Alliance, Alfter, Germany). *Working Group 3; E2 Service Model (E2SM); Technical Specification; Release 004 Version 07.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[146] (O-RAN Alliance, Alfter, Germany). *Working Group 3; Near-RT RIC Architecture; Technical Specification; Release 004 Version 07.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[147] M. Dryjański (RIMEDO Labs, Poznan, Poland). *Network Slicing in O-RAN*. Accessed: Mar. 17, 2025. [Online]. Available: https://tinyurl.com/remnsoran

[148] C. Li and A. Akman, "O-RAN use cases and deployment scenarios," O-RAN Alliance, Alfter, Germany, White Paper, Feb. 2020.

[149] P. Lédl et al., "O-RAN town: Piloting a high-power multivendor open RAN solution in a brownfield network," deutsche Telekom AG, Bonn, Germany, White Paper, Feb. 2023.

[150] F. Mehran, C. Turyagyenda, and D. Kaleshi, "Experimental evaluation of multi-vendor 5G open RANs: Promises, challenges, and lessons learned," *IEEE Access*, vol. 12, pp. 152241–152261, 2024.

[151] (O-RAN Alliance, Alfter, Germany). *Working Group 4; Management Plane Specification; Technical Specification; Release 004 Version 17.01*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[152] (O-RAN Alliance, Alfter, Germany). *Working Group 4; Control, User and Synchronization Plane Specification; Technical Specification; Release 004 Version 17.01*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[153] R. Holloway. "What exactly is O-RAN open fronthaul and its APIs?" Accessed: Apr. 10, 2025. [Online]. Available: https://tinyurl.com/ofhaul

[154] C. Adamczyk, "Challenges for conflict mitigation in O-RAN's RAN intelligent controllers," in *Proc. Int. Conf. Softw. Telecommun. Comput. Netw. (SoftCOM)*, 2023, pp. 1–6.

[155] J. Armstrong, E. Fallon, and S. Fallon, "Pre-emptive conflict detection architecture for O-RAN service management and orchestration," in *Proc. IEEE Int. Conf. Ind. 4.0 Artif. Intell. Commun. Technol.*, 2024, pp. 335–340.

[156] T. Chen, I. Chih-Lin, and T. Melodia, "O-RAN's role in shaping 6G: Industry perspectives on open and smart RAN," *IEEE Wireless Commun.*, vol. 32, no. 1, pp. 10–12, Feb. 2025.

[157] K. Kondepu, B. R. Tamma, and V. Gudepu, "O-RIDE: Open-RAN innovation and deployments exploration," in *Proc. 26th Int. Conf. Distrib. Comput. Netw.*, 2025, pp. 426–429.

[158] G. Brown, "The role of the RAN intelligent controller in open RAN systems," Heavy Reading, New York, NY, USA, White Paper, Oct. 2020.

[159] "Open radio access network architecture on AWS," Amazon Web Services, Seattle, WA, USA, White Paper, Dec. 2022.

[160] S. Köpsell et al., *Open RAN Risk Analysis*, Secunet, Essen, German, Feb. 2022.

[161] J. F. Santos, A. Huff, D. Campos, K. V. Cardoso, C. B. Both, and L. A. DaSilva, "Managing O-RAN networks: XApp development from zero to hero," *IEEE Commun. Surveys Tuts.*, early access, Feb. 7, 2025, doi: 10.1109/COMST.2025.3539687.

[162] B. Brik et al., "Explainable AI in 6G O-RAN: A tutorial and survey on architecture, use cases, challenges, and future research," *IEEE Commun. Surveys Tuts.*, early access, Dec. 2, 2024, doi: 10.1109/COMST.2024.3510543.

[163] S.-Y. Lien et al., "Universal vertical application adaptation for O-RAN: Low-latency RIC and autonomous intelligent xApp generation," *IEEE Commun. Mag.*, vol. 62, no. 5, pp. 80–86, May 2024.

[164] A. Elyasi, A. Ashdown, K. M. Rumman, and F. Restuccia. "O-RAN xApps: Survey and research challenges." Feb. 2025. [Online]. Available: https://ssrn.com/abstract=5237491

[165] M. Hoffmann et al., "Open RAN xApps design and evaluation: Lessons learnt and identified challenges," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 473–486, Feb. 2024.

