A Survey on Open Radio Access Networks: Challenges, Research Directions, and Open Source Approaches

Wilfrid Azariah *, Fransiscus Asisi Bimo*, Chih-Wei Lin*, Ray-Guang Cheng*, Rittwik Jana†, Navid Nikaein‡,
* Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taiwan
† Google LLC, USA
‡ Department of Communication Systems, EURECOM, France

***Abstract*—The open radio access network (RAN) aims to bring openness and intelligence to the traditional closed and proprietary RAN technology and offer flexibility, performance improvement, and cost-efficiency in the RAN deployment and operation. This paper provides a comprehensive survey of the open RAN development. We briefly summarized the RAN evolution history and the state-of-the-art technologies applied in open RAN. The open RAN-related projects, activities, and standardization is then discussed. We then summarize the challenges and future research directions required to support the open RAN. Finally, we discuss some solutions to tackle these issues from the open source's perspective.**

***Index Terms*— Open RAN, O-RAN, Open Source**

I. Introduction

Radio access network (RAN) is one of the main components in cellular networks [1]. The RAN connects the user equipment (UE) to the core network (CN). The RAN component has evolved throughout the years as a solution to the growing number of subscribers and rising user demands. The first version of RAN is distributed RAN (D-RAN), followed by centralized RAN (C-RAN), and lastly, virtualized RAN (vRAN). However, the vRAN solution hardly meets the expectations of the current 5G network requirements. The next generation of RAN is looking to opening the the interfaces in the RAN ecosystem.

The cellular network industry offers the next generation of RAN as a solution to fulfill 5G network requirements. Open RAN is proposed as an evolution of the vRAN [2]. To answer the challenges faced by vRAN, open RAN breaks the closed nature of the previous RAN generations. Open RAN promises to deliver the expected quality of service (QoS) and quality of experience (QoE) of the latest 5G network requirement while also preventing the economic expenditure to skyrocket. While advancing the previous RAN generations, open RAN still faces challenges to this very day.

Motivation of this work: Open RAN movement and its goal to make an open-interface system has existed for these past few years. Even though open RAN's development has been around for years, there is still a lack of reference or research that covers information about open RAN holistically. The RAN advancement history and the open RAN concept can be found in [1] and [2]. Open 5G network projects from the RAN and CN side, as well as its problems are explained in [3]. The O-RAN Alliance architecture is summarized in [4], but it does not include the bigger picture of the open RAN movement. O-RAN Alliance architecture is also discussed in [5] and [6], including its advantages and limitations that could be addressed in future research. But still, these references could not serve the complete picture of open RAN's history, present condition, and what opportunities open RAN can open up in the future.

Contributions of this work: Through this survey, we explain in detail about open RAN technology, the further evolution of vRAN as written above. We will first provide the bigger picture of the open RAN movement, including the most current landscape of open RAN in terms of projects and activities. Then, this paper will describe the open RAN components and their implementation. We will narrow down the study to focus on O-RAN Alliance standards and architecture as a reference. Also, this paper will explain what challenges are faced in open RAN and future research possibilities for the next-gen technology. Besides that, this paper will suggest some solutions to open RAN challenges based on open source approaches.

Organization of this paper: The rest of this survey paper is structured into three main parts. First of all, we explain why we need open RAN from sections II to V. Then, we list the challenges and research directions for the open RAN in section VI. After that, we address some issues of challenges and research directions from open source's perspective in section VII. Section II summarizes the RAN evolution history. Section III provides the background of the open RAN movement and the technologies related to the open RAN field. Section IV reviews the projects, activities, and standards related to open RAN. Section V discusses the O-RAN Alliance architecture. Section VI describes the challenges and future research directions of the open RAN field and the RAN technology in general. Section VII will address some of the challenges in section VI from open source perspective. Our survey's broad outline is shown in Fig. 1.

II. Evolution from Traditional RAN to vRAN

This section will summarize the evolution history of RAN. The real-world implementation of the RAN is better known as the Base Station (BS) [1]. There are two major units of a BS, namely Radio Unit (RU) and Baseband Unit (BBU). RU is

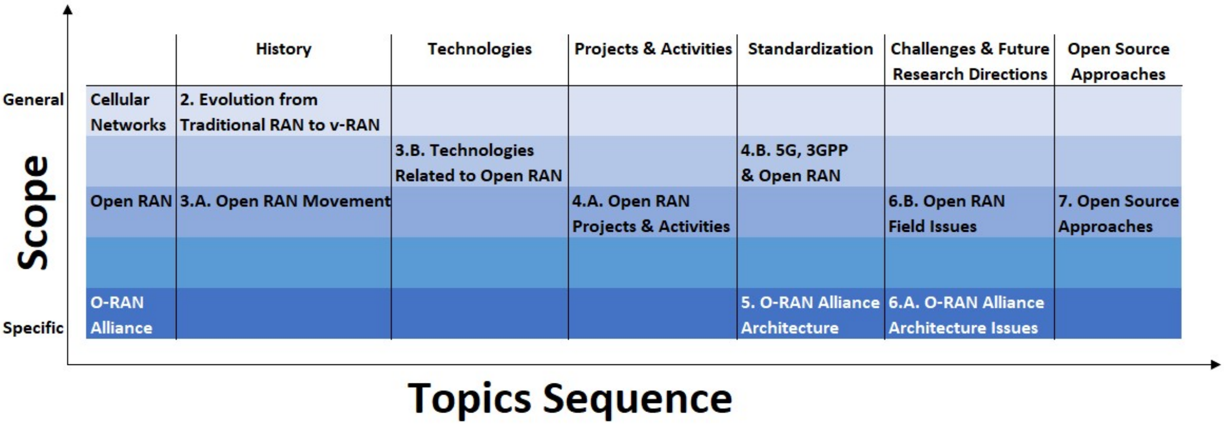

Fig. 1. Overall organization of our survey on open RAN

responsible for transmission and reception. Meanwhile, BBU is responsible for radio management, resource utilization, and other operations.

A. Traditional RAN or D-RAN

The first version of RAN was equipped with an integrated system between RU and BBU. BBU was usually installed in a room, right below BS. RU could either be installed in the room or at the top of a tower, with the purpose to enable the RU to support connectivity in a large area [1]. In this context, the RU can also be called the remote RU (RRU). Either way, the distance between RU/RRU and BBU is short. This traditional RAN framework was called D-RAN [7]. D-RAN is simple to implement because a high-speed interface between RU and BBU is not needed. Each RAN is independently operated. As the number of UEs increased, the network densified, and more BS are built. Cellular network providers spent a significant amount of money to rent the BS space and also the cooling system to operate the network. Thus, began the search for a solution to reduce the Operating Expenditure (OpEx) and the C-RAN framework is proposed.

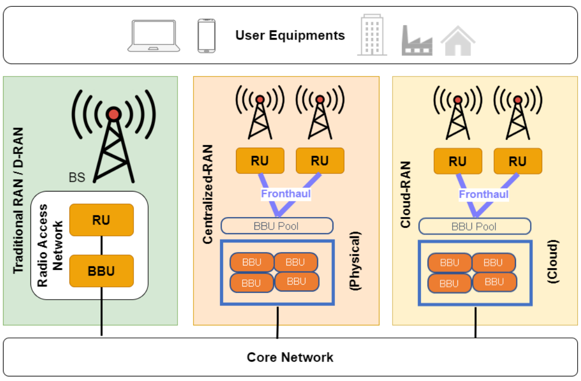

Fig. 2. RAN architecture per generation

B. C-RAN

The "C" in C-RAN could be an abbreviation of "Centralized" or "Cloud". However, both have the same idea, which is to cluster the BBUs into a pool [2], [8]. Initially, Centralized RAN is proposed to reduce the space rental cost and the air conditioner power consumption of the BBU by pooling BBUs from different BS into a single physical location. The fronthaul (FH) interface is introduced to connect the BBU pool and RRUs. Fig. 2 shows how the FH interface (FHI) connects one BBU pool to many RRUs. Despite reducing the overall OpEx, Centralized RAN requires the FH to have high bandwidth and low time latency requirement [9], [10].

The real difference between Centralized RAN and Cloud RAN is the cloud system. In centralized RAN, the BBUs are pooled in a physical location. In Cloud RAN, BBUs from each BS were pooled in a cloud server [1]. Cloud RAN is superior because the cloud control in Cloud RAN made it easier for the number of BBUs to be changed with time. The cloud also increased the baseband processing by exploiting general-purpose processors [2]. Cloud RAN further reduced energy consumption, increased network throughput, improved network scalability, reduced Capital Expenditure (CapEx), and reduced OpEx [7], [11].

The C-RAN had been acknowledged as one of the most potential technology that is able to fulfill the 5G technical requirements in radio access [2]. However, it still has limitations, such as the huge FH overhead, trust problems, security problems, and single point failure. These problems had forced the C-RAN framework to shift its focus to more advanced computing technologies. Some of these advanced technologies were virtualization and edge computing. Virtualization is the key to the shifting from C-RAN to vRAN.

C. vRAN

Virtualization means creating virtual instances over abstracted physical hardware [2]. In the context of vRAN, the virtual instances are network resources. Virtualization in RAN is heavily related to concepts such as software-defined networking (SDN) and network function virtualization (NFV) [7]. Simply, vRAN brings virtualization to Cloud RAN. In the cloud server, multiple virtual BBUs (vBBUs) are deployed. vBBU can be deployed on either a Virtual Machine (VM) or container. This system enables orchestration and resource scaling in vRAN. Ease to scale up and down network resources led to lower energy consumption, dynamic capacity scaling, efficient use of network resources, improved service reliability, and better service quality [2], [12]. Approximately, 50% of data-processing resources required can be reduced when virtualization is applied on Cloud RAN [13].

Overall, vRAN minimizes the operational and investment costs for network operators. However, vRAN has to deal with the properties of the wireless channel. Network resources have to be shared and distributed to different RRUs fairly and efficiently while taking QoS requirements into consideration. This made the vRAN system to be significantly complex. vRAN also still uses proprietary interfaces that prevent interoperability and a multivendor environment. As a result, vendor lock-in and monopoly prevent the network equipment price to be cheaper.

III. Open RAN Movement and the Related Technologies

In this section, we will introduce the history of the open RAN movement and the technologies used to build open RAN. The open RAN term had been proposed since May 2002 [14]. However, the current widely known open RAN term is an industry movement in wireless telecommunications to develop disaggregated and interoperable RAN [15]. Disaggregated and interoperable are the two fundamental pillars of open RAN. The two fundamental pillars will give a great benefit to us because open RAN's architecture will enable a multivendor environment and can run on general-purpose processors. Open RAN should be supported by several technologies. These technologies will be explained deeply in the next subsections.

A. *Open RAN Movement*

The growing cellular network market demand forces operators to optimize each component of the network, including the RAN. There are several challenges that should be done to deploy a good and reliable RAN. One of these challenges is the need to make a flexible network, resulting in a customized QoS. 5G introduces new 3 use cases domain, namely Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), and Massive Machine Type Communication (mMTC) [16]. Every application and every wireless device has its own specific QoS requirements [1]. The current monolithic RAN could not satisfy the diverse need of these use cases. If we force all of these use cases to connect to one specific network all the time, it will reduce the QoS and the QoE [17].

Network upgrade is needed to fulfill these requirements. The existing RAN should be more flexible and more adaptable. Better flexibility and adapting technique will enable the RAN framework to fulfill those QoS requirements in each application. This flexibility can only be achieved by disaggregating RAN components [1]. The RAN framework should be disaggregated or to be split based on every application's functionality. This enables the RAN to behave differently to each particular QoS requirement of each application, and this will lead to a smarter and more versatile RAN. RAN's architecture should be capable of supporting multiple levels of QoS in handoff scenarios at one time [14].

Another challenge for RAN is the need to combine the existing NFV with artificial intelligence (AI). The trend of NFV has existed since the development from C-RAN to vRAN. However, the virtualization concept proposed in vRAN is complicated, needing mechanisms for management and orchestration [2]. Management and orchestration can be done by leveraging AI. RAN needs NFV with AI embedded in it. As we know, AI has impacted the computing and networking world significantly until today [18]. AI can be used to analyze the enormous amount of data generated by the RAN. Then, the information gathered can be used to anticipate problems in the network and take necessary action to ensure network QoS is delivered to the users. AI can make the NFV system smarter, more optimized, and improved. Further explanation about NFV can be seen in the next subsection.

It is already stated that RAN's improvement is happening with the introduction of cloud-native principles. We shall develop on the cloud-native approach by also improving RAN's software and hardware. All of the previous generations of RAN had one similar weakness: operators lack control toward innovation in their software and hardware. Before open RAN, RAN's software and hardware were proprietary and remained optimized to one vendor only [15]. This caused an economical monopoly, extreme overhead, and led to unsupervised systems because the only vendor that can supply both the hardware and software was that one vendor. Operators are heavily dependent on vendors to provide innovation and new features on RAN technologies. Let's take AI and NFV in the previous paragraph as an example. AI embedded software is needed to complete the NFV advancement process. However, operators could not implement AI-embedded NFV because they are dependent on whether their partner vendors offer this solution or not. So, before operators can freely innovate their services, operators must regain control of their network components.

Last but not least, the challenge for RAN that still remains is the lack of efficient interconnection between each other parts of the network architecture [17]. This was caused by the uneven standards that used for cellular networks around the globe due to the proprietary interfaces used. Because of these different characteristics of each vendor's interfaces, hardware and software from different vendors could not interconnect easily. This led to plenty of waste of wireless infrastructures and spectrum resources. When operators try to ignore these differences and treated all the cellular networks the same, their users will experience low QoS and QoE of the network.

These challenges can be solved by utilizing the two fundamental pillars of open RAN in its architecture. As written before, the fundamental pillars of open RAN are disaggregation and interoperability by means of standardized interfaces. The interoperable of open RAN means that all of open RAN's architecture's interfaces are open [3]. The open RAN's requirements defined by Mobile Wireless Internet Forum (MWIFs) state that RAN vendors should provide an architecture that supports operation, administration, and network management based on open interfaces and Internet Engineering Task Force (IETF) network management protocols [14]. Basically, MWIFs require every telecommunication company applying open RAN in their connectivity to follow such architecture based on open interfaces and IETF standards.

Besides interoperability, another fundamental pillar of open RAN is disaggregation. Generally, disaggregation means the software is detached from its hardware [15]. Vendor develops their software to be compatible with other vendors' hardware. Through disaggregation, RAN software can run in general-purpose processors and further extends its interoperability. RAN's software can now run on any vendor's hardware. Because softwarization is done in RAN, acceleration techniques in software development can also be implemented, such as continuous integration and continuous deployment (CI/CD). In turn, the time to market and deployment time of RAN can be reduced. Disaggregation and softwarization will be explained in further detail in the next subsection.

Open RAN has open interfaces in its architecture and this has become one of open RAN's fundamental pillars. Open interface has a slightly different meaning than open source. Interfaces are boundaries that are shared between two or more separate components of a system. These boundaries are used in a system to exchange information from one to another component. In the RAN context, interface means the connections between specific parts of the RAN, such as the FHI. Open interface means that these connections are standardized and the specifications can be freely seen by everyone. One of the most notable examples of an open interface is the internet where everyone uses the same protocol to communicate with each other. There are other terms that have the same meaning as the "open interface" terms, such as "open API", "open standard", "open technology", or even just "open". When a vendor declared that their solution is "open", it usually means that the vendor offers an open interface.

Open source has source code as its subject. Source code is a pack of codes that are usually written with programming languages where programmers create or modify their software. Open-source software has an open and non-proprietary implementation. It has an Open Source Initiative compatible license that allows anyone who has the software's source codes to re-distribute, modify the software, or even derive codes from the software [19]. Popular examples of open-source software are Linux operating system and Mozilla Firefox. In the RAN context, open-source code RAN software means that the software source code to run the RAN network function can be accessed by everyone. Open source RAN software are rare because RAN software is the products being sold by vendors.

The implementation of open RAN has given us a lot of advantages. Open RAN can solve its predecessors' problems, thanks to its two fundamental pillars. Until this part, there have been many advantages of open RAN that have been conveyed in this paper. We will highlight the main advantages that open RAN possess. Summarizing all of the advantages written above, simply we can say that the open RAN framework has two key target advantages, which are improving efficiency in its performance and its cost.

The openness pillar has made open RAN's performance improved. It has also made open RAN easier to configure and modify. Open RAN deletes the heterogeneous network connections. This makes each network connection can interconnect easily and efficiently, which leads to increased QoS and QoE; and a reduced amount of wireless infrastructures and spectrum resources that are put to waste. The proof of this performance can be seen from Rakuten, which has built the world's first and currently the only commercial-scale, also the only nationwide network that uses open RAN standards and architectures [20]. The proof was supported by Umlaut who has evaluated several performance metrics by comparing Rakuten Mobile data in Tokyo to other operators in major cities around the world. The report is decidedly positive, with a score of 920 out of 1.000, and a "very good" rating. Compared to other cities, Rakuten is able to provide connectivity at a much higher speed than other mobile operators. This proves that the application of open RAN in network architecture can improve network connectivity's performance.

Besides improving performance, open RAN can also make the network connectivity be more cost-efficient. Because the framework includes AI in its architecture, the open RAN framework will be able to facilitate automated operational network functions [2]. Open RAN is built to perform some complex tasks that can never be done before by its predecessor. The existence of machine learning (ML) and AI reduces RAN dependency on human force, thereby reducing operational costs. Cost reduction will be much more significant in the future with the development multivendor environment. The disappearance of a monopoly system between hardware and software in open RAN framework makes its cost will be more efficient. Operators can change a particular part that needs an upgrade or is broken without needing to change the whole network [15].

Rakuten serves as the proof. Rakuten Mobile 4G tech's CapEx and OpEx are approximately 40% and 30% lower than traditional RAN deployment's expenditures [20]. Furthermore, it is also mentioned that the cost spent by Rakuten is much more efficient when deploying a 5G network, saving up to 50%. This cost efficiency can benefit rural regions to enjoy economical network connectivity. These two advantages can be reached through open RAN's openness and intelligence fundamental pillars. By solving the previous RAN problems, the network's performance will be improved and its cost will be more efficient.

The development of open RAN continues until the time this paper was written. The progress milestone that open RAN has achieved until this year is shown in Fig. 3. In summarizing the milestone, we compile information from announcements

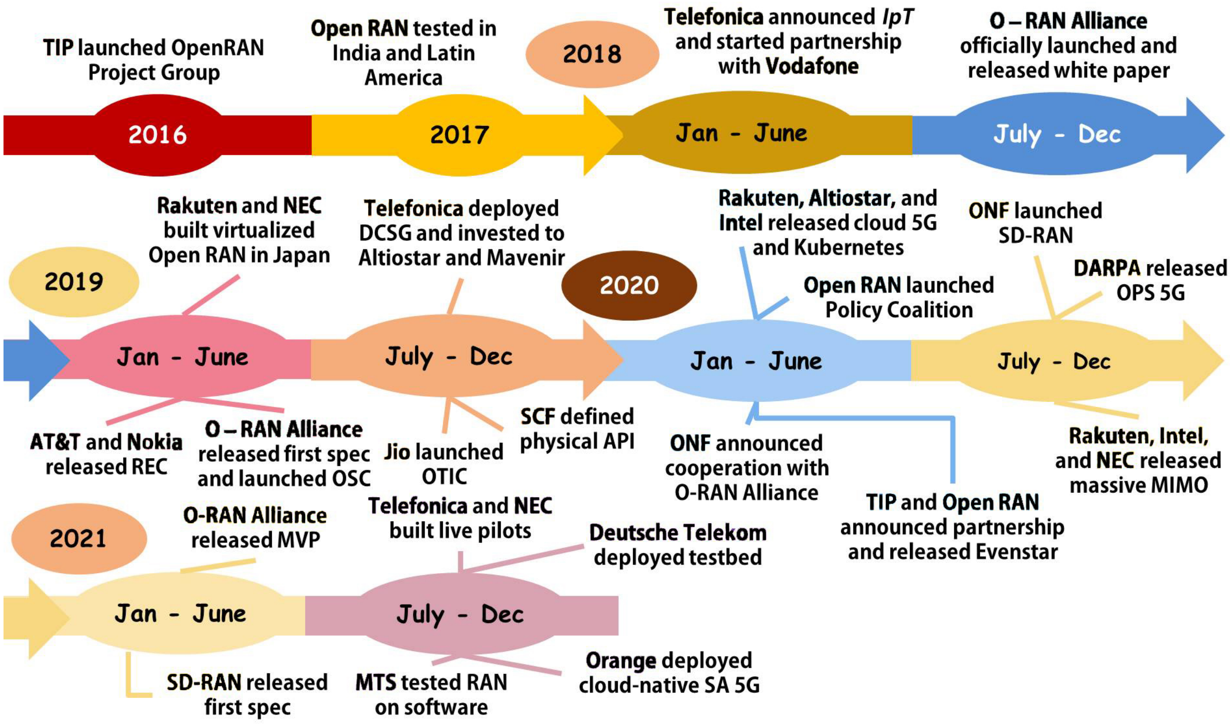

Fig. 3. Open RAN's milestone

and reports from many press releases; white papers; news; and annual reports [15], [20] - [25].

Open RAN started to develop rapidly in these last six years. The first development of open RAN occurred in February 2016, when Telecom Infra Project (TIP) formed the OpenRAN Project Group [15]. Since then, TIP has brought together more than 500 mobile network operators (MNOs), suppliers, vendors, developers, and integrators that are using open source technologies and open approaches [15], [21]. Further details about TIP OpenRAN project group are written in the next section. A year after the formation of the community, the first trials of open RAN started in India and Latin America [15]. Starting from 2018 until 2021, communities and companies all over the world have done more open RAN's deployments. These are the brief descriptions of open RAN movement development in half-year time period.

1) **Jan. - Jun. 2018:** At Mobile World Congress (MWC) 2018 in Barcelona, Telefonica announced that they released a project called as *Internet para Todos* (IpT), which means *internet for all*. IpT aims to deploy network connectivity in several Latin American countries' rural areas [15]. The Spanish telecommunication company also declared its collaboration with Vodafone to deploy a joint Request for Information (RFI) system, which was made to evaluate software-based technologies for open RAN, running on top of commoditized hardware.
2) **Jul. - Dec. 2018:** O-RAN Alliance released its first white paper ever [22]. The white paper contains O-RAN's vision to drive the development of an open and intelligent RAN. A month later, O-RAN Alliance was officially launched with the help of Linux Foundation (LF). Since then, O-RAN Alliance has been working officially to fulfill its purpose, which is to create standardizations of the open RAN. Further information about the O-RAN Alliance and its standardizations will be explained in the next sections.
3) **Jan. - Jun. 2019** O-RAN Alliance released its first specification. The first specification consisted of open FH, including control plane (CP), user plane (UP), synchronization plane, and management plane. They also collaborated with LF to launch O-RAN Software Community (OSC), with a mission to create open-source software for RAN. Rakuten also announced the deployment of the world's first virtualized, cloud-native 4G network, and highly secured cloud-native 5G network [15], [21]. In this deployment, Rakuten collaborated with other companies such as Altiostar, Nokia, Cisco, Airspan, and NEC Corporation to provide hardware, software, and virtualized cores to build the network's open interface and open vRAN architecture in Japan [15], [21], [23]. On the other hand, AT&T and Nokia released Radio Edge Cloud (REC) appliance [21].

4) **Jul. - Dec. 2019:** Small Cell Forum (SCF) enabled small cells in the open RAN ecosystem by stating a clear definition of the physical API that provides an open interface between 2 layers [15]. In India, Reliance Jio Infocomm released Open Test and Integration Center (OTIC) with other global operators, vendors, and integrators [21]. The center was made to help original equipment manufacturers (OEMs) and other software developers test their products' alignment with the O-RAN Alliance's specs. Telefonica also commercially released the TIP-incubated-Disaggregated Cell Site Gateways (DCSG) ecosystems in Germany and Ecuador.
5) **Jan. - Jun. 2020:** Open Networking Foundation (ONF) started to work together with O-RAN Alliance [21]. TIP's OpenRAN released the Evenstar program, which focuses on building RAN reference designs for 4G and 5G networks [15]. In the same month, O-RAN Alliance and TIP announced a liaison agreement about open RAN's further development [15], [21]. Rakuten, Altiostar, and Intel released cloud-native, container-based 5G RAN and Kubernetes plugins [21]. Rakuten also launched the world's largest vRAN network [24]. A group of 30 major companies release the open RAN policy coalition to promote open RAN technology to policymakers, with the main purpose to educate governments about what open RAN is and its benefits in network connection [15], [21].
6) **Jul. - Dec. 2020:** ONF announced the formation of a new project group called Software Defined-RAN (SD-RAN), focusing to create open-source software platform and multi-vendor solutions for 4G and 5G RAN deployments [21]. Rakuten also deployed its world's first 5G massive multiple input multiple output (MIMO) with O-RAN Alliance's specs compliant RRU [21]. The Defense Advanced Research Project Agency (DARPA) of the USA released an open, programmable, and secure (OPS) 5G network, with Cryptography for Hyper-scale Architectures in a Robust Internet of Things (IoTs) (CHARIOT), a new type of IoTs [21].
7) **Jan. - Jun. 2021:** SD-RAN released its first software specification for open RAN. The version 1.0 specification is a cloud-based exemplar platform for the open RAN that is suitable for the O-RAN Alliance's architecture. O-RAN Alliance also introduced a Minimum Viable Plan (MVP) for its architecture and its application in commercial networks [25]. The O-RAN's MVP was written in [25].
8) **Jul. - Dec. 2021:** Orange released the 5G standalone (SA) network with end-to-end (E2E) experimental cloud network. This is the first 5G SA network that is deployed in Europe, with the first planned testing done in France. Deutsche Telekom activated an open RAN testbed, the first testbed deployments done in Europe, specifically in Germany, featuring massive MIMO radios. Telefonica and NEC Corporation built open RAN live pilots in four global markets: Spain, Germany, UK, and Brazil. The Russian company Mobile TeleSystem (MTS) tested open RAN architecture in domestic software developed by Skolkovo Institute of Science and Technology.

B. Technologies Related to Open RAN

To make an open and intelligent RAN architecture, open RAN is supported by several technologies or approaches. Disaggregation, SDN, NFV, functional split, cloudification, automation, intelligence, network slicing, open source, and mobile edge computing (MEC) are some of them. These technologies and approaches will be introduced briefly in this subsection. Please look into the original reference for further understanding of the topic.

RAN disaggregation has several definitions. It can mean the disaggregation between RAN's software and hardware or the vertical and horizontal RAN disaggregation [26], [27]. The former is a separation process between the network's software and hardware that occurred in a network architecture. As explained before, RAN's software and hardware used to be proprietary and integrated. Through disaggregation, the term RAN's functionality softwarization emerges where the software does not depend on specific hardware and vice versa. The main differences between non-disaggregated and disaggregated network architecture are shown in Fig. 4. Software used to be integrated with hardware, whether it is in the CN or RAN. In a disaggregated network architecture, software and hardware are now separated. Through softwarization, further separation of control and data is done, thus the the term SDN emerges [3], [28], [29].

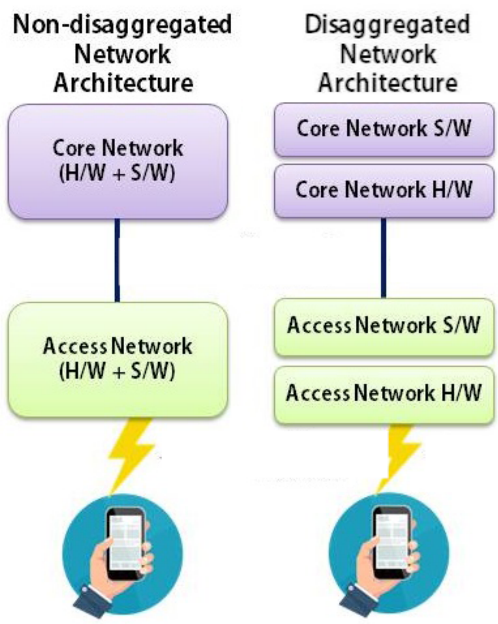

Fig. 4. Comparison between a non-disaggregated and disaggregated network architecture

Vertical and horizontal RAN disaggregation is also known as functional splits. Horizontal functional split is a selection process of the appropriate centralization level in RAN framework [30]. It separates the integrated BBU into two separate units: Central Unit (CU) and Distributed Unit (DU). Horizontal functional split also refers to the DU and RU separation. The degree of centralization in the horizontal functional split is flexible [31]. However, trade-offs should be considered when choosing these functional split options [10]. On the other hand, the vertical split is the separation between the CP and UP of the RAN. The CP and UP splitting (CUPS) is the extension of the SDN concept. The difference between vertical and horizontal split can be seen in Fig. 5. Further explanation about functional splits will be discussed in section IV.

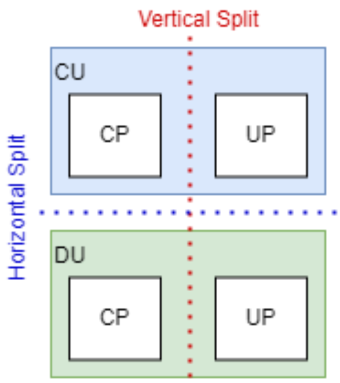

Fig. 5. Vertical and horizontal functional split of open RAN

Virtualization enhances softwarization [32]. Virtualization enables the softwarized RAN to be sub-divided into smaller parts within single hardware. There are two virtualization technologies, which are hypervisor-based and container-based virtualization [2]. Because of virtualization, a hardware is able to contain functions of many virtualized units, which leads to a new term called NFV. NFV virtualizes all network services and functionalities, such as virtual CU or virtual DU. These virtualized functionalities are called Virtual Network Functions (VNFs) and they run on top of VMs. NFV uses hypervisors named Virtual Machine Monitor (VMM) or virtualizer. The VMM hosts and runs Virtual Machines (VMs) that host VNFs.

Another related technology is called network slicing. Network slicing branches from NFV. Network slicing is the concept to slice or partition the physical network to form virtual resources [3], [33]. Slicing is done in all parts of the network. Through the network slicing, there are multiple virtualized network that lies in the same physical network. Each virtualized network can flexibly be allocated for different use cases requirement in 5G. Different services can be allocated to each of the slices to satisfy specific user needs. Network resources can also be allocated dynamically according to each slice's needs. Network slicing contradicts the former "one size fits all" service model.

Disaggregated RAN allows its software to operate on different hardware. Network operators can use commercial off-the-shelf (COTS) server or a cloud instead of using dedicated hardware to run their software. This condition is called cloudification, where the hardware's roles are replaced by the cloud or COTS. Network operators might choose to use cloud or COTS because the deployment cost is cheaper than investing in their own hardware to run their software. There are also several types of cloud deployment. Choosing the right cloud deployment is the first step a network operator should make in cloudification. Network operators might want to choose public cloud deployment, where the cloud resources are owned and operated by a cloud service provider. A public cloud may have lower costs and be easier to manage, but the network operator has less control over the cloud resources and the level of security might be low. Network operators can also choose private cloud deployment, where the cloud resources can only be used exclusively by one business or organization. Private cloud deployment allows for more flexibility, customization, and control. However, it requires advanced technical skills and is more costly. Network operators can also choose the hybrid cloud deployment, where data and functions move back and forth between private and public environments.

Since there is a cloud deployment option, operators and network providers can choose different approaches to provide their service, namely brownfield or greenfield strategy. In greenfield strategy, they have to make their network from the ground, which means that the operators should deploy the software and hardware from scratches. They do not use existing third-party cloud infrastructure. After they do the groundwork, they move those network components to the new cloud infrastructure. The greenfield strategy is the opposite of the brownfield strategy, where many of a network's functions of the previous architecture are retained. Simply, it means that in a brownfield project, the operator needs to upgrade or add new features to an existing cloud network and uses some legacy cloud components. Brownfield projects are mostly done after a greenfield project, with the purpose to develop or improve an existing application infrastructure. Operators and network providers should consider trade-offs between the two approaches.

Through softwarization, the RAN's characteristics and behaviors are easier to reprogram, thus making the RAN programmable. The advantages of open RAN programmable behavior is its network performance will be more optimized, the network resources can be dynamically allocated, the software and hardware functions can be controlled automatically, and novel algorithm can be leveraged to improve their own networking system performance [24]. The further advanced tech of programmability is called automation, where characteristics and behaviors of RAN are all reprogrammed automatically not by humans, but by computer programs.

There are two types of automation, namely orchestration and management. Orchestration is the automation of softwarized RAN deployment process. Management is the automation of RAN monitoring, configuring, coordinating, and task managing. The management system of RAN is called radio resource management (RRM). From RAN management, there is a new term called self-organizing network (SON). With SON, the RAN is requested to do its own self-configuration, which includes new deployment of nodes, performance optimization, and fault management [34]. The intelligence of RAN is proved through the presence of AI/ML embedded into the RAN. AI/ML turns RRM into a more intelligent one, namely RAN Intelligent Controller (RIC) [2]. The RIC facilitates the optimization of RAN through closed-control loops between RAN components and their controllers [35]. RICs have made the open RAN more adaptable, more effective, and also, more penny-wise. As shown on Fig. 6, these loops are operating at different times, ranging from 1 millisecond to thousands of milliseconds. Real-time RIC (RT RIC) loop works for less than 10 milliseconds. Near-RT RIC loop time range is between 10 milliseconds to 1 second. The Near-RT RIC is responsible for deploying AI-enabled optimization and giving feedback for UEs and cells. Non-RT RIC loop time range is one second or more. Non-RT RIC is responsible for releasing AI-enabled service policy and running analytics for the entire RAN.

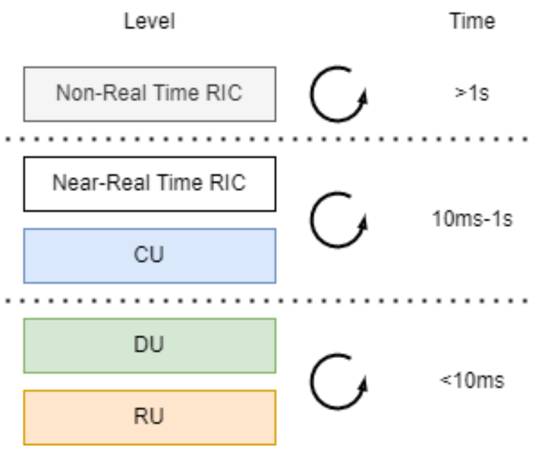

Fig. 6. Three types of RIC

We have already introduced the open source term. Even though not mandatory like open interface, open source accelerates improvement of the network infrastructure. Formerly, RAN software has been largely proprietary and developed for specific hardware. The virtualization concept introduced by vRAN caused a significant change in the way network systems were designed. This change has significantly increased the use of server-based platforms and virtualized network functions in vRAN, and also in open RAN, thus opening new chances for companies and network vendors to rebuild their RAN hardware and software to server-based platforms based on open-source software. Since then, there have been many open-source software developments, such as OpenStack and Kubernetes. Simply, open-source software has become a major solution to overcome challenges for RAN's software infrastructures. Its big contribution to RAN's software caused O-RAN Alliance and other communities and groups to develop open source solutions and adopted them to open RAN's significant functions. Although these open source solutions are still in the early deployments stage, the telecommunication industry is mostly expecting companies to use open source based software infrastructure solutions to run RAN applications.

Another related RAN technology is the smart network interface card (NIC). A smart NIC is a programmable accelerator that can increase the efficiency and flexibility of the data center network, data security, and data storage. Smart NIC can offload computation from its host processor [36]. Smart NIC can transparently unpack virtual switching data path processing for networking functions, such as network overlays, network security, load balancing, and telemetry. Through the use of Smart NIC or accelerator cards, a regular server performance and latency can be enhanced to meet the requirements needed to function as vBBU. Xilinx T1, T2, and T3 Telco Accelerator Card are some example of accelerators.

The other technology related to open RAN is MEC. MEC is a cloud service that runs at the edge of a network. It performs several specific tasks that would be processed in centralized core or cloud infrastructures. MEC moves the computing process closer to the users for enabling applications and services requiring some specific network characteristics that differ from other applications or services. MEC is capable to move content and functionalities to the edge; providing any private cellular network services by using its localized data process; deploying computational offloading to IoT devices; leveraging the proximity of edge devices to end users; and enhancing the privacy and security of mobile applications [3], [37]. Until today, there are several researches remaining for MEC, some of them are binary and partial offloading for MEC; MEC resource management system; MEC-open RAN-network slicing integration and their combined orchestrators; and MEC-embedded-RIC or called inter-near-RT RIC [34], [37].

Until this part, we have already known that open RAN is related to so many technologies. These technologies related to open RAN can improve costs in the future 5G network [38]. The term used to measure the total cost spent for deploying a network is called total cost of ownership (TCO). Simply, TCO is the sum of CapEx and OpEx needed when deploying a network. The reports also mentioned that the average TCO reduced from implementing these related technologies is 20%, if we compare it to 4G's TCO. This savings can be much higher than just 20%, because 5G network with these related technologies can give many business benefits that really significant for any business.

1) **SDN and NFV** The deployment of SDN and NFV concept in 5G network can save for about 25% 5G's TCO. This 25% savings are covered both from RAN and core virtualization. In the future 5G era, NFV will be a potential-generating tool, and maybe people will heavily rely on NFV when using 5G network.
2) **Automation and intelligence** Automation and intelligence are heavily related to AI. AI in 5G will be

expected to save about 25% 5G's TCO. We already know that AI in 5G has made open RAN able to do many things, including automation and deploying ML. However, it is predicted that AI in the future will still be an uncommon thing because the actual AI applications in real life is still rare.
3) **Network slicing** Network slicing will not be a part of a cost-saving element in 5G, but the one that can reduce cost is network sharing. Network sharing can make savings for about 40% TCO.
4) **Cloudification and open source** The deployment of cloud and open-sourced software in 5G network can save up to 5% 5G's TCO.

The cost efficiency mentioned in this section can also be further improved by implementing open RAN in 5G networks. Before open RAN era, the previous RAN hardware infrastructure was the part that had the highest cost of all parts in RAN infrastructure. Basically, to make good RAN hardware, there must be cabinets, radio antennas, baseband processing tools, power tools, cooling tools, and other tools. These tools had made for about 45% until 50% out of RAN's TCO. In 5G technology, the infrastructure cost has increased by 65% in some deployments. This 65% cost is probably the maximum possibility for any open RAN scenario. This cost can be less than 65% for some open RAN's lower-cost deployment scenarios.

This wide gap depends on some modifications used in the 5G open RAN architecture. One of those modifications is adopting C-RAN's architecture deployment into the open RAN. At least, this C-RAN adoption has saved 25%, compared to the previous generation of C-RAN cost, which is D-RAN. While C-RAN adoption has saved 25%, this number is predicted to be up to 45% in the future.

While the FH has made its way to reduce the RAN's TCO, the backhaul will not be similar to FH. Because 5G connectivity requires higher capacity with lower latency, backhaul has its cost increased by 55% in some higher open RAN deployments. This 55% number is not definitive, because the increased cost can probably be lower than this number in some lower open RAN deployments.

IV. Projects, Activities, and Standards Related to Open RAN

This section describes some important open RAN related projects and activities. The term "related projects" means how open RAN is implemented in several contributions and other projects from organizations and companies who believe in the vision of open RAN. However, even if these projects implement the open RAN framework, they may have different standardization and do not necessarily follow the O-RAN Alliance Standard. We will also explain some mobile communication standards that are tightly related to open RAN.