[166] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms," *IEEE Trans. Mob. Comput.*, vol. 22, no. 10, pp. 5787–5800, Oct. 2023.

[167] (O-RAN Alliance, Alfter, Germany). *Working Group 3; Y1 Interface: General Aspects and Principles; Technical Specification; Release 004 Version 01.01*. Oct. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[168] H. Lee, Y. Jang, J. Song, and H. Yeon, "O-RAN AI/ML Workflow implementation of personalized network optimization via reinforcement learning," in *Proc. IEEE Globecom Workshops*, 2021, pp. 1–6.

[169] X. Lin, L. Kundu, C. Dick, and S. Velayutham, "Embracing AI in 5G-advanced toward 6G: A joint 3GPP and O-RAN perspective," *IEEE Commun. Stand. Mag.*, vol. 7, no. 4, pp. 76–83, Dec. 2023.

[170] (O-RAN Alliance, Alfter, Germany). *Working Group 2; Non-RT RIC Architecture; Technical Specification; Release 004 Version 06.00*. Oct. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[171] "O-RAN minimum viable plan and acceleration towards commercialization," O-RAN Alliance, Alfter, Germany, White Paper, Jun. 2021.

[172] "Open RAN update," 5G Americas, Bellevue, WA, USA, White Paper, Nov. 2023.

[173] (O-RAN Alliance, Alfter, Germany). *Working Group 2; A1 Interface: General Aspects and Principles; Technical Specification; Release 004 Version 05.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[174] S. Kumar, "AI/ML enabled automation system for software defined disaggregated open radio access networks: Transforming telecommunication business," *Big Data Min. Anal.*, vol. 7, no. 2, pp. 271–293, Jun. 2024.

[175] T.-H. Wang, Y.-C. Chen, S.-J. Huang, K.-S. Hsu, and C.-H. Hu, "Design of a network management system for 5G open RAN," in *Proc. Asia-Pac. Netw. Oper. Manag. Symp.*, 2021, pp. 138–141.

[176] (O-RAN Alliance, Alfter, Germany). *Working Group 1; Decoupled SMO Architecture; Technical Report; Release 004 Version 3.00*. Oct. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[177] (O-RAN Alliance, Alfter, Germany). *Working Group 1; O-RAN Operations and Maintenance Interface Specification; Technical Specification; Version 04.00*. Feb. 2021. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[178] (O-RAN Alliance, Alfter, Germany). *Working Group 6; Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN; Technical Report; Release 004 Version 08.01*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[179] (O-RAN Alliance, Alfter, Germany). *Working Group 6; O2 Interface General Aspects and Principles; Technical Specification; Release 004 Version 08.01*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[180] *5G; Management and Orchestration; Architecture Framework; Technical Specification; Release 17 Version 17.2.0*, 3GPP Standard TS 28.533, May 2022.

[181] "Unlock innovation and fuel business transformation with network slicing," Juniper Netw., Sunnyvale, CA, USA, White Paper, Oct. 2022.

[182] *ETSI GR NFV-IFA 046; Network Functions Virtualisation (NFV); Architectural Framework; Report on NFV Support for Virtualisation of RAN; Group Report; Release 5 Version 5.1.1*, ETSI, Sophia Antipolis, France, May 2023.

[183] *ETSI GS NFV 006; Network Functions Virtualisation (NFV); Management and Orchestration; Architectural Framework Specification; Group Specification; Release 4 Version 4.4.1*, ETSI, Sophia Antipolis, France, Dec. 2022.

[184] "open network automation platform (ONAP) architecture overview," ONAP, San Francisco, CA, USA, White Paper, Jul. 2019.

[185] B. Han et al., "DMH-HARQ: Reliable and open latency-constrained wireless transport network," Jul. 2024, *arXiv:2212.03602*.

[186] *Solution Brief; IxNetwork Test Solution for Oran Fronthaul Transport Network; Enable 5G Success with Robust xHaul Transport Infrastructure*, keysight Technologies, Santa Rosa, CA, USA, Sep. 2022.

[187] H. S. Shah and K. Nevatia, "5G ORAN deployment: Field insights and best practices for automated and orchestrated provisioning," Cisco, San Jose, CA, USA, White Paper, Feb. 2024.