A. Projects and Activities

The first project that we will explain is the O-RAN Alliance. This community name is shortened to O-RAN. Because the term is tightly related is open RAN, many people assumed that O-RAN, like ORAN, is also a shorter version of open RAN. In addition, those two terms are used by industries interchangeably. There are also many journals that used these two terms at the same time, and with the same meaning. That is not the case. To make it clear, O-RAN is the short form of O-RAN Alliance, a name of a software community [15].

O-RAN Alliance is a community that has tried to standardize and detail the open RAN specifications so they can be implemented in real life [39]. O-RAN Alliance was founded in February 2018 [40]. To form O-RAN Alliance, five big mobile operator companies at that time gathered, which were AT&T, China Mobile, Deutsche Telekom, Orange, and NTT Docomo. The main purpose of making O-RAN Alliance is to promote an open and intelligent RAN. This alliance is a combined version of xRAN Forum and C-RAN Alliance [15], [40]. All these communities were involved in deploying open interfaces, big data intelligent control, and virtualization in RAN. These communities had also recognized the need to make an open network. After these communities were merged into O-RAN Alliance, these communities had become official members of O-RAN Alliance, but they still kept their original purposes to deploy a more reliable and faster network. The first O-RAN Alliance board meeting was held in June 2018, and the first work group (WG) meeting in September 2019 [40]. Until March 2019, the O-RAN Alliance consisted of 19 operator members and 55 contributor members, including major vendors and venture firms.

O-RAN Alliance has its own core principles. Basically, there are two core principles of O-RAN Alliance. These core principles are openness and intelligence [2], [22]. To achieve these two high-level core principles, O-RAN Alliance has given some reference designs about how the architecture of the open RAN should be [40]. These reference designs from O-RAN Alliance are called O-RAN vision, which consists of standardization design, virtualization design, and white box design [22], [41]. To be able to achieve the three visions, O-RAN Alliance divides itself into smaller groups. There are two kinds of groups in O-RAN Alliance. The first groups are the groups divided based on each work description, called WGs. There are ten WGs, and these ten WGs' objectives and chairpersons can be seen in Fig. 7. The second groups are groups divided not based on each work description, but based on their scopes, or their focuses. We call these second groups as focus groups (FGs). There are four FGs [39]:

1) **Standard Development Focus Group (SDFG)** SDFG is responsible to make the standardization of O-RAN Alliance and to make the main interface to other Standard Development Organizations (SDOs) that will be relevant for the alliance's work. SDFG also coordinates incoming and outgoing liaison statements.
2) **Test and Integration Focus Group (TIFG)** TIFG's task is to define O-RAN Alliance's overall approach for doing tests and integration, including coordination and specifications tests for WGs.
3) **Open Source Focus Group (OSFG)** is responsible

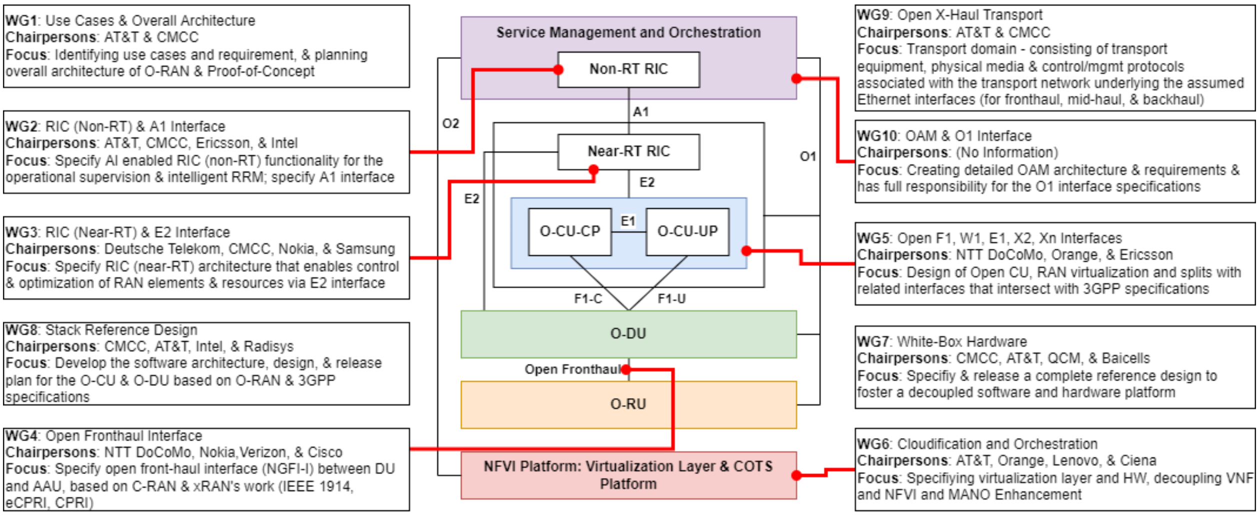

Fig. 7. O-RAN Alliance's WGs and their objectives

for dealing with open source-related issues for O-RAN Alliance. OSFG is the parents of OSC, which will be explained briefly.

4) **Security Focus Group (SFG)** is responsible for collecting security requirements and solutions from WGs.

To maintain the healthy relationships among every WGs and FGs, O-RAN Alliance formed a special committee, whose position is above all of these WGs and FGs. This committee is called O-RAN Technical Steering Committee (TSC). TSC consists of member representatives and the technical leader from every WGs, and this community represents both from members' side and contributors' side in making decisions for every technical innovation made by WGs and FGs for O-RAN Alliance. TSC has tasks to provide technical guidelines to every WGs and FGs; and approve O-RAN's specifications that have been made by WGs and FGs based on members' approvals and publications [39].

Another important committee within O-RAN Alliance is the MVP Committee (MVP-C). MVP-C is relatively new in O-RAN Alliance. MVP-C's main job is to prepare a MVP for O-RAN's WGs and the public. MVP will provide a priority list of work items and is used to coordinate WGs. MVP helps O-RAN Alliance to work more effectively. For the public, MVP gives a clear understanding of O-RAN's roadmap and priorities.

As written before, the third FG called OSFG is the parents of OSC. OSC is founded in April 2019 as a collaboration between O-RAN Alliance and LF [15], [42]. This collaboration is made to support the creation of RAN's specific software. OSC is responsible for dealing with open source-related issues for O-RAN Alliance [39]. The software community is focused on aligning a software reference implementation with the O-RAN Alliance's open RAN architecture and specifications [25]. Because of this focus, OSC's responsibilities are developing the open-source software, coordinating with other open-source communities, promoting the open-source software, and addressing wireless technology support for essential patents [25], [39], [43].

The OSC also has Technical Operating Committee (TOC) that has a similar position and tasks as O-RAN's TSC. OSC usually publishes two software releases every year [25]. This tradition has started since December 2019 by OSC, unveiling its first release called Amber. After Amber, OSC has done the Bronze, Cherry, Dawn, and the E release when this paper was written. OSC's timeline is shown in Fig. 8. Please refer to [44] to learn more about each release's progress as we will not dig into those details. The next release after E will be called F. The F release will be scheduled approximately six months after E, which is probably in June or July 2022 [45].

The workflow of OSC can also be seen in Fig. 8. In the figure, we can see that OSC and O-RAN Alliance work together in a loop. Firstly, ten WGs contribute to making RAN specifications, architectures, and reference designs. These specs and architectures are checked to 3GPP and 5G network standards. Besides checking to 3GPP and 5G standard, these specs and architectures are also given to TIFG. TIFG tests whether OSC's software aligns with the specifications defined by O-RAN Alliance. If there is any OSC code that works differently than the original specification, TIFG will request TOC to resolve these variance problems so that it aligns with the standard specification. After that, TOC prioritizes and negotiates with O-RAN Alliance about solving these variance problems. The result of this negotiation is passed to the Requirements and Software Architecture Committee (RSAC). On the other hand, O-RAN Alliance's WG1 also gives recommendations to RSAC about RAN specs or design features for inclusion in a particular release. Until this process, we can see that RSAC receives two things: negotiations from TOC and specs recommendation from WG1. When receiving the inputs, RSAC selects those recommendations based on

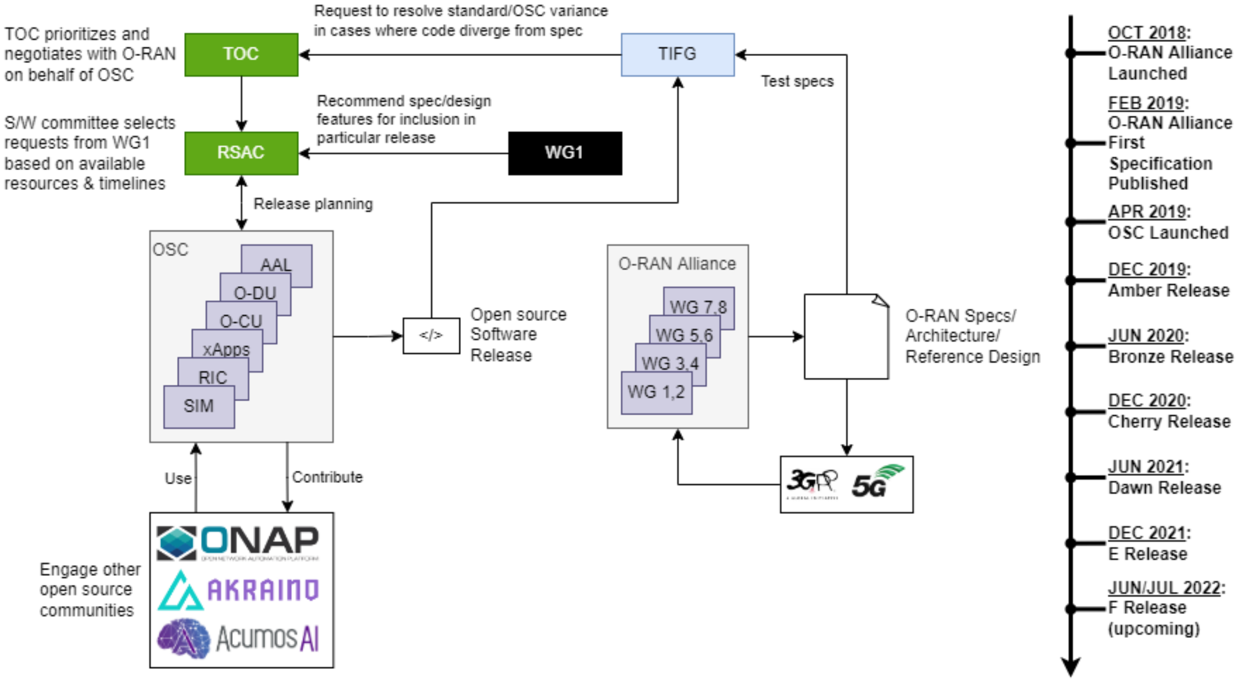

Fig. 8. OSC's workflow and timeline

available resources and software release timeline. Therefore, RSAC and OSC together develop their release planning. This release planning is implemented by OSC to produce the software releases every 6 months. OSC also interacts with other open source communities, as illustrated in the figure, such as Open Network Automation Platform (ONAP), Akraino, and Acumos AI. OSC contributes to these communities and at the same time, OSC also uses the software made by these communities. From the software release, the process starts again.

O-RAN Alliance supplies further support to the open RAN movement by providing OTICs and conducting the O-RAN Global PlugFest. OTIC provides a working laboratory environment for open RAN E2E system testing, certification, and badging [46], [47]. OTIC's environment is designed to be vendor independent, open, collaborative, and secure. Currently, there are 6 OTICs in Berlin, Torino, Madrid, Paris, Taoyuan, and Beijing [46]. Also, OTIC's physical space can be used to host O-RAN related activities, such as PlugFest. PlugFest is a worldwide testing and integration event to demonstrate the O-RAN ecosystem [46]. The scope of PlugFest includes testing using the O-RAN Test Specifications; Validation and Demo of the O-RAN architecture elements, concepts, feature packages, reference implementation, and reference design; and O-RAN certification and badging dry-run. The third O-RAN Global PlugFest has just finished in November 2021 [47].

Besides O-RAN Alliance related projects and activities, there are other projects that stem from the open RAN movement. OpenAirInterface (OAI) is one of them. Started in 2009, OAI is an open software for RAN and CN developed by Eurecom. Since 2014, OAI is managed by OAI Software Alliance. This alliance originated in France and is divided into several project groups. OAI 5G RAN project group aims to develop an open source 3GPP compatible 5G next generation node B (gNB) RAN stack software [48] for its community members. Currently, OAI provides software-based implementations of evolved node Bs (eNBs), UEs and Evolved Packet Core (EPC) that are suitable with Long-Term Evolution (LTE) Release 8.6 [3]. Besides OAI 5G RAN, OAI also has Mosaic5G project group with other projects, such as FlexRIC, FlexCN, and Trirematics.

OAI RAN's source code is written in C programming language, and this source code is distributed under OAI Public License, which is the combination of Apache License v2.0 and fair, reasonable, and non-discriminatory (FRAND) clause [3], [49]. In this OAI framework implementations are compatible with Intel x86 architectures running the Ubuntu Linux OS. These implementations also require some modifications, mostly in kernel and BIOS-level modifications. These requirements should be fulfilled to make the OAI-RAN platform can perform in real-time, including installing a low-latency kernel, disabling power management, central processing unit (CPU) frequency scaling functionalities, and can also be used for field experimentation and evaluations with emulated wireless links [3], [50].

The second related project is srsRAN. srsRAN is a free open-source software radio suite for 4G and 5G started in 2014 [51]. The project was formerly known as srsLTE and was originally developed by a startup called Software Radio System. Originally, srsLTE provides software implementations of LTE. Similar with OAI-RAN, the software implementations done by srsLTE is in eNB, UEs, and EPC that are suitable with LTE Release 10 [3]. The additional features for srsLTE are worked in 3GPP's Release 15. Also similar with OAI-RAN, srsLTE's software codes are written in C and C++ programming languages. srsRAN codes are distributed under GNU Affero General Public License (AGPL) version 3 or commercial license. The software is also compatible with Ubuntu plus Fedora Linux distributions. Different than OAI-RAN, srsLTE does not require kernel or BIOS-level modifications. srsLTE does need to disable the CPU framework scaling to do the real-time performance. srsLTE has now transformed to srsRAN as the project expands focus to 5G New Radio (NR) [51].

The third project is the TIP OpenRAN Project Group. The project group has been started in February 2016 [15]. At that time, vRAN was widely discussed but still, that version of vRAN was considered impractical to implement and deploy for commercial traffic [24]. The project group consisted of MNOs who felt that the telecommunication industry was lacking innovation and the equipments needed for deploying network connectivity were extremely costly, especially in a highly concentrated or closed ecosystem [15]. TIP OpenRAN project group is an initiative to develop RAN solutions based on open interface that can run on general-purpose hardware [24]. This project's scope includes multiple generations of the mobile communication system, from 2G to 5G. However, unlike previous projects, this project is closed source [3]. OpenRAN project group has goals to encourage innovation through building an ecosystem that can enable openness of the network; enable multi-vendor and software-based RAN; and reduce network deployment and maintenance costs. [15], [24]. Every work done by TIP OpenRAN Project Group is reviewed and approved by TIP technical committee. The members of the project group are mostly companies who have become members of TIP. Some of them are Vodafone and Telefonica, along with Intel as TIP co-chair [24].

The fourth open RAN related project is ONAP. ONAP is an open source project started in 2017 that provides a common platform for telecommunication companies, providers, and operators to design, implement, and manage differentiated services [52]. ONAP project is one of LF projects [3]. The project automates 5G by using SDN and NFV technologies. ONAP includes all the Management and Orchestration (MANO) layer functionality that compliant with NFV architecture from European Telecommunication Standards Institute (ETSI). ONAP includes network service design framework and fault, configuration, accounting, performance, and security (FCAPS) functionality.

The fifth open RAN related project is SD-RAN. SD-RAN is one of ONF's projects. ONF is a community of several telecommunication operators that contribute to their own open source codes and frameworks for their networks' deployments [3]. ONF always works closely with O-RAN Alliance, TIP, Broadband Forum, LF, and Open Compute Project in every network deployments [27], [53]. Started in 2020, SD-RAN's purpose is to build open source RAN components. In particular, SD-RAN is an ONF's component project that focuses on building the near Real-Time (RT) RAN Intelligent Controller (RIC) and xApps [3], [5], [54]. The xApps-SDK code made by SD-RAN is shared with OSC. SD-RAN has made its initial release in January 2021, called SD-RAN v1.0. SD-RAN v.10 is a cloud-native exemplar platform for software-defined RAN based on O-RAN Alliance's specs.

In addition to the projects above, there are also testbeds that can instantiate softwarized 5G networks by leveraging some of the open source components above. POWDER-RENEW, COSMOS, and Colosseum are a few examples [3]. The 3 testbeds are part of Platforms for Advanced Wireless Research (PAWR) program [3], [55]. POWDER-RENEW and COSMOS have worked with O-RAN and can support indoor and city-scale outdoor scenarios[3]. On the other hand, Colosseum is advertised as the world's most powerful wireless network emulator. Colosseum can support a large-scale network emulator scenario with with 256 programmable software radios [3], [55].

B. 5G, 3GPP, and Open RAN

In the previous subsection, 3GPP is mentioned several times. 3GPP is a standard development organization for the cellular telecommunications technologies [56]. Originally, the scope of 3GPP's technical specifications and reports are for the 3G mobile system. These standards are structured in a system called "Releases". However, the scope was then extended beyond 3G, including the 5G mobile system that we have been talking about. 3GPP was tasked to do some scheduling and formulation toward the 5G technology standard together with International Telecommunication Union Radiocommunication Sector (ITU-R), a community that is fully responsible for radio communication [57]. To standardize the 5G framework, 3GPP has deployed Release 15 (Rel-15), Release 16 (Rel-16), and Release 17 (Rel-17) [58]. These releases introduce the new 3 use cases domain of 5G, which are eMBB, URLLC, and mMTC. Further explanations about these releases can be seen in [57] and [58].

From these releases, 3GPP introduced a new term called NR. NR is a new radio access technology (RAT) that 3GPP developed for 5G technology. Simply, NR is the term for 5G RAN. 3GPP defines some specifications for the 5G NR technology, such as Standalone (SA) configuration, Non-Standalone (NSA) configuration, CU - DU functional split, and CUPS [57], [59]. We will briefly explain these specifications.

The 5G NR configuration can support either SA or NSA configuration. Like its name, the NR SA means the 5G NR framework works independently without having another generation technology connected. NR NSA is the opposite of the NR SA. The NR NSA still connects to 4G CN or EPC to support the NR. 3GPP also defines the interface differences between NSA and SA configurations. SA and NSA have their CN and RAN involved differently in each configuration:

1) **NSA Configuration** The NR's eNBs connect to the gNB through an X2 interface. Another interface used in NSA is S1 interface, connecting eNBs and gNBs to EPC. NSA uses EPC as its CN, and its users can connect both to eNB or gNB to deliver the network.
2) **SA Configuration** SA enables operation service to be provided solely on the basis of gNB. The gNB connects to other gNBs with Xn interface and connects to the 5G core (5GC) through NG interface. SA uses 5GC as its CN. SA's UEs will only connect to gNB. CU and DU are parts of gNB.

As we have mentioned in the previous section, there are 2 kinds of functional split. The CU - DU functional split is called horizontal functional split. 3GPP defines the interface between the CU and DU as the F1 interface. Another further specification that 3GPP defines is the vertical functional split. Vertical functional split is done to separate the UP and CP radio protocol. UP carries the network user traffic. UP protocol stacks consist of three layers, which are packet data convergence protocol (PDPC), radio link control (RLC), and medium access control (MAC). CP radio protocol contains radio resource control (RRC) layer. RRC is responsible for controlling or configuring the lower layers, broadcasting information from a terminal to a cell, and deploying RRC connection functions. Further details about UP and CP can be seen in [59].

After many years of planning, the 5G technology has applied in real life in 2019 [38]. The 5G launch already existed for about 36 deployments across Asia, Europe, and North America. The telecommunication companies had also worked hard to implement 5G in real life. However, the 5G technology still has so many elements that needed improvements today. One of these improvements that should be done is optimizing 5G network costs. As we have seen from the previous section, the open RAN movement is also started due to this concern.

3GPP	O-RAN
Functions • Management & Orchestration • CU-CP/CU-UP • DU	Additional Functions • SMO • Non-RT RIC • Near-RT RIC
Interfaces • E1 • F1-C/F1-U • NG-C/NG-U • X2-C/X2-U • Xn-C/Xn-U • Uu	Additional Interfaces • A1 • E2 • O1 • O2 • Open Fronthaul • O-Cloud Notification interface

Fig. 9. Comparison between 3GPP and O-RAN network functions and interfaces

It might be confusing to grasp the urgency of the open RAN movement formation since the 5G RAN standards have been developed by 3GPP. Previously, we have explained how 3GPP defined the RAN and its interfaces standard. These interfaces' standards are open and standardized, such as the Uu, S1, and X2 interface [15]. Although 3GPP has provided the standardization for 5G RAN's framework, this standardization is still insufficient to provide a clear and definite standard for open interfaces. For example, 3GPP does not provide the standardization for the FHI. 3GPP also defines the X2 interface as an optional interface so most vendors did not implement this interface. Due to the lack of standardization, the implemented FHI and X2 interfaces are full of proprietary specifications in case the vendor did implement it.

The unclear RAN standard was not enough to provide good service to existing mobile terminals and smartphones, especially if a multivendor environment is introduced. It was necessary to create a new movement in the telecommunication industry by cooperating with other companies in industries and release new RAN functions and interfaces standards that can enhance its value. That is why the open framework network movement was started. To fulfill this necessity, NTT DoCoMo and other operators gathered to create the O-RAN Alliance that we have explained in the previous section. Fig. 9 summarize the new additional functions and interfaces that O-RAN Alliance provides to supplement the 3GPP specifications.

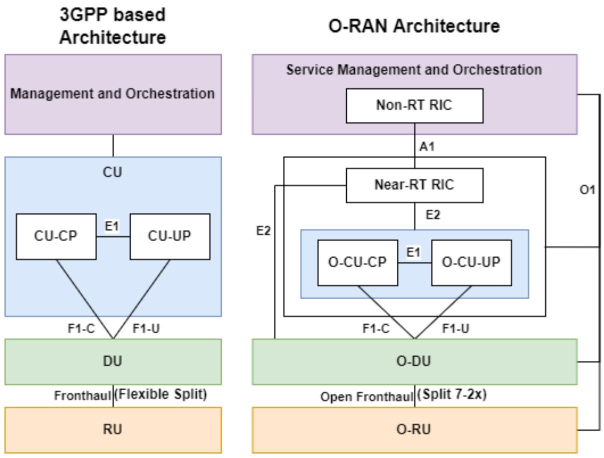

Fig. 10. Comparison between 3GPP and O-RAN architecture

We will explain the functions and interfaces in Fig. 9 in the next subsection. As an introduction to the following section and to better understand the difference between 3GPP and O-RAN specifications, Fig. 10 shows how the functions and interfaces are positioned in the RAN architecture. We can see that the specifications that O-RAN Alliance provides are more detailed than the more generalized 3GPP specifications. This is important, particularly in the multivendor environment. For example, the X2 interface becomes essential for multi-vendor networks to function seamlessly [15]. In addition, we know from the previous section that the 5G RAN needs to support SA and NSA deployment. The current 5G deployments are still NSA and it will be challenging to move to the future 5G SA RAN deployment if the X2 interface is not open [15]. Open RAN's multi-vendor RAN system dream also helps reduce costs for mobile network operators by introducing more software vendors and diversifying the supply chains. In spite of that, open RAN standards still need to comply with 3GPP specifications because 3GPP's specification is the universal standard for the 5G RAN technology.

However, open RAN's principles are not only applicable to the 5G RAN technology. Open RAN movement's purpose is to define and build RAN solutions based on general-purpose, vendor-neutral hardware and software-defined technology in 2G, 3G, 4G, and 5G [15]. Even though open RAN can be used for other cellular generations, the current O-RAN Alliance's standardization still focuses on 5G because it is the latest trend that the telecommunication industry and companies follow. We will introduce how open RAN is deployed in 5G RAN in the next section about O-RAN Alliance architecture. In the following section, we will see how the interfaces and functional split that is specified by 3GPP for the 5G NR are implemented in the O-RAN Alliance architecture.

V. O-RAN Alliance Architecture

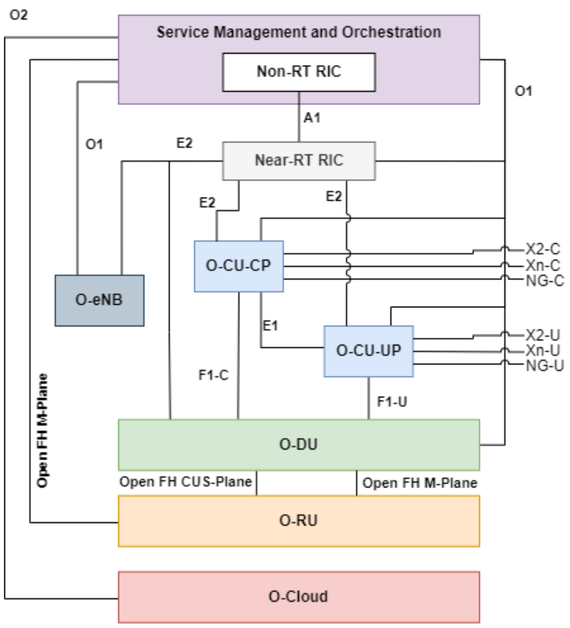

Fig. 11. Logical architecture of O-RAN

In this section, we will describe how open RAN system works by referencing O-RAN Alliance's standardizations. As written before in the previous section, we have known that O-RAN Alliance has given many standardizations for the open RAN framework. All of those standards and reference designs from O-RAN Alliance will be explained deeper in this section. The standardizations of O-RAN Alliance are used in this section because O-RAN Alliance has provided detailed reference design specifications, completed with clear documentation for each reference design. Since the O-RAN Alliance reference design is an open design, this section will only summarize the reference design that is already available publicly. For a further and deeper understanding of the reference design, please refer to the original publications made by O-RAN Alliance.

Besides the whole progress that O-RAN Alliance has made, O-RAN Alliance is also supported by OSC so that we can directly see how the standards are implemented into open-source software. OSC has also made rapid progress in implementing the standardized O-RAN system compared to the other open source RAN projects. However, this rapid progress still has not reached full completion. Because of this, in this section, we will also mention what features or components are still missing in OSC's implementation from the O-RAN Alliance's reference architecture. In the closing part, we will also mention several use cases of O-RAN Alliance's standardized architecture concept.

Generally, the O-RAN Alliance's architecture for open RAN consists of 7 main parts: Service Management and Orchestration (SMO), Near-RT RIC, O-RAN CU (O-CU), O-RAN DU (O-DU), O-RAN RU (O-RU), O-RAN eNB (O-eNB), and O-RAN Cloud (O-Cloud). To picture how these building blocks are connected to each other in the O-RAN's architecture, we should start by looking at the O-RAN reference architecture that WG1 has provided. The reference can be seen in Fig. 11.

The reference architecture shown in Fig. 11 is designed to enable the next generation of RAN infrastructures [6]. This architecture is the basic, or the foundation of making the open, virtualized, and AI-embedded RAN. The idea of making such RAN has long been thought of by many operators. The O-RAN's architecture will be a standardized reference system, plus a complimentary reference system for other architectures made by other parties, such as 3GPP and other RAN-related organizations. From Fig. 11, there are several main parts of the architecture that we have mentioned before. We will discuss each of them in brief.

A. SMO

SMO is responsible for RAN domain management in the O-RAN architecture [60]. In the nutshell, SMO functions as the Management and Orchestration framework in O-RAN [4]. SMO is needed to provide RAN support in FCAPS; Non-RT RIC; and O-Cloud management, orchestration, and workflow management. The SMO framework has several components and interfaces such as the Non-RT RIC, A1, O1, O2, and open FH M-Plane. The relationships between these functions can be seen in Fig. 12.

The O1 interface is used so that SMO can support other O-RAN network functions' FCAPS. On the other hand, O2 interface is used by SMO for O-Cloud management, orchestration, and workflow management. Open FH M-Plane is used for FCAPS functionality from SMO specifically to O-RU. The Non-RT RIC function is to support intelligent RAN optimization and perform intelligent RRM in non-RT intervals. The interface between Non-RT RIC and Near-RT RIC is called

A1. Basically, the Non-RT RIC will get the information from the elements of the RAN, create or choose an ML training model for RAN optimization, and pass it to the Near-RT RIC through A1 interface. The A1 interface itself can support 3 types of services, which are policy management, enrichment information, and ML model management.

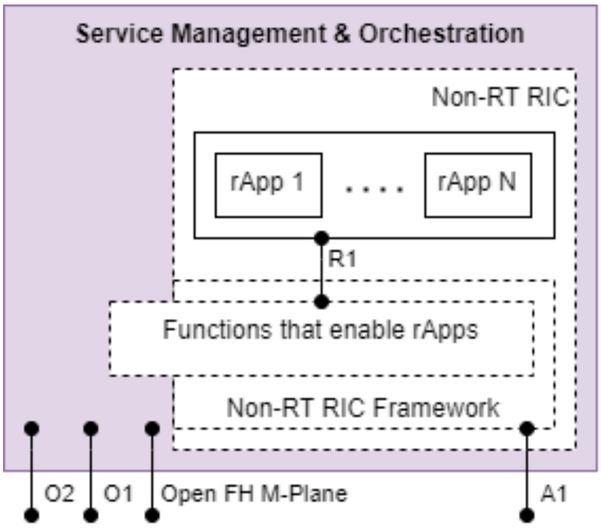

Fig. 12. O-RAN's SMO framework

As can be seen in Fig. 12, the Non-RT RIC is further divided into 2 sub-functions. The first function is the Non-RT RIC Applications (rApps), which are applications to perform RAN optimization [5]. Outside rApps, there are internal SMO framework functionalities that terminate the A1 interface and expose services to the rApps through R1 interface. Non-RT RIC framework also hosts data management and exposure; and AI/ML workflow. There is also an internal messaging infrastructure for communication between Non-RT RIC with other SMO functions.

O-RAN's operation and maintenance (OAM) is directly related to O-RAN's O1 and O2 interface of SMO. O-RAN's OAM architecture points out management services, managed functions, and managed elements supported in O-RAN. OAM logical architecture is as shown in Fig. 13. From the figure, we can conclude that the O2 interface will manage the O-Cloud while the O1 interface will manage all the other O-RAN building blocks from Near-RT RIC to o-eNB. Currently, OAM itself has 2 use cases: O-RAN service provisioning and O-RAN measurement data collection. Service provisioning focuses on supporting connectivity in/between physical network function (PNF) and VNF. Operators can also operate and maintain the network through service provisioning. O-RAN measurement data collection focuses on collecting measurement data from the network so that the Non-RT RIC can do AI/ML training/inference/analyzing for optimization.

Moreover, there are actually several OAM models and deployment options, which are flat, hierarchical, and hybrid management models [61]. In the flat management model, all the O-RAN architecture entities are managed by the SMO through O1 interface. The O-eNB, O-CU, O-DU, and O-RU are called managed functions (MFs) or managed elements (MEs). In the hierarchical management model, some MFs are allowed to manage over lower-level MEs. An example would be the O-DU managing the O-RU. Lastly, the O-RU management responsibility could be shared between O-DU and SMO. This model is called the hybrid management model.

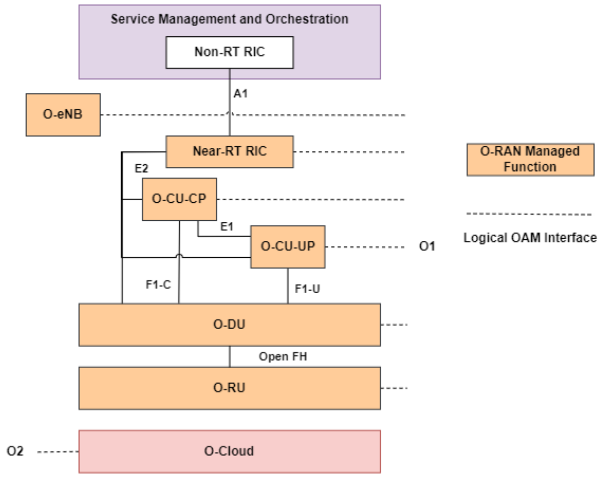

Fig. 13. O-RAN's OAM architecture

Looking back to section III subsection A, we can conclude that the WGs that are in charge of the components inside SMO are WG2 and WG10. Inside OSC, there are also several companies that are in charge of the development of the SMO code [44]. SMO in general is developed by VMWare. Ericsson focuses on the development of Non-RT RIC. In addition, OAM is developed by Highstreet Technologies. As a whole, the O-RAN SMO standard is currently still under development and the specs are not complete yet. Specifically for the Non-RT RIC, the ML specs are not finished yet. Nevertheless, OSC's SMO in Dawn release can already support the O1 interface. OSC's development of O1 interface has also reached a point where the fault supervision management service is finished. A1 Policy Management and Enrichment Information are also ready to use in the Dawn version. Outside of O-RAN Alliance and OSC, we also have already mentioned several SMO-related projects in section III subsection A.

B. Near-RT RIC

The Near-RT RIC is a logical function that allows near-real-time control and optimization over the RAN elements and resources through data collection and actions over its interface [60], [62]. In short, Near-RT RIC will receive the policy from Non-RT RIC, follow the policy, and adjust the RAN parameters based on the given policy. To do this, Near-RT RIC has A1, O1, and E2 interfaces. Data collection and actions are done by Near-RT RIC through the E2 interface connected to E2 nodes. The controlling actions over the E2 nodes are directed by the policies and enrichment data that the Near-RT RIC receives from the Non-RT RIC via A1 interface, as we have explained before. In addition, O1 interface enables FCAPS from the SMO to Near-RT RIC. Near-RT RIC's architecture can be seen in Fig. 14.

An E2 node is a general term for a logical node terminating the E2 interface [60]. Specifically, the E2 node could be an O-CU, O-DU, or O-eNB. From the Near-RT RIC perspective, the E2 interface is a one-to-many connection [4]. Only CP protocols are done in the E2 interface. The E2 functions include 2 category groups: Near-RT RIC Services and Near-RT RIC support functions. Near-RT RIC services contain report, insert, control, and policy. On the other hand, Near-RT RIC support functions cover E2 management and Near-RT RIC service update.

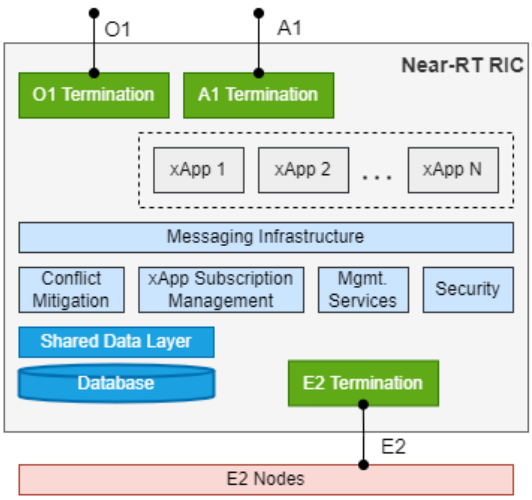

Fig. 14. Near-RT RIC internal architecture

In Fig. 14, we can see that the Near-RT RIC has several components. These components can be divided into 2 major groups: the RIC platform and the xApps. The RIC platform includes all the supporting termination and management components. The database stores and provides data from or to xApp applications and from E2 nodes. The xApps subscription management combines all subscription and data distribution operations. Conflicting interactions from different xApps are resolved by the conflict mitigation function. Security function protects the Near-RT RIC from hazards coming through the third-party xApps. Management services cover FCAPS and life cycle management for xApps. Messaging infrastructure provides a common message distribution system for different components of the Near-RT RIC.

Conversely, the xApps component is the main component of the Near-RT RIC. xApps is a set of applications where each xAPP provides specific RAN functions for the corresponding E2 node. Examples of these functions are RAN data analysis and RAN control [5]. xApps are considered as third-party applications and can be implemented by multiple microservices [4]. The xApps component is used to enhance the RRM capabilities of the O-RAN architecture. xApps can also provide basic info such as configuration data, metrics, and controls.

From section III subsection A, we can conclude WG3 is in charge of developing the Near-RT RIC and E2 interface specifications. In OSC, there are 2 main companies that are in charge of the development of the Near-RT RIC code [44]. RIC platform is developed by Nokia. AT&T focuses on the development of xApps. However, China Mobile, Viavi, Samsung, HCL, and ONF also contribute to the xApps development. In Cherry release, admission control, traffic steering (TS), key performance indicator (KPI) monitor, QoE prediction (QP), anomaly detection (AD), and measurement campaign xApps are already available. For the Dawn release, the Load Prediction, and Security xApp are going to be available. Although there are already numerous xApps that are available, the E2 interface is still incomplete. To this date, the E2 interface still cannot do the control service or modify any RAN parameters in the E2 node. The E2 interface can only do the report and indication services.

C. O-CU

The is a logical node hosting the functions of RRC, SDAP, and PDCP protocols of the BS [60], [63]. In O-RAN architecture, O-CU can be further divided into O-CU-CP and O-CU-UP. O-CU-CP consists of the CP part of PDCP and RRC. Conversely, O-CU-UP covers the UP part of PDCP and SDAP. O-CU has a lot of interfaces which include the E1, E2, F1, NG, O1, X2, and Xn interfaces. The F1, NG, X2, and Xn interfaces can each be divided into the control and user interface. O-CU's architecture and its interfaces can be seen in Fig. 15.

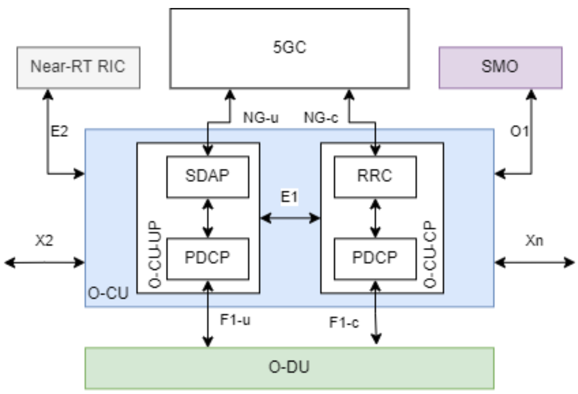

Fig. 15. O-CU architecture

E1, F1, NG, X2, and Xn are all interfaces defined by 3GPP. O-RAN Alliance reuses the E1 specification defined by 3GPP and adopts the E1 interface to bridge O-CU-CP and O-CU-UP. The F1 interfaces are used to connect O-CU to the O-DU. Inversely, the NG interfaces link O-CU to the 5GC. X2 interfaces connect the O-CU to other eNB or en-gNB in EN-DC configuration. Lastly, Xn interfaces link the O-CU to other gNB or ng-eNB. Both X2 and Xn interfaces are adopted from 3GPP with interoperability profile specifications added.

In section III subsection A, we can see that WG5 and WG8 are in charge of developing the O-CU and its interface specifications. In OSC, China Mobile is the main company that is in charge of developing the O-CU code [44]. In the plan for Dawn release, the O-CU-CP and O-CU-UP are already integrated. The O-CU code also has supported the F1, E1, and Ng messages. The E2 interface has been supported for handover and E2AP. Last, O1 interface is supported. Nonetheless, the latest status indicates that OSC's O-CU code development will be postponed. In exchange, O-RAN Alliance is using Radisys's commercial O-CU for testing purposes.