[188] A. Garcia-Saavedra and X. Costa-Pérez, "O-RAN: Disrupting the virtualized RAN ecosystem," *IEEE Commun. Stand. Mag.*, vol. 5, no. 4, pp. 96–103, Dec. 2021.

[189] M. Dryjański. "O-RAN deployment scenarios." Accessed: Mar. 18, 2025. [Online]. Available: https://tinyurl.com/remorands

[190] *5G; System Architecture for the 5G System; Technical Specification; Release 18 Version 18.5.0*, 3GPP Standard TS 23.501, Mar. 2023.

[191] *5G; Management and Orchestration; Concepts, Use Cases and Requirements; Technical Specification; Release 17 Version 17.4.0*, 3GPP Standard TS 28.530, Apr. 2023.

[192] K. Park, S. Sung, H. Kim, and J. Jung, "Technology trends and challenges in SDN and service assurance for end-to-end network slicing," *Comput. Netw.*, vol. 234, Oct. 2023, Art. no. 109908.

[193] I. Badmus, A. Laghrissi, M. Matinmikko-Blue, and A. Pouttu, "End-to-end network slice architecture and distribution across 5G micro-operator leveraging multi-domain and multi-tenancy," *EURASIP J. Wireless Commun. Netw.*, vol. 2021, no. 94, p. 23, Apr. 2021, doi: 10.1186/s13638-021-01959-7.

[194] S. Pal and R. Armada, *Presentation; OSM #10 Ecosystem Day; OSM for 5G O-RAN*, ETSI, Sophia Antipolis, France, OSM, Arendal, Norway, 2020.

[195] M. A. Habibi et al., "Enabling network and service programmability in 6G mobile communication systems," in *Proc. IEEE FNWF*, Oct. 2022, pp. 320–327.

[196] *5G; NR and NG-RAN Overall Description; Stage-2; Technical Specification; Release 16 Version 16.4.0*, 3GPP Standard TS 38.300, Jan. 2018.

[197] C. Rotsos et al., "Network service orchestration standardization: A technology survey," *Comput. Stand. Interfaces*, vol. 54, no. P4, pp. 203–215, Nov. 2017.

[198] (ONAP, San Francisco, CA, USA). *E2E Network Slicing Use Case*. Accessed: Mar. 17, 2025. [Online]. Available: https://tinyurl.com/onapnsuc

[199] J. Ordonez-Lucena, P. Ameigeiras, L. M. Contreras, J. Folgueira, and D. R. López, "On the Rollout of network slicing in carrier networks: A technology radar," *Sensors*, vol. 21, no. 23, p. 8094, 2021.

[200] (O-RAN Alliance, Alfter, Germany). *Working Group 6; O2ims Interface Specification; Technical Specification; Release 004 Version 08.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[201] F. Salahdine, Q. Liu, and T. Han, "Towards secure and intelligent network slicing for 5G networks," *IEEE Open J. Comput. Soc.*, vol. 3, pp. 23–38, 2022.

[202] (O-RAN Alliance, Alfter, Germany). *Working Group 9; Xhaul Packet Switched Architectures and Solutions; Technical Specification; Release 003 Version 08.00*. Jun. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[203] H. Baba et al., "5G xHaul sharing as slice implementation with inter- and intra-operator orchestration," in *Proc. 6th IEEE Conf. Netw. Softw. (NetSoft)*, 2020, pp. 356–358.

[204] (O-RAN Alliance, Alfter, Germany). *Working Group 9; Xhaul Transport Requirements; Technical Specification; Version 01.00*. Feb. 2021. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[205] (O-RAN Alliance, Alfter, Germany). *Working Group 9; WDM-Based Xhaul Networks; Technical Specification; Release 004 Version 05.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[206] D. Camps-Mur et al., "5G-XHaul: A novel wireless-optical SDN transport network to support joint 5G backhaul and fronthaul services," *IEEE Commun. Mag.*, vol. 57, no. 7, pp. 99–105, 2019.

[207] S. Mondal and M. Ruffini, "Optical front/mid-haul with open access-edge server deployment framework for sliced O-RAN," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 3, pp. 3202–3219, Sep. 2022.

[208] Y. Rekhter and E. C. Rosen, "BGP/MPLS IP virtual private networks (VPNs)," Internet Eng. Task Force, RFC 4346, Feb. 2006.