D. O-DU

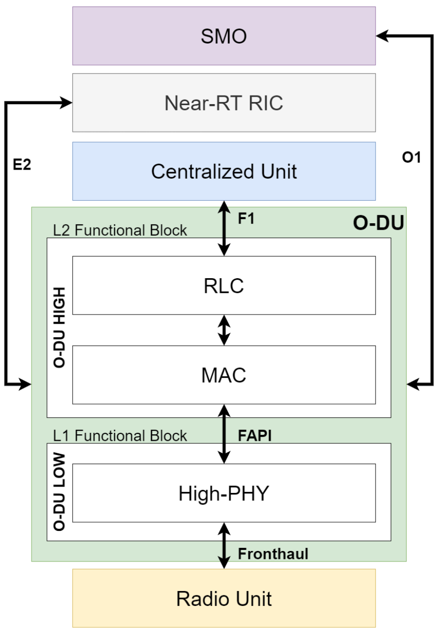

Fig. 16. O-DU architecture

RLC, MAC, and high physical layers are hosted in the O-DU logical node [60], [63]. Data segmentation/integration, scheduling, multiplexing/demultiplexing, and other baseband processing are in O-DU. In general, O-DU is divided horizontally into two parts: O-DU high and O-DU low. O-DU high handles the layer 2 (L2) functional blocks, which are the RLC and MAC layer, while O-DU low covers the layer 1 (L1) functional block or the high physical. The functional application platform interface (FAPI) standard interface specified by SCF is used so that these 2 parts can communicate with each other. Other interfaces related to the O-DU are the E2 interface, F1 interface, O1 interface, and open FHI (O-FHI). O-DU's architecture and interfaces to other components can be seen in Fig. 16.

In OSC's Cherry release, the 5G O-DU High is divided into 6 different threads as shown in Fig. 17 with different colors. The first thread will host the overall O-DU High thread, including the O-DU utility and common functions. The second thread is known as the DU_APP. DU_APP will cover the config handler, DU manager, UE manager, evolved-GPRS Tunneling Protocol (EGTP) manager, and the ASN.1 codecs. The F1 interface is strongly related to the DU_APP. The third thread will cover the 5G NR RLC UL. The fourth thread will cover the 5G NR RLC DL and most of the 5G NR MAC functions. This includes the 5G NR scheduler (SCH) functions. However, the lower MAC handler will be inside a different thread, which is the fifth thread. Another important note is we can see that there is a separate block for the MAC_SCH. This is done to indicate that the MAC_SCH will be detached from the 5G NR SCH. The last thread will handle the Stream Control Transmission Protocol (SCTP) manager. We will not explain the details of these threads, so please refer to [63] for further explanations.

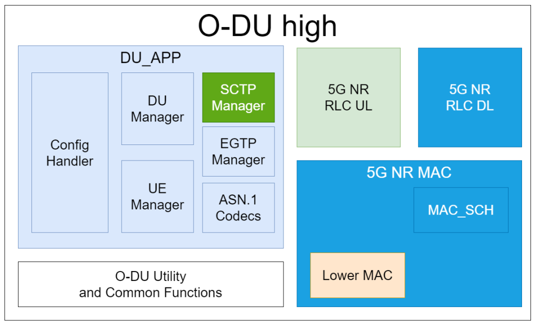

Fig. 17. O-DU high thread architecture in Cherry release

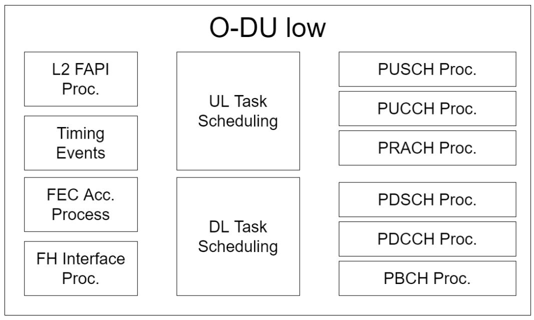

Fig. 18. O-DU low processing blocks

Since O-DU Low handles the high physical, it is constructed mostly by L1 processing blocks as shown in Fig. 18. For the shared and control channel processing, we can see that the downlink (DL) and uplink (UL) are each handled by a separate block, which totals to 4: Physical DL Control Channel (PDCCH), Physical DL Shared Channel (PDSCH), Physical UL Control Channel (PUCCH), and Physical UL Shared Channel (PUSCH). Both Physical Broadcast Channel

(PBCH) and Physical Random Access Channel (PRACH) have independent blocks. The task scheduling is divided into 2 blocks: one for UL and one for DL. Task scheduling will manage all queued tasks and begin the processing operations correspondingly. Then, we have the L2 FAPI processing block to handle the FAPI protocol L2 interface request/response. On the O-RU side, we have the FHI procedure block. We also have the forward error correction (FEC) acceleration processing to handle FEC operations such as passing FEC requests to hardware and invoking callback function on acceleration processing completion. Last, we have the timing events processing block to handle timing-related operations. Please refer to [63] to understand further details on each of these processing blocks.

The O-RAN Alliance architecture adopts the Functional Split 7.2x for the lower layer split between the O-DU and the O-RU [60], [64]. This split is done between the resource element mapping and FFT/iFFT stacks [65]. Fig. 19 shows the location of split 7.2x by using PDSCH as an example. There are 2 variations of this Split 7.2x which differ whether "7.2a" precoding is done in O-DU or "7.2b" precoding is done in O-RU. This split is chosen by O-RAN Alliance while considering the trade-offs in RU complexity and interface throughput. The details of the trade-offs are studied in [10], [31], [65], and [66].

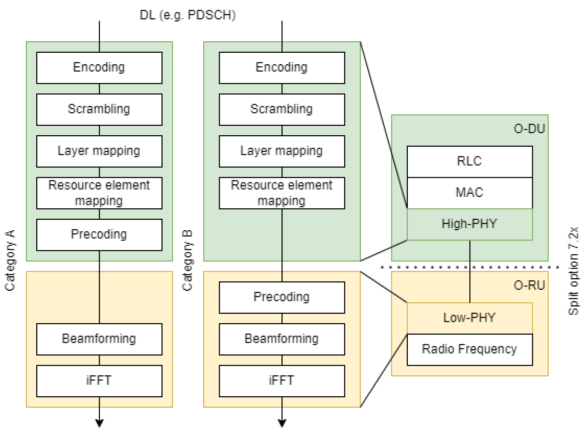

Fig. 19. PDSCH in lower layer split 7.2x

Connecting this O-DU and O-RU functions split is the O-FHI. There are 4 planes inside the O-FHI, which are the Control, User, Synchronization, and the Management Plane. As we have known, the CP is a protocol for passing control signals, including scheduling and beamforming commands, configuration parameters, configuration requests, status, and responses. UP transfers the user data, such as the DL, UL, and PRACH In-phase/Quadrature (IQ) data. Synchronization Plane (S-Plane) is used for achieving timing synchronization between multiple units of equipment. For S-Plane, there are several clock models and synchronization topologies that are extensively explained in [64]. Finally, Management Plane (M-Plane) handles maintenance and monitoring signals. The data flows in the O-FHI are shown in Fig. 20.

Section III subsection A has clearly explained that there are separate WGs working for the development of O-DU and O-FHI. WG8 is in charge of developing the O-DU specifications. O-FHI specifications are developed by WG4. In OSC, each O-DU high and O-DU Low has a different company that is in charge of developing their respective code [44]. O-DU high is developed by Radisys while O-DU low is developed by Intel. In Dawn release, O-DU high will already support time-division duplexing (TDD) on numerology 1. O-DU high will also support closed-loop automation. O-DU low and O-DU high pairwise testing in timer mode has been accomplished in Cherry. Radio mode testing should be achieved by Dawn release. Last, O-DU high will also support O1 interface for health check. Nevertheless, the latest status indicates that the MAC SCH separation will be postponed by Radisys to focus on finishing multi UE and other functions in O-DU high.

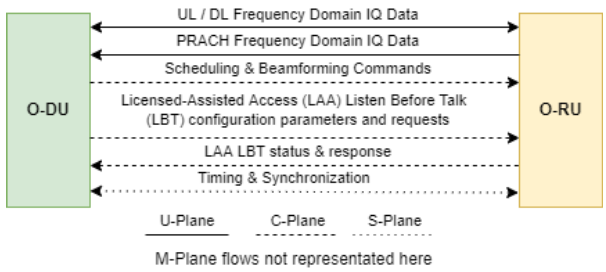

Fig. 20. Fronthaul data flows

For the O-FHI, Intel has finished implementing the C, U, and S plane into O-FHI library in OSC repository. O-RAN adopters can directly use this open-source library in developing their O-RAN compliant DU and RU. However, the M-plane has not been developed yet. In fact, the M-Plane specifications are still under development.

E. O-RU

The O-RU hosts low physical layer and radio frequency processing functions [60], [64]. Each O-RU can have 1 or more panels. Each panel will consist of 1 or more TX-antenna-arrays/RX-antenna-arrays. A TX-antenna-arrays/RX-antenna-array is defined as a logical construct for data routing and is related to the physical antenna described in the O-RU construction. An array element will include 1 or more radiators which specification is written in IEEE Std 145-1993. As we have explained, the O-FHI and the O1 interface connect O-RU to O-DU and SMO respectively.

O-RU supports the beamforming functionality [64]. Beamforming allows a wireless signal to be directed specifically to a receiving device using multiple antennas. Requiring multiple antennas means that this beamforming process can be done in MIMO. We will discuss the benefits of beamforming in section V. O-RAN Alliance architecture specifies that the beamforming commands are implemented between O-DU and O-RU

[34], [67]. O-RAN has included the digital, the analog, and the hybrid beamforming technologies in its specifications. O-RAN Alliance has also proposed four methods for beamforming, which are predefined-beam, weight-based dynamic, attribute-based dynamic, and channel information-based beamforming method [64], [67].

In predefined-beam beamforming, the O-RU has predefined a table of beam indexes containing beam weights. The O-DU will send the information on which index should the O-RU use. Unlike the former, weight-based dynamic beamforming allows the O-DU to generate the beam weights in real-time for the O-RU to use. Nonetheless, this method requires the O-DU to know the specific antenna characteristics of the O-RU. Sparing O-DU to know the details, attribute-based dynamic beamforming lets the O-DU instruct the O-RU to use a specific beam index associated with definite beam attributes. O-RU is in charge of generating the beamforming weights based on these attributes. Similar to the former, channel information-based beamforming also grants the O-RU to generate the beamforming weights. These weights will be created based on the channel information provided by the O-DU. For the mathematical description of these beamforming methods, please look into [64].

Since the RU is majorly hardware-based, O-RAN Alliance only focuses to develop the software standards to control this component. Vendors need to follow the specifications provided by O-RAN Alliance in making their RUs so that they can be considered O-RU. For the Dawn release, OSC promises that the failure detection and recovery use cases simulator will be available for the O-RU.

F. O-eNB

The O-eNB is defined as an eNB or ng-eNB that supports E2 interface [60]. In the nutshell, O-eNB provides the O-DU and O-RU functions in an integrated node while still keeping the O-FHI between them. The O-eNB will have the E2, NG, O1, S1, X2, and Xn interfaces.

G. O-Cloud

The O-Cloud is a cloud computing platform that will host O-RAN functions [60]. O-Cloud can consist of individual or collection of physical infrastructure nodes of the O-RAN architecture such as Near-RT RIC, O-CU, or O-DU. In order to work properly, O-Cloud should also host the supporting software components and the appropriate management and orchestration functions. Inside the O-Cloud, computing resources, such as general-purpose CPUs and shared task accelerators, are pooled and brokered by an abstraction layer [4].

The O2 interface will link O-Cloud to the SMO. The O-Cloud organizes the services relating to O2 interface into 2 groups: Infrastructure management services and Deployment management services. Infrastructure management services are responsible for the management and deployment of the cloud infrastructure. On the other hand, the deployment management services are responsible for life cycle management of the virtualized/containerized deployments. In OSC, the development of O-Cloud is still in initial process. Currently, OSC does not have a repository yet for the O-Cloud.

H. Use Cases

In this subsection, we will introduce the O-RAN architecture's use cases. The O-RAN use cases are generally divided into phase I and phase II. These use cases have been explained deeply in one of O-RAN Alliance's white papers [68]. Several new use cases and updated deployment scenarios of O-RAN architecture are also mentioned in [25]. This subsection will slightly explain these use cases, including those new ones.

In phase I, there are two key use cases, which are white box hardware design and AI-enabled RAN [68]. For the former, its use case is low-cost RAN white box hardware design. The white box has been initially developed by O-RAN Alliance's WG7 as a method to reduce the cost of 5G deployment, and generally, the whole industry chain [22], [68]. Releasing this white box can also reduce the difficulty of deploying the 5G tech, thus making it more attractive for small and medium enterprises in the telecommunication industry to modify the design [68]. The white-box hardware design is mostly focused in O-DU and O-RU, plus an additional FH gateway. The white box hardware design requires support of O-RAN's O-FHI, O-DU open-source software, and facilitates NFV in the cloud.

For AI-enabled RAN, there are many use cases described in phase I [68]. These are TS, QoE optimization, QoS-based resource optimization, and massive MIMO optimization. TS means a process of mobile load balancing [68]. It is a widely used network solution to achieve optimal traffic distribution based on desired objectives. TS is becoming more and more important nowadays, as the mobile network has grown rapidly in these past years. If we're still relying on the traditional optimization system, we will get a lot of disappointment, because the available connection is inefficient, has low quality, low QoE, and slow feedback response. Because of that, the O-RAN architecture has developed the TS leveraging O1, A1, and E2 interfaces, thus making the network more flexible and agile; and reducing network intervention.

Another phase I use cases for AI-enabled RAN is QoE optimization. As we know, 5G network can provide several applications that have high bandwidth consumption and extreme sensitivity to latency, such as applications in Cloud Virtual Reality (VR) [68]. The current QoE framework can not support such applications, because of dynamic traffic volume and fast fluctuating radio transmission capabilities. To reduce this problem, O-RAN Alliance deploy a new QoE framework, a vertical application made for specific QoE prediction and QoE-based proactive closed-loop network optimization. This closed-loop optimization is done in real-time, involving the software-defined RIC. Before QoE is reduced, the radio resources can be allocated more to users and apps who needed the most urgent radio resources. Because of its capability for adapting to users' needs, the QoE is optimized and the radio resources are utilized in a more efficient way.

QoS-based resource optimization is the next phase I AI-enabled RAN use case. The network needs careful planning and configuration to support the diverse 5G services and applications. However, the traffic demand and radio environment are dynamic. Default configuration and planning might not be sufficient to provide highly demanding requirements in extreme situations. QoS-based resource optimization could be used in these situations to optimize RAN resource allocation between users when all user's requirements could not be fulfilled. Analytics function in Non-RT RIC might conclude that it is better to prioritize a group of users through close examination of congestion situations.

The last phase I use case for AI-enabled RAN is massive MIMO optimization. The MIMO framework has been considered one of the key technologies for 5G [68]. Aside from being a key technology, the massive MIMO has also become a reason why the 5G technology requires energy more than the 4G [38]. To overcome this problem, O-RAN architecture has tried to optimize the massive MIMO framework. The objective of this optimization is to improve cell-centric network QoS and user-centric network QoE in a multi-cell deployment area [68]. If needed, this optimization can provide a multi-vendor massive MIMO deployment area with multiple transmission points. However, this area can only be provided by MIMO when using specific operators. When applying this kind of optimization, O-RAN architecture has several advantages, including the possibility to apply and combine non-RT and near-RT analytics; apply ML; and make decisions for various tasks related to Non-RT and Near-RT RIC. Because of this optimization, the Non-RT and Near-RT RIC can oversee the current traffic, coverage, and interference situation in a whole cluster of cells. When doing this massive MIMO optimization, the O1, A1, and E2 interfaces will give necessary data, policy, and configuration exchanges between two components in the O-RAN architecture.

In phase II, [68] mentioned several AI-enabled RAN use cases. The first use case for the interfaces is RAN slice service level agreement (SLA) assurance [68]. The network-level slicing occurred in 5G as a process that can provide E2E connectivity and data processing tailored to specific business requirements. These requirements include several network capabilities, such as support for high data rates and low latency. These capabilities are specified based on SLA, which was agreed upon between the MNO and the customer. Mechanisms to ensure slice SLAs and prevent any violations are becoming more popular. By combining the AI and Ml models into SLA assurance mechanisms through Near-RT and Non-RT RIC, O-RAN Alliance can raise the possibility of SLA slice assurance being fulfilled. This importance of SLA assurance mechanism can lead to further research. Network slicing process still has so many opportunities to be explored by MNOs, thus making the network more efficient and changing the way how operators do their telecommunication business.

Another phase II AI-enabled RAN use case is context-based dynamic handover management for vehicle-to-everything (V2X). Nowadays, the application of V2X concept in real life has increased rapidly. Because of this rapid increase, the V2X applications can be found easily, such as in self-driven vehicles, seat entertainment, and traffic assistance. These applications of V2X can increase traffic safety, reduce emissions, and save more time when riding [68]. The V2X works based on real-time information about the driver, the road, and the traffic conditions. Regarding these V2X applications, the O-RAN architecture can do several things, such as collecting and doing maintenance of the past traffic and radio conditions, deploying and evaluating AI and ML-based applications that can detect or predict abnormal users' behavior, monitoring the traffic and the radio real-time condition, and deploying real-time xApps that predict and prevent anomalous situations in UEs.

The AI-enabled RAN can also be used in flight path-based dynamic UAV resource allocation use case. Like V2X, the application of UAV has been applied in many industries, such as agricultural plant protection, power inspection, police enforcement, police enforcement, geological exploration, and environmental monitoring [68]. The 5G technology can provide a high-speed network for several applications that need low-altitude UAVs. In O-RAN architecture, the Non-RT RIC can provide necessary information about the aerial vehicle and everything around that vehicle, including flight path information, climate information, flight limitation area, and space load information. To provide such information, the RIC can deploy Unmanned Traffic Management (UTM) for constructing and training AI and ML models that should be applied in RAN. This information can also be provided when Near-RT RIC performs radio resource allocation for on-demand coverage for a UAV, by considering the radio channel condition, flight path information, and other information.

Another phase II AI-enabled RAN use case related to UAV is for UAV radio resource allocation. In UAV control scenario, the O-RAN architecture meets the need for wireless resource adjustment [68]. Rotor UAV usually flies at low altitude and low speed, while carrying mounted cameras and sensors. These UAVs have asymmetry requirements between UL and DL services. For UL, the 5G network has to support service to receive the 4K high-definition video UP data from the UAV. On the other hand, there is only a small amount of control data interaction requirement. This introduces a new requirement for the gNB. In addition, the existing network management platform could not optimize parameters for individual users. In O-RAN architecture, Near-RT RIC function module provides RRM functions and radio resource requirements for different terminals to O-CU and O-DU

Besides AI-enabled RAN use cases published in phase II, there is also virtual RAN use case. The use case for virtual RAN network is RAN sharing. RAN sharing is an efficient and sustainable way to reduce the network deployment costs while increasing the capacity and the coverage of the network [68]. O-RAN architecture can accelerate the development of RAN sharing solutions because of its open and multivendor architecture. The O-RAN architecture also enables each operator to separately control their own VNFs in shared hardware.

Remote O1, O2, and E2 interfaces can be introduced so that each operator could use their own SMO and Near RT RIC to independently monitor and remote control their O-DU in shared network resources. O-RAN architecture introduces freedom and ease in the RAN sharing process

In the previous paragraphs, we have explained several AI-enabled RAN use cases. O-RAN prioritized some of these use cases in [25]. The prioritized AI-enabled RAN use cases are TS; QoS and QoE optimization; RAN slicing and SLA assurance; and massive MIMO optimization. There are also new additional AI-enabled RAN use cases introduced, such as multi-vendor slices, O-RAN dynamic spectrum sharing, RAN slice resource allocation optimization, local indoor positioning in RAN, massive MIMO single user/multi user-MIMO grouping optimization, O-RAN signaling storm protection, congestion prediction and management, and industrial IoT optimization. Another new use case introduced is the O-DU pooling use case.

VI. Challenges and Future Research Directions

This section summarizes the challenges that operators may face when deploying and developing the O-RAN architecture and the future research directions in response to the challenges. We will divide this section into 2 parts. First, we mention issues regarding the O-RAN architecture building blocks, covering the design and implementation challenges and future research directions for each of them. After that, we will mention issues about the open RAN field as a whole.

A. O-RAN Alliance Architecture Issues

In this subsection, we will mention some issues regarding the O-RAN Alliance architecture and also its implementation in OSC. A common problem that occurs upon the entire O-RAN building blocks is that the standard specifications are not finished being written yet by the O-RAN Alliance. We have mentioned several nodes that still have not complete specifications in section IV, which are SMO's Non-RT RIC ML, O-FHI M-Plane, O1 interface, and O2 interface. OSC's code development for the currently available specifications is also not finished yet, as mentioned in section IV. Some of them have not or have just started like the O-Cloud. Some of them are postponed so that OSC can focus on developing other nodes, such as the O-CU and the O-DU MAC SCH. Others are waiting for the official specification release, such as the M-Plane.

SMO has an issue regarding copyright with 3GPP. Since some of the O-RAN specifications also reference 3GPP specifications, the occurrence of this issue is inevitable. Currently, this copyright issue focuses on the OAM part. However, O-RAN Alliance WG1 also plans to resolve this issue on the overall O-RAN architecture level. Besides these implementation problems, SMO also faces some design challenges and future research directions that we will explain in the following paragraphs.

Network densification causes management and optimization of neighboring cell relationships to be more complex. O-RAN-based proactive automatic neighbor relations (ANR) optimization is a well-known application of SON that is used to manage neighbor cell relationships, or called neighbor cell relation table (NCRT) [69]. An ANR technique is proposed that identifies the trends before handover failure, marks the cells for handover prohibition, and identifies time-based trends to remove and add back the cells in the NCRT table by [69]. The ML-embedded ANR model is made as an rApp. The rApp has tasks to give several data updates to its neighbor cell. This data will also be used by rApp to improve default cell removal and prohibit any possible handover policies by generating a new policy. The new policy will be sent by rApp to xApp in Near-RT RIC. After that, the xApp executes the policy on open RAN network nodes. This design will minimize the handover failure, reduce the operational cost, thus increase the performance and the QoS of the network.

One of the problems in network optimization/automation is that even though the objective is clear, there is currently no practical way to select which intelligence model should be deployed, where to deploy it, and which RAN parameters to be controlled. An rApp called OrchestRAN is proposed [70]. The OrchestRAN framework offers to solve this. MNOs can directly specify high-level control/inference objectives and OrchestRAN will automatically compute the optimal set of algorithms and where to execute them. OrchestRAN prototype is tested on Colosseum. Experiment results show that OrchestRAN is able to instantiate data-driven services on demand at different network nodes and time scales. OrchestRAN successfully achieved minimal control overhead and latency.

Another method proposed by researchers for SMO is to improve the RRM. The proposed intelligent RRM scheme is called cell splitting. The application of cell splitting is done by [71], where the cell splitting is involved in the open RAN architecture to prevent cell congestion. Long short-term memory (LSTM) recurrent neural network (RNN) is utilized to learn, thus making several predictions about the traffic pattern that would possibly arise from a real-world cellular network in a densely populated area. From these traffic patterns, LSTM will detect any cell that will potentially be congested. LSTM model is trained at Non-RT RIC, by using long-term data gathered from the RAN. This cell splitting architecture has several open problems needed to be solved in the future, such as the dynamic capability and flexibility of the model, the complex inference model, security risks, the requirement of extreme data rate, and interoperability of multi-vendor elements.

Near-RT RIC's current implementation of the TS xApp is relying on the currently available AD and QP xApps. TS is critical in the RAN. We think that better ML models with higher accuracy should be applied in the AD and QP xApps. This will improve the overall TS xApp performance. Another problem arises from the design of the present O-RAN RIC. RIC is not able to understand the connections between the same subscriber or other UE. Instead, RIC only owns high-level insights in a connection. This increase the risk for false-positive alarms/actions, especially without IMSI/IMEI correlation across connections. Some more suggestions are also raised by [72] to improve xApp and RIC platform. These suggestions include smaller xApp package, support for time-series handling, support for message queue, scaling out method, support for more AI technologies (online training, reinforcement learning (RL), and federated learning), and support for hardware acceleration.

O-RAN's RIC consumes a non-negligible amount of resources. Additionally, O-RAN's RIC has forward compatibility limitation because it is coupled to specific implementations, such as Redis or Prometheus databases. As a result, applications must poll these databases frequently in order to find new agents. FlexRIC is introduced by [73] to answer these problems. FlexRIC is an software development kit to build service oriented controllers. It is made up from a server library and an agent library. FlexRIC has two innovative service model with proof-of-concept prototypes for RAN slicing and traffic control. Results show that FlexRIC uses 83% less CPU than O-RAN's RIC while cutting the round-trip time in half.

There are several existing research relating to Near-RT RIC xApps. A method that enables fault-tolerant xApps in the RIC platform, known as RIC fault-tolerant (RFT) is proposed by [74]. The RFT is deployed so that the Near-RT RIC is more tolerant towards faults, thus preserving high scalability. Results show that the RFT can meet latency and throughput requirements as the number of replicas increases [74]. Another xApp called NexRAN is proposed by [75]. NexRAN is an open source xApp that can perform closed-loop controlled resource slicing in O-RAN ecosystem. NexRAN is executed on O-RAN Alliance's RIC to control the RAN slicing in srsRAN. NexRAN testing is done in POWDER research platform. Testing results show that NexRAN can successfully perform the RAN slicing use case. An RRM xApp based on RL is proposed by [76]. The xApp dynamically adapts the per-flow resource allocation, modulation, and coding scheme to meet the traffic flow KPI requirements based on the network status. The xApp testing is done on OAI LTE RAN in a small laboratory setup. Future testing should be done to determine the performance in largescale, heterogeneous, and non-stationary scenarios.

O-DU or the DU generally is one of the most discussed topics by wireless mobile communications enthusiasts. In section IV, we have mentioned that O-DU is responsible for L2 and some L1 processing. It is also mentioned that accelerators are included in the O-DU specifications. Back to section II, we also have pointed out that through NFV, it is possible for the operator to use a virtual DU (vDU). However, there is currently a problem with the DU implementation. Since vDU is deployed in COTS server, the current majority of vDUs rely on x86 architecture. The reason for this is that most COTS servers still use the x86 architecture. The problem is x86 architecture cannot meet the O-DU requirements under special cases like meeting the numerology 4's 1 us latency requirement and handling heavy UE traffic load. In these cases, the x86-based processing can get extra support from accelerators.

Regarding accelerators, O-RAN Alliance has already defined some specifications for the use of accelerators in the new documentation. In [63], the term lookaside and inline accelerators are explained. However, these specifications still lack the necessary details that are needed for standardization. The method to access or utilize those accelerators; the accelerator's APIs design and usage; and the definition of whether the accelerator is "dedicated or not" are not defined yet by O-RAN. Since accelerators are used in O-DU low and Intel is the company responsible for developing OSC's O-DU, the only accelerator that is compatible with the O-RAN code reference in the market right now is only the field-programmable gate array (FPGA) by Intel. Nevertheless, there are vendors that implement the DU independent of Intel such as the use of Nvidia graphics processing unit (GPU) and Xilinx's FPGA [77] as accelerators. Qualcomm will also introduce its own accelerator card to bring more competition to the market.

It will not be surprising that in the future, there will be a competition between the use of lookaside and inline acceleration. Survey shows that larger companies prefer inline accelerators while smaller companies would choose otherwise [78]. Intel will probably prefer the lookaside acceleration method that they are currently using. ARM-based accelerators probably must use the inline method unless they want to prepare a powerful CPU for doing this. Aside from this competition issue, the other challenge will be concerning the inline accelerator stack split. O-RAN Alliance should determine whether they want to standardize the accelerator stack split or give freedom to the vendors to determine the split each of them uses.

The use of accelerators makes vDUs more expensive than regular DUs [4]. Another solution to x86's need for accelerators is changing to reduced instruction set computer (RISC) or RISC-V based processors like ARM or C5 that have more powerful computation power for the L1 and L2 processing that O-DU needs to handle. This solution will eliminate the need of using accelerators. Besides having better computation power, ARM-based processors are also more cost-efficient compared to traditional processors [79]. We will dig into details about this comparison in the next subsection.

Aside from the accelerator issue, there is interesting research that studied the RU - DU resource assignment problem. The construction and maintenance cost of the O-DU is proportional to the amount of O-DU resource the cluster of connecting O-RUs use over the operating time. Operators are looking for intelligent assignment strategies for minimizing this resource, thus saving energy and providing free physical resources to support other services. The research done by [80] mentioned that an efficient RU and DU resource management can be achieved by embedding a deep RL (DRL) based self-play approach to RAN architecture, thus making the operation of RAN more cost-effective. Future research could be directed towards RU - DU resource assignment with multiple resources and large-scale RAN which would lead to cost and latency problems.

O-FHI's M-Plane is currently still being defined, just as we have mentioned before. Regarding this problem, a design of M-plane for the next 5G O-FHI is proposed [81]. It is planned that M-plane will support the initialization, configuration, and management of the O-RU. They also embedded several functions for M-plane, including a "start up" installation, software configuration and performance, software fault, file management, and many more. The M-plane is designed based on O-RAN's YANG model and includes O-RU controllers, which is powered by network configuration (NETCONF) protocol over Secure Shell (SSH).

As it happens, O-RAN Alliance has released a few specification documents for the M-Plane, which are [82], [83]. Security in the M-Plane has received attention. The prior specification uses SSH version 2 with a simple public key and password authentication. This approach has weak security, does not meet industry best practices and violates USG guidance [47]. Transport layer security (TLS) with public key infrastructure (PKI) X.509 certification usage in the M-Plane was recommended for an improved security [47], [84]. Nonetheless, this new additional specification is optional for the operators to consider. For further details, please look into [82].

While O-RAN Alliance has already specified the security protocol for the M-Plane, no requirements are defined about the CUS-plane. The reason might be that adding a security protocol would compromise the latency budget and performance. The usage of MAC security (MACsec) for the CU-plane security of the O-FHI is proposed [85]. MACsec will be added with an ephemeral key exchange in the CP while MACsec authentication only is proposed for the UP. MACsec is actually an optional security protocol in the eCPRI, so this proposal is well aligned with the O-RAN Alliance specifications. MACsec is based on IEEE 802.1X and provides a simple and fast security solution compared to the internet protocol security (IPsec) standard in 3GPP that is not dedicated to speed. Theoretically, the throughput for small packets using MACsec is almost double the IPsec. Future research direction could be directed to prove this through field tests. Another option to be considered is using WireGuard.

As we have mentioned before, 3GPP defines a flexible functional split for DU and RU. O-RAN Alliance chooses the split 7.2 option in its architecture. Research done by [86] shows that flexible functional splits have more benefits compared to fixed split options. The research focused on the placement of RAN slices in a multi-tier 5G open RAN architecture. The problem was formulated mathematically considering different functional split options for each network slice separately. Results showed that flexible functional split outperforms fixed functional split in utilizing physical network resources and satisfying different network slice requirements. Another work also suggested a dynamic functional split to solve the timing requirement of connecting an O-DU or O-RU with a DU or RU that does not follow O-RAN delay requirements [6]. Further research should be done to determine whether O-RAN Alliance should upgrade its architecture specification to support a flexible functional split.

O-RU has an issue related to the O-FHI. Various vendors have followed the O-RU specifications that O-RAN has supplied. Nevertheless, some of these claims are false. In spite of claiming that their RU are O-RAN compliant, these vendors still design their RU with their own proprietary standard. Some of them do not even use the O-FHI library that OSC has provided. This poses a huge problem because when the RUs want to be tested with the O-DU, they will fail. As a solution, these vendors will modify existing or create new codes either on the DU or the RU side. While this method solves the problem, following the O-RU standard at the start of the RU designing process is more recommended.

One interesting research direction for the O-RU is related to the beamforming process. The beamforming process has become one of the solutions to reduce total power consumption in joint optimization schemes [87]. It was made to decrease the number of BBUs required to handle network traffic, maximize the total input from users by defining physical resource blocks, enhance the precision of radio connection, increase throughput and number of parallel connections in a given cell area [67], [87]. It can also minimize energy costs in mobile cloud and network considering QoS. Despite all these benefits of beamforming, this process still needs high cost and complex design when applying it in a real environment. To overcome these problems, 3GPP and O-RAN Alliance have developed several functional split options. As explained in section IV, O-RAN's architecture uses the 7-2x split option, which leads to the more simplified communication interference between O-RAN units, especially in FH. In section IV, we have also introduced several beamforming methods proposed by O-RAN Alliance.

The last method proposed by O-RAN Alliance, the channel information-based beamforming method, is a digital beamforming process and can be considered as the best method out of these four methods. This method offers steerable beams, by using antennas that have dedicated radio signals and radio paths, enabling it to provide a high number of beams that can be transmuted dynamically. These beams are generated by O-RU based on the channel information from O-DU. However, implementing the pre-coding in O-RU can produce inter-RU interference. To overcome this problem, a zero-forcing algorithm (ZF) is proposed [67]. ZF is an algorithm that can be implanted in the beamforming process and solves the inter-RU interference problem. The ZF beamforming is a beamforming method that can mostly be used in massive MIMO. This method allows transmitting and receiving data to specific users while eliminating interference from these users. It is concluded that ZF can really reduce the channel interference in its interface so that the high transmission bitrate can fulfill the low latency requirements by utilizing the advantage of low latency in the FHI. However, ZF implementation still needs several further research. Even though ZF can really reduce the channel interference and the low latency requirements can be fulfilled, it is still unexplained whether the channel status information (CSI) also has met the low latency requirements for URLLC or not.

O-Cloud might not be mentioned as frequently as other parts of the O-RAN architecture in research papers. However, there is research that proposed deployment strategies for the virtualized O-RAN units in the O-Cloud [88]. With embedded intelligence, O-RAN is considered a SON. O-RAN system has to maintain its performance and availability autonomously while maintaining the quality of service. Self-healing is a key feature in intelligently handling and managing failures and faults in the network. The proposed optimized deployment strategy aimed to minimize the network's outage while complying with the performance and operational requirements. Binary integer programming was used to optimize the placement of the VNFs and their redundant ones. The model evaluation showed that the proposed strategy could maximize the network's overall availability.

B. *Open RAN Field Issues*

This subsection will discuss broader issues that are related to the open RAN field. Overall, there are three challenges that network operators may face regarding the open RAN field, which are implementation and standardization problems; improving performance and cost; and secure open RAN. This subsection will be closed by other open issues that may probably occur in an open RAN architecture.

Implementation and standardization problems is the first issue that we will discuss. If we talk about standardization, it is directly related to the O-RAN architecture, because as mentioned in previous sections, O-RAN Alliance has the most progress in developing the open RAN framework compared to the other open RAN projects. O-RAN Alliance has also provided detailed reference design specifications, complete with very clear documentation for each reference design. Even though O-RAN has provided such clear documentation, but still, the alliance lack robust, deployable, and well-documented software. Most of these O-RAN frameworks can not be used in real life, in actual networks [3]. This is caused by several reasons, such as the open source components proposed by O-RAN Alliance are still incomplete, require additional integration, require further development for actual deployment, or contain components that lack robustness.

The difficulty of keeping pace with recent architecture standards is also a challenge. The cellular network community feels pressured because they have to keep updating themselves to the specifications and technologies that have just recently been developed in new communication, networking, and programming standards [3]. The most widely known example of these is the NR and the mmWave architecture which were introduced as 5G enablers by 3GPP. Both of these architectures are currently being deployed in closed source commercial networks. The current problem with these developments is the RAN software libraries (OAI-RAN and srsRAN) are not fully developed for supporting NR yet, because these developments still require further coding and compatibility testing. The same thing goes for mmWave, where testing of the mmWave architecture for 5G technology is still not possible because of complexity. This complexity is caused by the lack of accessible open hardware for the mmWave software to run, and it needs several beam management testing for the software to be able to run properly.

With the problems above, operators face dilemmas if they choose to migrate their current RAN assets to open RAN deployments. These decision points are summarized by [89]. The first question is whether operators should use the current open RAN standard specifications, which are still incomplete as we have mentioned, to open up the interfaces in vRAN or wait for the specifications to be more mature. The next question is whether operators should introduce open RAN first in smaller networks before targeting macro RAN. While there is no simple answer and each operator will take different approaches, AT&T recommends that open RAN implementation be introduced in incremental modules [47]. Operators can launch open RAN modules to their architecture bit by bit before totally changing the whole of their RAN architecture to open RAN. Survey shows that small cells will be the initial focus of the vRAN deployment by operators [89]. Testing out in a localized network first before going to the main macro RAN would be a wise resolution [47], [89].

The other question that is related to keeping pace with the open RAN standard is whether operators should migrate their RAN assets to disaggregated and virtualized options simultaneously [89]. The counterargument is that it is less risky to successfully disaggregate the BBU first before considering cloud deployment. While the benefits of using both disaggregation and virtualization concurrently are attractive, survey shows that 25% of operators will implement them separately at first [89].

Through openness, open RAN has unlocked the possibility of interoperability amongst different vendors. Although economically beneficial, this opportunity poses new emerging challenges. Different vendors will use a wide range of components. It would be difficult to establish the location of bottlenecks that reduce the overall performance while using the heterogeneous components. In addition to the multiple vendor challenge, the requirement that open RAN should also be able to interact with legacy 4G equipment will rocket the challenge even higher.

Some questions are also raised regarding the multivendor environment in open RAN. When the RAN experiences a problem, how do operators decide which vendor should solve it? What is needed and how the operator takes this decision in a multivendor environment remain open matters. There is also a concern on whether operators should directly deploy multivendor open RAN or start working with a single vendor first before reorganizing the supply chain to a multivendor environment at a future stage. Survey shows that about 40% o operators expect to work with only 1 to 2 established vendors at the beginning [89].

Integration challenge is the next issue relating to open RAN implementation. Ericsson mentioned that it really concerned about open RAN's hardware integration [15]. With the complex multivendor ecosystem that open RAN offers, operators ranked integration issue as the second major obstacle with a 55% vote [90]. We know that openness has become a fundamental pillar of open RAN from the very start [14], where in an OpenRAN architecture there are hardware and software come from many different vendors, thus making it a multi-vendor architecture. Because the architecture involves components from different vendors, every service provider and mobile operator has to ensure that their networks continue to operate smoothly with all these different parties in their network [15].

There are actually 4 different open RAN system integration models to tackle this challenge [15]. In the first model, MNO will do the integration themselves, just like how Rakuten and Vodafone did. However, this approach requires the MNO to possess great in-house expertise. Integration will be done by the service provider in the second model. This model would perform well if the service providers have a lot of experience working with multiple resources. The third model assigns the integration role to the hardware or software vendor. An example of this is shown by Fujitsu who provides support in RU to DU integration for Dish [15]. Even so, it would be a challenge to ask vendors to integrate on behalf of their competitors [15].

The use of service from a system integrator, a new actor, is the last model. The system integrator is a third party that is not associated with a certain hardware or software vendor, yet works closely with the vendors to make sure their heterogeneous products work together [15]. Telefonica has provided an example of this by using services from Everis, a Spanish system integrator [15]. The appearance of this new role opens fresh business opportunities for companies who want to join the ecosystem. NEC, currently the most engaged system integrator, and others have sensibly taken this chance [91].