[209] G. Dawra et al., "BGP overlay services based on segment routing over IPv6 (SRv6)," Internet Eng. Task Force, RFC 9252, Jul. 2022.

[210] C. Filsfils et al., "Segment routing policy architecture," Internet Eng. Task Force, RFC 9256, Jul. 2022.

[211] R. Rokui, H. Yu, L. Deng, D. Allabaugh, M. Hemmati, and C. Janz, "A standards-based, model-driven solution for 5G transport slice automation and assurance," in *Proc. 6th IEEE Conf. Netw. Softw. (NetSoft)*, 2020, pp. 106–113.

[212] "Transport networks for 5G," 5G Americas, Bellevue, WA, USA, White Paper, Aug. 2023.

[213] *GSTR-TN5G—Transport Network Support of IMT-2020/5G; Technical Report*, Int. Telecommun. Union, Geneva, Switzerland, Feb. 2018.

[214] (O-RAN Alliance, Alfter, Germany). *Working Group 9; Management Interfaces for Transport Network Elements; Technical Specification; Release 004 Version 09.00*. Oct. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[215] K. G. Szarkowicz et al., "A realization of network slices for 5G networks using current IP/MPLS technologies," Internet Eng. Task Force, Internet-Draft draft-ietf-teas-5g-ns-ip-mpls-07, Nov. 2023, work in Progress.

[216] A. Farrel et al., "A framework for network slices in networks built from IETF technologies," Internet Eng. Task Force, RFC 9543, Sep. 2023.

[217] *5G; Management and Orchestration; Provisioning; Technical Specification; Release 16 Version 16.6.0*, 3GPP Standard TS 28.531, Aug. 2020.

[218] "E2E network slicing architecture, version 2.0," GSM Assoc., White Paper, London, U.K., May 2022.

[219] (O-RAN Alliance, Alfter, Germany). *Working Group 6; Cloudification and Orchestration Use Cases and Requirements for O-RAN Virtualized RAN; Technical Specification; Release 004 Version 12.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[220] N. Aryal et al., "Moving towards open radio access networks with blockchain technologies," in *Proc. 5th Conf. Blockchain Res. Appl. Innov. Netw. Services (BRAINS)*, 2023, pp. 1–9.

[221] (O-RAN Alliance, Alfter, Germany). *Working Group 1; Use Cases Analysis Report; Technical Report; Release 004 Version 16.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[222] J. Groen et al., "Implementing and evaluating security in O-RAN: Interfaces, intelligence, and platforms," *IEEE Netw.*, vol. 39, no. 1, pp. 227–234, Jan. 2025.

[223] (O-RAN Alliance, Alfter, Germany). *Working Group 10; O-RAN Operations and Maintenance Architecture; Technical Specification; Release 004 Version 14.00*. Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[224] E. Kirstel. "Open RAN challenges and testing needs for real-world performance." Accessed: Mar. 17, 2025. [Online]. Available: https://tinyurl.com/SpOranCh

[225] Y. Gopikrishna. "5 key lessons from integrating and deploying open RAN." Accessed: Mar. 18, 2025. [Online]. Available: https://tinyurl.com/OpenRANDep

[226] F. A. Bimo et al., "OSC community lab: The integration test bed for O-RAN software community," in *Proc. IEEE FNWF*, 2022, pp. 513–518.

[227] P. Curwen and J. Whalley, "Will open RAN remain open?: A regular column on the information industries," *Digit. Policy Reg. Gov.*, vol. 23, pp. 642–644, Dec. 2021.

[228] N. Aryal, E. Bertin, and N. Crespi, "Open radio access network challenges for next generation mobile network," in *Proc. 26th Conf. ICIN Workshops*, 2023, pp. 90–94.

[229] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "Intelligent and QoE-aware open radio access networks," in *Proc. 3rd URSI AT-AP-RASC Meeting*, 2022, pp. 1–4.

[230] J. M. Batalla et al., "On deploying the Internet of Energy with 5G open RAN technology including Beamforming mechanism," *Energies*, vol. 15, no. 7, p. 2429, 2022.

[231] P. Porambage, M. Christopoulou, B. Han, M. Asif Habibi, H. Bogucka, and P. Kryszkiewicz, "Security, privacy, and trust for open radio access networks in 6G," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 332–361, 2025.