To guarantee that all providers and operators can operate their networks smoothly in open RAN, the RAN needs a new approach. This approach should prioritize a software-driven and open-minded ecosystem from hardware vendors, software vendors, system integrators, tower companies, real estate owners, regulators, industry bodies, and mobile operators. The integration of open RAN needs to be built for a software-centric world. In this software-centric world, each software talks to all physical components to deliver scalability and innovation; and change the game for how open networks are integrated. This approach will be mostly focused on open RAN's software, where the software will make those components smarter and interoperable. This approach will also help those components to be integrated and maintained remotely by using a software upgrade, so operators and vendors do not need to climb network towers anymore. To integrate open RAN, there are two levels of integration needed to be done, which are open RAN ecosystem integration and open RAN software system integration. Ecosystem integration means the system integrator will be responsible for integrating across the entire architecture, including integrating open radios and BBU software. Meanwhile, software system integration is a process that mainly focused on COTS hardware. In this integration, open RAN can achieve its automation, by using the same DevOps tools and the same CI/CD principles, thus making the integration more simplified.

Integration and interoperability are critical challenges in open RAN. Significant effort and collaboration are needed to test the open RAN system. Meta mentioned that a standard development environment; standard optimization metrics; and standard test and validation methodologies are required to unlock the full potential of open RAN [47]. This is where TIP, O-RAN Alliance, and other open RAN standard organizations play an important role to help mitigate this challenge. Testing methodologies, testing centers, WGs, and plugfests have been made as we have discussed in section III. Future research should be done in this direction. OSC itself offers standardization labs called OSC Community Labs [47]. Further information on OSC Community Lab will be explained in section VII.

A pipeline for designing, training, testing, and experimental evaluation of DRL-based control loops in open RAN was proposed by [92]. The proposed pipeline is called ColO-RAN. ColO-RAN addressed the development and adoption of DRL in real networks. This problem was caused by the unavailability of large-scale datasets and experimental testing infrastructure. ColO-RAN enables ML research using O-RAN components. The capabilities of ColO-RAN were showcased on Colosseum and Arena testbeds. Performance evaluation of ColO-RAN was done by developing three xApps to control the RAN slicing, scheduling, and online training of ML models. ColO-RAN and its related dataset will be made publicly available to research communities around the world.

The last issue relating to open RAN implementation cross-examine open RAN's credibility in the promise of a multivendor environment and lower entry barrier for new vendors. It is impossible to avoid Intel's products if an operator wants to build virtualized open RAN network in 2021. Wider than the giant company, this problem roots in the fact that the current industrial-grade virtualization in COTS server still mostly runs in x86-based processors. To make things worse, the majority of leading open RAN software companies have all based their products on FlexRAN which is x86 dependent. Fortunately, the open RAN ecosystem is expected to be more diverse in the future. Nvidia's BlueField-3 Arm-based data processing unit will compete against x86 processors. We also have mentioned several other competitors in the previous subsection. Amazon has already used ARM processors in some of its COTS and other companies will presumably follow soon.

As with any other technology, the cellular network market and industry will try every possible solution available that gives them better performance and lower price. ARM claims that its server platform would provide up to 60% lower upfront infrastructure costs, up to 35% lower ongoing infrastructure costs, and up to 80% cloud infrastructure cost savings compared to traditional server [79]. Looking at the performance difference between ARM and traditional servers, inferring they are x86 based, ARM could become more popular in the future. Furthermore, Arm has officially joined the 5G open RAN policy coalition. Eventually, ARM vendors would emerge such as NXP and Qualcomm. Nevertheless, ARM products presently are more custom built and majorly are not general processors. In addition, Arm still lacks the flexible industrial-grade virtualization capability that rivals x86.

Improving the performance & cost is the common theme of the evolution process of the RAN as we have explained in section I. This improvement process continues endlessly, including the current open RAN era. One of the problems related to improving O-RAN and open RAN performance is the resource scheduling process. The resource scheduling process is heavily related to network slicing. We have learned that network functionality is abstracted from its hardware and software components [3]. This abstraction leads to the formation of slices, and each slice is given a set of functions, services, and system resources, making it an independent dynamically created logical entity that can support UEs with multiple needs [93]. However, due to increased demands, each slice has been forced to do its own dynamic management. The management process has become critical and more challenging, especially when it comes to resource scheduling. Because of this reason, the slice needs new mechanisms that needed to be developed, which requires a more optimized way of allocating resources.

There has been a lot of research that discusses improving resource scheduling and network slicing process in open RAN. Recent research by [93] provided a new algorithm in 5G's slices called dynamic scaling multi-slice-in-slice-connected user equipment services for system resource optimized scheduling (DMUSO) algorithm. Like its name, DMUSO algorithm uses a concept where in a UE, slices are connected to each other, so this concept is called multi-slice-in-slice connected UEs. This concept connects slices to services and resources to slices, not among slices in a UE. DMUSO algorithm can improve the 5G system performance, by learning the service demands, data rates, resources, bandwidths, efficiency, transmission rates, and channel that are related SLAs for a user equipment's specific level in network slicing. Results showed that DMUSO achieves efficient and optimized system resource scheduling, with significant performance gains from 4.4 times to 7.7 times, compared to other methods and algorithms. Further researches related to DMUSO are still open, and [93] still has to analyze the effect of varying user equipment mobility on resource scheduling across slices in the future.

There is also a resource allocation improvement issue in open RAN. A resource allocation optimization model that can minimize the cost of updating a RAN infrastructure is proposed by [12]. The model allows the architecture to have a hybrid combination of different hardware and software generations, by taking costs involving links, cell sites, and the capacity limit of RAN resources into account. Another method proposed to solve this resource allocation problem is hierarchical orchestration. Hierarchical orchestration is proposed to address over-simplified resource allocation and limited support for different network segments for the current one-size-fits-all orchestrators in E2E networks [94]. The E2E network is divided into three segments, which are RAN segment, transport network segment, and core network segment. Every segment has its own distributed orchestrator. These distributed specialized orchestrators enable independent management of each segment.

Hierarchical orchestration also introduces hyperstrator, a higher-level orchestrator to coordinate the distributed orchestrators and deploy slicing process across multiple network segments. The hyperstrator works as a central point, and it is interacting with the whole E2E network. Therefore, hyperstrator has many tasks related to its orchestrators, one of them is to ensure cohesive performance across segments and slices for guaranteeing consistent QoS of the network slicing. From this proposed architecture, experiment results show that the hierarchical orchestration approach can leverage the capabilities of existing orchestrators and their communities to achieve a remarkable resource allocation in E2E networks. However, this research still more further studies. Future research should be directed towards identifying requirements, classes, and relations of ontology for hierarchical orchestration protocols. Also, future research should define a systematic translation, make a model, evaluate, and assess the new hierarchical orchestration model.

A low-complexity, closed-loop control system for Open-RAN architectures to support drone-sourced video streaming applications is proposed by [95]. Flying drones has a higher likelihood of having line-of-sight propagation compared to UEs which can lead to performance degradation in high data rate transmission. The control system jointly optimizes the drone's location in space and its transmission directionality to minimize its uplink interference impact on the network. The proposed system was prototyped and tested in a dedicated outdoor multi-cell RAN testbed. Numerical simulations are also used to evaluate the system. Results show that the control scheme achieves an average 19% network capacity gain compared to traditional BS-constrained control solutions.

Building a 5G network that is energy efficient is also an important issue. Increased network capacity, geographical coverage, and increased traffic demands require network densification and will lead to more energy consumption. The negative impact of exhaustive energy consumption will not only degrade business profits but also impact the environment. State-of-art applications of ML techniques used in the 5G RAN to enable energy efficiency are reviewed by [96]. Recent research focuses on functional split technique for green open RAN [97]. An RL-based dynamic function splitting (RLDFS) technique that decides on the dynamic function splits among DUs and the CU in an open RAN system to make the best use of RES supply and minimize operator costs was proposed. Performance evaluation was done using a real data set of solar irradiation and traffic rate variations. Results showed that the proposed RLDFS method makes effective use of renewable energy and is cost-efficient.

Other researchers also proposed F-RAN, or fog-computing-based RAN to improve RAN's performance and cost. This is another type of RAN, which its architecture is based on fog computing. Fog computing is a term for an alternative to cloud computing that puts a substantial amount of storage, communication, control, configuration, measurement, and management at the edge of a network [98]. This fog computing can be applied to C-RAN to alleviate the constraints of FH and high computing capabilities in BBU pool [9]. F-RAN has several advantages, such as rapid and affordable scaling. This makes F-RAN architecture much more adaptive to the dynamic traffic and radio environment. Even though F-RAN is considered an excellent solution for covering C-RAN's weaknesses, it still has some challenges and open problems

Another method proposed by researchers to improve RAN's performance and cost is called as Air-Ground Integrated Mobile Edge Networks (AGMEN). AGMEN is proposed by [99] to integrate UAV-assisted network densification. In AGMEN, multiple drone cells are deployed in a flexible manner to provide agile RAN coverage for the temporally and spatially changing users and data traffic. There are also several key components in AGMEN, such as multi-access RAN with drone cells and UAV-assisted edge caching. However, AGMEN still has many challenges related to the device heterogeneity and dynamic nature of the architecture. Further research for AGMEN relates to mobile routing, multi-dimension channels, and UAV scheduling.

The other method proposed by [100] to improve RAN's performance and cost is PlaceRAN. PlaceRAN focuses on minimizing the computing resources and maximizing the aggregation of radio functions to optimize the placement of radio functions in virtual NG-RAN planning. PlaceRAN uses disaggregated RAN combination (DRC) concept and multi-stage problem formulation, thus enabling the management of units and protocols as a set of radio functions. The result achieved by [100] showed that PlaceRAN can reduce the number of DRCs in the network, by taking functional split options, one-way tolerated latency, cross-haul bandwidth, and passion bandwidth into account.

Secure Open RAN is the next popular challenge that is widely discussed. Every network connectivity should be deployed as secured as possible, and so does the open RAN and the 5G network. As users of network connectivity, we surely do not want to experience any fraud or data-stealing that will threaten our safety. Therefore, a secure open RAN should be seen as a big challenge, and this challenge should be discussed in further research.

Open RAN promotes the advantages of its disaggregation and interoperable pillars. On the other hand, these pillars also introduce new challenges regarding security. Disaggregation and virtualization is the theme of open RAN. However, decoupling of the software from the hardware expanded the security threat surface [101]. Major virtualization technologies are vulnerable to security attacks, including MEC, SDN, NFV, network slicing, and cloud [102]. O-RAN Alliance's SFG has noticed that decoupling, containerization, and virtualization vulnerabilities can be exploited by threat actors [103]. Other works have also mentioned some of the vulnerabilities relating to disaggregation, such as insufficient identity, credential, and access management; insecure interfaces and APIs; system vulnerability; and shared technologies vulnerabilities [3], [102]. The interoperable pillar of open RAN also introduces new security challenges. One of them is different vendors might use different degrees of security in their products [6]. Although vendors should practice security best practices, not all vendors will implement adequate secure management interfaces.

The same case happens with the standards developed by O-RAN Alliance. In O-RAN Alliance, there are two fundamental pillars, which are openness and intelligence. Each pillar introduces new security challenges. Additional interfaces and functions in O-RAN architecture add the area of threat surface [5], [6], [103]. Incorporating AI/ML in O-RAN architecture, which is known as the intelligence pillar, also adds a new surface threat targeting AI/ML-related functions in O-RAN. Other works have also discussed the new security challenges of using AI/ML in mobile networks [104], [105]. Another additional threat surface comes from using open source code. Actually, open-source software gives both advantages and disadvantages when it comes to security [3], [106]. However, extra security measures should still be taken due to the fact that adversaries could easily access the same open-source code used in O-RAN system and exploit its vulnerabilities [6], [106].

Thankfully, as we have explained in section III, O-RAN Alliance is working carefully to address the security challenges with its SFG. SFG specifies the security requirements in O-RAN system in [107] and [108]. Hopefully, O-RAN Alliance can develop a secure architecture, framework, and guidelines for its open RAN standards. Other references have also discussed some defense mechanisms to increase software security in addition to these requirements and protocols, such as authentication and access control; cryptography; secure virtualization; anonymity and obfuscation; resilience assurance; lightweight security based on physical layer; and intrusion detection mechanisms [102], [105], [109], [110]. SFG has also made modeling and analysis of security threats in the O-RAN architecture in [103]. SFG summarizes the threat surfaces, agents, potential vulnerabilities, and threats. Some parts of this analysis are also discussed by [111]. SFG specifies a total of 49 threats, divided into 7 categories, which are threats to O-RAN system, O-Cloud, open-source code, physical, 5G radio networks, and ML system threats.

Focusing on O-RAN system, there are a total of 35 threats [103]. We will mention some of the main threats for each O-RAN component. A common threat to all the O-RAN components is the exploitation of insecure design, weak authentication, and weak access control in O-RAN components or network boundaries to compromise O-RAN components' integrity and availability. In the FH, the attacker could penetrate O-DU and beyond through accessing the O-FHI or the attacker could target one or more planes in O-FHI. In O-RU, attacker could set up a rogue O-RU. In Near-RT RIC, attacker could exploit xApps vulnerabilities or create malicious xApps. In Non-RT RIC, attacker could exploit rApps vulnerabilities or penetrate the Non-RT RIC to cause denial of service (DoS).

Overload DoS attack could also target SMO.

O-RAN TIFG has also provided some E2E Security Test Specification in [112]. From E2E perspective, the O-RAN system is just another gNB. Therefore, following security requirements, threats and test cases outlined in 3GPP TS 33.511 for gNB is a must for any O-RAN standard adopter. Other than that, TIFG also specifies optional test cases for S-Plane, C-Plane, and Near-RT RIC A1 interface; and O-Cloud. These test cases cover DoS, fuzzing, blind exploitation, and resource exhaustion types of attacks. Other references have also identified some of these threats and attacks, such as data breach; hijacking attacks; malicious insiders; data loss; DoS and distributed DoS (DDoS); abuse and nefarious use of services or resources; and caching attacks [102], [104], [109], [113].

Another method proposed to improve the network security is deploying zero trust architecture (ZTA) [101], [114], [115]. Integrating ZTA to 5G and 6G technologies has become a critical need, both in tactical and commercial applications [115]. ZTA is a solution to address security requirements in a network with unreliable infrastructure. ZTA has a dynamic risk assessment and trust evaluation as its key elements. Intelligent ZTA (i-ZTA), an AI-embedded-ZTA, is proposed by [115] to provide secured information in unreliable network infrastructure. ZTA can provide necessary computational resources and seamless, reliable, and robust connectivity.

Besides ZTA, another method proposed to deploy a secure open RAN is called blockchain (BC). BC is a network system that can establish transactional faith among peer entities on decentralized peer-to-peer (P2P) platforms while overcoming the vulnerability of centralized ledger host [116]. BC has also emerged as a potential tool in designing a self-managed and scalable decentralized network [109], [116]. The BC technology has been discussed by many researchers in open RAN and its various potential application scenarios, including IoT, MEC, smart grid, vehicular network, and smart city. BC can offer pseudonymity, which is one of the defense mechanisms to keep its users' privacy [117]. There are several main benefits if we integrated BC to open RAN and its applications, one of them is reduced communication cost and reduced delay that are required to establish agreements facilitating a real dynamicity in RAN sharing [109], [116], [117]. However, incorporating BC into RAN still needs to be further studied. Future research should mainly focus on improving the latency, stability, and scalability of BC; developing a green mining mechanism for power-limited node devices; and finding ways to lower the cost of incorporating BC into RAN technology [109], [116], [117].

Other Open Issues relating to the open RAN field will be discussed in the following paragraphs. A lot of challenges and future research directions for open RAN and O-RAN has been covered in the previous paragraphs. However, there are still issues related to open RAN that do not belong in previous categories. This last part will discuss those challenges and future directions.

As written in the previous implementation and standardization problems paragraphs, we can see that one challenge that needs to be researched deeply is the lack of accessible open hardware for software to run. This is caused by limited contributions to RAN open-source software. It seems like there are more network operators and vendors who are trying to develop open-source CN and MANO frameworks, such as OMEC [118], Magma [119], Free5GC [120], and Open5GS [121]. At the same time, they are not really trying to develop RAN-related project frameworks, like OAI-RAN or srsRAN. The development of RAN is basically done mostly by researchers and small companies that have limited manpower and resources [3].

This limitation should be addressed because limited contributions results in limited feedback for the open RAN standard developing organizations. This poses a serious problem since the organizations cannot verify the viability of the standard that has been developed. The major vendors and operators should be more involved in developing RAN-related projects because nowadays the lower layers of RAN stack have become more and more important each day. These lower layers have become sources of intellectual property and product-bearing revenues for telecommunication businesses.

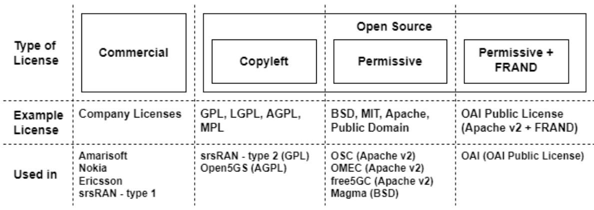

Fig. 21. Open source 4G/5G software license types

Relating to this contribution problem is the different licensing model for 3GPP technologies. There are different licensing models for 4G/5G technologies software, which can be seen in Fig. 21. As we have mentioned, most RAN software are closed-source, commercial, or copyright, such as software from Amarisoft, Nokia, and Ericsson. Open source software itself can be divided into permissive and copyleft licenses. Both of them allow developers to copy, modify, and redistribute code freely on open source or commercial products. However, copyleft license obligates the developer to open their altered source code under open source so it can be publicly available, such as the GNU General Public License (GPL), GNU Lesser General Public License (LGPL), GNU Affero General Public License (AGPL), and Mozilla Public License (MPL). This requirement is undesirable for companies using the source code because they have to share their trade secret. Permissive license does not require the developer to publicly release their source code if they use permissive licensed open source software. However, some permissive licenses do not retain the intellectual property rights. This creates potential future problems for companies using the open source software for their commercial products. Examples of this type of license are Berkeley Software Distribution (BSD), Massachusetts Institute of Technology (MIT), and Apache v1.1 License. Apache License v2.0 is an example of another type of permissive license that retains the intellectual property rights.

From these different licensing models, we can see that companies can suffer a loss in using the wrong kind of open source software. Tackling this obstacle, OAI has its own software license model to take OAI-RAN to be licensed under OAI public license, thus allowing their users to easily use their open source software. The OAI public license allows parties to contribute to the source code while retaining their intellectual property rights. This effectively provide incentives for the community to use OAI as a reference implementation in their research and development or productization.

Responding to this problem, O-RAN Alliance has also made its initial step to deploy and open and softwarized RAN using Apache License v2.0. Hopefully, through the initial step that O-RAN Alliance took, other wireless communities, operators, and vendors can follow O-RAN's path by increasing their support toward the development of complete and truly open RAN, and also to open source RAN-related projects. Fortunately, more organizations are now expected to contribute to open RAN. IEEE Standard association has initiated an open RAN Industry Connections Activity to help the existing open RAN industry efforts [122]. Hopefully, this program will accelerate and facilitate collaboration between organizations and individuals in the open RAN field.

The next open issue related to RAN is the latency issue. Rakuten's open RAN deployment in Japan is a real-life example of this challenge. Even though the overall performance of Rakuten's open RAN mobile network is rated "very good", the latency score shows room for improvement relative to major operators in assessed cities [20]. Latency problem is more serious in URLLC network slice that requires low latency. Using vBBU could result in higher latency [71]. Backup strategies for several use case scenarios should be explored in the next research.

A gradient-based scheme is proposed by [123] to overcome the latency problem. This scheme is used to solve the minimum delay function placement and resource allocation for open RAN. There are many layers of RAN and open RAN allows these layers to be split and deployed as virtual functions and openly communicate with each other layer for service provisioning. An E2E mobile operator employing open RAN is modeled by [123], with a hierarchical mobile network architecture consisting of local, regional, and core data center layers. These three hierarchical layers are used to add flexibility in resource allocation and increase reliability, while taking the advantages of open RAN. In addition, the case where the traffic of a service function requests (SFRs) traverses multiple chains via a single path through containerized network functions (CNFs) is also modeled. From this formulation, a gradient-based solution was proposed to achieve the optimal solution efficiently. This solution is applied through an algorithm called the gradient-based minimum delay (GBMD) algorithm. This algorithm can serve up to 90% E2E network delay decrease.

Another issue that still exists in open RAN is the scalability issue. We already know that open RAN is already more scalable than its predecessors, thanks to NFV, SDN, and automation that caused the RAN more scalable and more flexible, both in its management and in its orchestration [3]. Open RAN's framework relies heavily on AI-embedded software to maintain its scalability and flexibility [124]. RIC is also responsible for making open RAN scalable. Both Non-RT and Near-RT RIC has made the RAN more scalable and adaptable. The Near-RT RIC provides a secure and scalable platform, thus making it possible to control its xApps in a flexible way [22]. OSC's Dawn release has also deployed a Bouncer xApp, specifically to load and test xApps' scalability [44]. All of these are done with the purpose to make open RAN more scalable, leading to a more effective RAN, especially in its performance.

However, the scalability in open RAN still needs to be improved. There are several concerns regarding the open RAN scalability [21], [90]:

1) Operator who would want to adopt the cloud-native open RAN need to expect the cloud scaling challenges [21]. Scaling the DU virtualization across servers will be a real-time problem. Additional scaling constraints will emerge if operators take the use of accelerators into consideration. Furthermore, containerized orchestration alone cannot solve the high network availability and reliability operational goals. The applications will need additional built-in state synchronization and data integrity consideration. Moreover, specific failover and availability mechanisms will be required at the protocol level.
2) Most open RAN enthusiasts assume that the architecture will bring significant economic savings while ignoring the cost of operating the complex architecture [90]. We have explained that there are a variety of new business roles involved in the open RAN ecosystem, such as the system integrator. The question is whether the service expense for these roles and OpEx in a multivendor architecture at a large scale will exceed the cost-saving promised? The total TCO with large-scale operations is still undetermined [21].
3) The existing gap between scalable automation and AI capability is another challenge to be considered [21]. The AI/ML technologies implementation is relatively new in the telecommunication context. Using AI/ML in telecommunication grade operations will require significant resource expense. While interface standardization is currently available, data access, pipelines, and validation cannot fully scale yet due to a lack of standardized network configuration and performance data exchange. Furthermore, large-scale AI/ML assets deployment and operations by operators in live networks are still rare and complicated. To successfully scale and operate AI/ML use cases, harmonization of expertise from telecommunication, data science, and cloud/big-data field is

required [21].

To answer the scalability challenge, OAI provides the Trirematics operator as a solution. Trirematics operator is an intelligent software operator for RAN and CN deployment scenarios [125]. Trirematics' orchestration and management framework is cloud-native and extensible. Trirematics include features, such as intelligence, agility, automation, abstraction, maintenance, and observability. Trirematics makes intelligent and agile decisions in deployment and operation of the E2E network. It automates the lifecycle of the network entities and abstracts complexities in 5G environment while providing enough diagnosis and control for its users. It also has extensive observability features, such as log processing, alerting, metric processing, and health monitoring.

Rakuten Mobile's commercial deployment in Japan is proof that open RAN is ready for prime time. Dell'Oro Group's research report states that open RAN is expected to acquire more than 10% of the RAN market share by 2025 [126]. It is predicted that the whole vRAN market will see a 60% compound annual growth rate (CAGR) and the open RAN sector will experience 124% CAGR between 2020 to 2026 [89]. It is forecasted that the total cumulative open RAN revenues will hit 10 to 15 billion USD between 2020 to 2025 [127]. While the majority respond enthusiastically, [4] view this result as proof to doubt the open RAN adoption by the market. Whether these data and predictions be viewed positively or negatively is in the hand of each impacted party. Huawei clearly states that they do not support open RAN or vRAN. Open RAN is seen to possess many challenges such as standardization, OpEx, and security issues that we have explained above. In addition, there might be political reasons involved behind the open RAN movement. This notion was researched and open RAN was considered a geopolitical hijacking case [128]. By considering open RAN as a social imaginary of openness and trustworthiness, open RAN can be weaponized to outcompete rivals by parties ranging from industry to government. Whether this is true or false is still debatable and hopefully, the truth will shine as time proceeds.

Whichever the case, AI/ML's implementation will play an important role in the future of the cellular network. The 5G technology challenges show that it is also required to implement more AI/ML in the future network technology. Actually, O-RAN Alliance is applying this by building ML directly into the network architecture through RIC. We have also mentioned how researchers use ML to provide solutions for the problem in O-RAN architecture and open RAN generally. Several additional benefits of implementing AI/ML into 5G communication systems are mentioned in [129]. First, AI/ML can help with problems related to the complicated and dynamic wireless communications channel. AI can help in channel measurement data processing, channel modeling, and channel estimation. AI can also help in physical layer research of the 5G network. Massive MIMO arrays create enormous volumes of data which is well suited to be analyzed by ML systems. ML can also improve the performance in signal processing. Data-driven localization using ML algorithms in 5G systems can also be a valuable application. AI can also improve network management and optimization in 5G systems. Model-based optimization can be replaced with data-driven optimization in ultra-dense networks. AI/ML technologies also enable a paradigm shift of wireless networking from reactive to proactive design.

However, [129] also mentions that there are some limitations of the current AI/ML technologies that prevent them to be directly applicable to the current 5G system. For example, ML algorithms are primarily developed for systems and applications that do not require high-frequency performance. On the other hand, 5G network requires high data processing rates for URLLC. Existing ML algorithms also rely on powerful processing hardware. On the contrary, communication systems are full of devices constrained by storage capabilities, computational power, and energy resources. Adjustments and innovations in AI technology should be done before it could be fully compatible with the needs of 5G network. One interesting research direction is to develop distributed ML algorithms for 5G network. Previously, we have mentioned how MEC and fog computing could be solutions to improve the performance of open RAN and 5G networks. The main idea is to move from centralized to distributed systems. Parallel to this trend, ML algorithms in communications applications should also move from centralized to distributed. For example, cloud, fog, and edge computing networks can each be supported by a lightweight deep learning model. Furthermore, decentralized and centralized algorithms could be used in parallel, while balancing complexity, latency, and reliability.

The 5G technology system is required to meet users' demands from all over the world, by delivering connectivity anytime and anywhere [130]. Besides that, there are a lot of smart devices and advanced technologies that have been increasing rapidly, and these have also led to the increasing need for 5G connectivity systems. The 5G network is also demanded to fulfill multiple requirements. That is why the development of optical wireless communications has been rapidly increasing because it could be a potential solution to support high data rate [131].

Moreover, the 5G technology has also been required to provide services not only in big cities and metropolitan areas but also in remote and terrestrial areas, the areas that are uncovered or under-served geographically. Non-Terrestrial Network (NTN) can be an effective solution for 5G to provide omnipresent services by achieving global network coverage [130]. NTN is networks, or segments of networks, that use an airborne or space-borne vehicle for transmission [132]. 3GPP has also made several researchers and activities to support NTN as part of 5G technology, especially 5G NR architecture. NTN was specifically introduced by 3GPP in Rel-17 [58]. The NTN in 5G NR is still in the developing process, with a future guarantee from 3GPP that the 5G network can be operated in frequencies more than 52 GHz. We will know about this guarantee furthermore in Rel-18.

The 3GPP standardization has already completed the first 5G NR standard, specifically in Rel-15 [58], [130]. They have also started to work on several solutions to support NTN in 5G NR systems [130]. RRM has become one of the major issues in 5G NR, and this management strategy is relevant to offering tight cooperation between satellite and terrestrial networks. Through RRM, researchers are finding efficient resource allocation methods to decrease interference and provide real-time video services by implementing effective link adaptation procedures. Meanwhile, mobility management is also important to keep the network service's continuation by achieving seamless handover over heterogeneous wireless access networks. The novelty of 5G architecture embedded with NTN is in form of SDN and NFV-based-NTN; internet of space things (IoST); cognitive NTN; and non-orthogonal multiple access (NOMA) based NTN [130].

VII. Open Source Approaches

In this section, we will address some of the issues pointed out in previous sections from open source perspective. Open source has brought solutions into the information technology (IT) industry. With open RAN, the border between IT and telecommunication industry is getting blurry. Open source could be the key to solving open RAN issues and accelerating innovation in RAN technologies.

A. Mix and Match of Open Source Projects

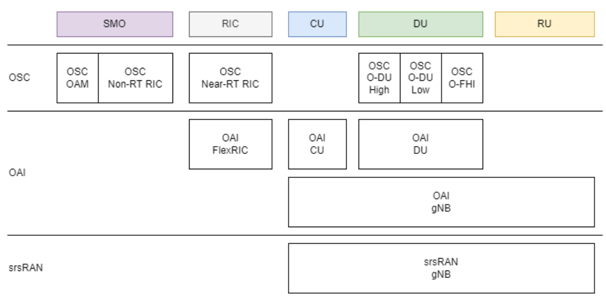

Fig. 22. Open-source software from OSC, OAI, and srsRAN

In section V, we have mentioned that OSC has not fully realized the implementation of the whole O-RAN architecture. O-CU and O-RU projects are postponed, resulting in the unavailability of open-source software for the O-CU and O-RU. OSC's software could not be used to run the RAN E2E function because there are missing functions in the architecture. A possible solution to this problem is filling these gaps using other open-source projects, such as OAI and srsRAN. For example, we can use OAI's CU to replace OSC's O-CU function. Open source nature enables this mix and match flexibility. As a matter of fact, OSC has adopted this approach in developing its SMO. OSC uses products from ONAP to fill in several functions in their SMO, such as OAM VNF message stream (VES), A1 adapter, and A1 policy management service [44]. Fig. 22 shows the open-source RAN components software that are currently provided by OSC, OAI, and srsRAN. This mix and match approach is not perfect because only OSC is fully compliant with O-RAN Alliance standard. Further adjustments should be made for OAI and srsRAN software in order to properly use them in an O-RAN system.

B. Accelerating Innovation

Open source RAN software enables the RAN software to be accessible to the majority of people, thus bringing innovation to the RAN technology. Currently, academics and research institutions can use open source RAN platforms, such as srsRAN, OAI, and OSC, to provide solutions to challenges faced in by the industry in RAN technology. As a result, more and more research are done and papers published leveraging the open source RAN platforms [35], [49], [50], [74], [75], [92], [133]. Some of these researchers also open up their work to be used for free, thus widening the open source RAN platforms available. These researches help provide feedback on the shortcomings of open source RAN platforms and also on O-RAN Alliance specifications. The solutions proposed by these researches can also be used for new feature development in O-RAN specification. It can be expected that open source will be one of the research directions for the future of open RAN.

Currently, most of the 5G products use commercial software provided by software design house companies. Open source RAN software is not used for commercial products. Open source can be used as a reference design or a proof of concept. However, with the rise of open source RAN software platforms, we are starting to see utilization of open source software in commercial open RAN products. BubbleRAN is an example of a company that uses open source as its core technology [134]. BubbleRAN is a startup company that aims to accelerate innovation in the area of multi-vendor 5G through open source. BubbleRAN uses OAI's experimental RT wireless platforms and Mosaic5G's agile 5G platforms to develop their products. All of BubbleRAN's products would offer the basic functionalities provided in the open source platforms mentioned before. Extra features are provided by BubbleRAN on their commercial products to support the features of particular solutions required by their customers.

C. Interoperability and Security Testing

As we have mentioned in section VI, interoperability testing and security testing of open RAN are important issues to be addressed. Currently, the only available solution to ensure an interoperable and secure open RAN system is the standardized architecture description and testing procedure provided by O-RAN Alliance. There is no concrete testing system yet for vendors to ensure that their open RAN products are interoperable and secure. Open-source testing software could be the solution to the open RAN interoperability and security testing problem. The cybersecurity domain has implemented open-source testing tool solutions, such as OSSEC, Kali Linux, and OpenVAS. We think that the future direction of open RAN security testing can also be directed into open source solutions, following the example of cybersecurity testing trend. These open source security testing tools can provide the basic security testing requirements to meet the test case procedure requirements provided by O-RAN Alliance's TIFG in [112]. Companies can still sell advanced testing services or products that provide extra features compared to the open source counterpart.

The same concept could be done for interoperability testing. Open source interoperability testing tools can be developed to provide the minimum testing requirement for open RAN products. This solution can be used by vendors to ensure that their open RAN products are interoperable with open RAN other products. However, it does not mean that the system integrator role will vanish due to the existence of open-source interoperability testing tools. The open RAN multivendor ecosystem is still complex whether open-source interoperability testing tools exist or not. Although network operators can use open-source interoperability testing tools, using the services from system integrators might provide ease rather than forming an internal interoperability-focused team. System integrators can also provide advanced services for interoperability testing.

D. OSC Community Labs

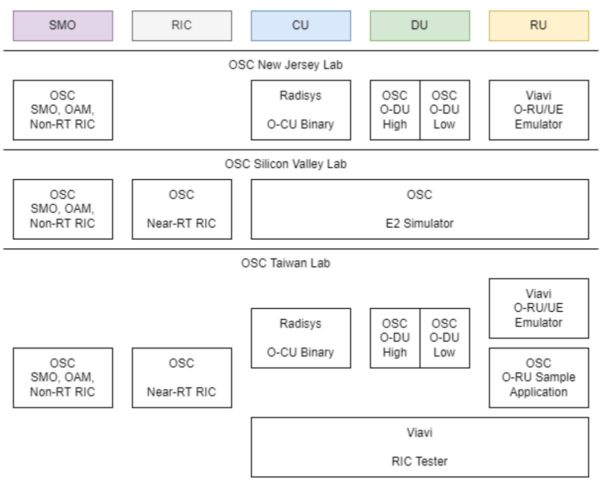

Fig. 23. Logical resources of OSC Labs

OSC Community Labs or OSC Labs are integration and testing platforms offered by OSC. OSC labs provide the hardware and software resources for OSC members to test their software. OSC labs' software follows the O-RAN specification. Currently, there are three OSC labs [47]. The first lab is located in New Jersey and is maintained by AT&T. It has full computing, storage, and networking support. It focuses on E2E integration and testing of OSC software components. It can also be used as a platform to demonstrate OSC projects. OSC New Jersey Lab is accessible since early 2021. Two other labs, one in Silicon Valley and one Taiwan, are still under construction. OSC Silicon Valley INT Lab is maintained by China Mobile Technology USA Inc. It mimics OSC New Jersey Lab's software and networking environment. It already supports RIC platform and xApp testing. OSC Taiwan Lab is located in Taipei. It is jointly maintained by National Taiwan University of Science and Technology (NTUST), the National Yang Ming Chiao Tung University (NYCU), and Chunghwa Telecom (CHT). It also targets to mimic OSC New Jersey Lab. It has already done several O-DU testing and integration. Fig. 23 shows the logical resources owned by each lab. OSC labs are important platforms for open source open RAN integration and testing purposes. OSC labs can be leveraged for future research in improving open RAN.

VIII. Conclusion

Open RAN is accelerating innovation and disrupting the cellular industry ecosystem. It aims to overcome the present challenges faced by the cellular network industry, through disaggregation and interoperability. Leveraging state-of-the-art technologies and approaches, open RAN promises to bring a multivendor ecosystem, flexibility, cost efficiency, and performance improvement. A lot of projects, activities, and standardization efforts have been done for open RAN. However, there are a lot of challenges in developing and implementing the open RAN idea into reality.

We have presented a comprehensive overview of open RAN and its importance; discussed its challenges and potential research directions; and address some of the challenges from open source perspective. We first summarized the evolution history of the RAN from traditional to vRAN. This helps identify the differences and development of the RAN technology. Then, we introduced the open RAN movement and the technologies that are related to the open RAN generation. After that, we discussed projects, activities, and standards that are related to open RAN. This helps to elaborate on the current condition of open RAN development. Next, we briefly explained the standardized open RAN architecture from O-RAN Alliance and its use cases. After that, we discuss challenges and future research directions for each component of O-RAN's standard architecture. A number of broader issues are also mentioned relating to open RAN and the future of RAN, including implementation and standardization; performance and cost improvement; security; and other open problems. Finally, we discuss how open source could potentially be the solution for open RAN challenges. Hopefully, this survey can serve as a useful starting reference for academia and industry to pursue further in-depth study on open RAN.

List of Abbreviations

2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GC	5G Core
AI	Artificial Intelligence
API	Application Program Interface
BBU	Baseband Unit
BS	Base Station
C-RAN	Centralized/Cloud RAN

CapEx	Capital Expenditure
CN	Core Network
COTS	Commercial Off-The-Shelf
CP	Control Plane
CPU	Central Processing Unit
CU	Central Unit
CUPS	Control Plane and User Plane Splitting
D-RAN	Distributed RAN
DL	Downlink
DU	Distributed Unit
E2E	End-to-End
eMBB	enhanced Mobile Broadband
eNB	evolved Node B
EPC	Evolved Packet Core
FG	Focus Group
FH	Fronthaul
FHI	FH Interface
FPGA	Field-Programmable Gate Array
gNB	Next-generation Node B
GPU	Graphics Processing Unit
LF	Linux Foundation
LTE	Long-Term Evolution
MANO	Management and Orchestration
MEC	Mobile Edge Computing
MIMO	Multiple Input Multiple Output
ML	Machine Learning
mMTC	massive Machine Type Communication
MNO	Mobile Network Operator
MVP	Minimum Viable Plan
NFV	Network Function Virtualization
NIC	Network Interface Card
NR	New Radio
NSA	Non-Standalone
OAI	Open-Air-Interface
ONAP	Open Network Automation Platform
ONF	Open Networking Foundation
OpEx	Operational Expenditure
OSC	O-RAN Software Community
OTIC	Open Test and Integration Center
PNF	Physical Network Function
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Networks
RIC	RAN Intelligent Controller
RT	Real Time
RU	Radio Unit
RRM	Radio Resource Management
RRU	Remote Radio Unit
SA	Standalone
SDN	Software-Defined Networking
SFG	Security FG
SMO	Service Management and Orchestration
SON	Self-Organizing Network
TIFG	Test and Integration FG
TIP	Telecom Infra Project
TCO	Total Cost of Ownership
UE	User Equipment
UL	Uplink
UP	User Plane
URLLC	Ultra Reliable Low Latency Communications
vBBU	virtual BBU
VM	Virtual Machine
VNF	Virtual Network Function
vRAN	virtual RAN
WG	Work Group

REFERENCES

[1] S. K. Singh, R. Singh and B. Kumbhani, "The evolution of radio access network towards Open-RAN: Challenges and opportunities," *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Apr. 2020.

[2] L. Gavrilovska, V. Rakovic, and D. Denkovski, "From cloud RAN to open RAN," *Springer Wireless Personal Communications*, vol. 113, pp. 1523–1539, Mar. 2020.

[3] L. Bonati *et al.*, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Computer Networks*, vol. 182, Issue 1, pp. 1 - 32, Dec. 2020.

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

[5] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," 2022, arXiv:2202.01032.

[6] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access network–what O-RAN can and cannot do!" 2021, arXiv:2111.13754.

[7] B. Brik, K. Boutiba and A. Ksentini, "Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 228-250, 2022

[8] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2282-2308, Thirdquarter 2016.

[9] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-Constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152-160, Apr. 2015.

[10] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 105-111, Oct. 2015.

[11] T. X. Tran, A. Hajisami, and D. Pompili, "Cooperative hierarchical caching in 5G cloud radio access networks," *IEEE Network*, vol. 31, no. 4, pp. 35-41, July 2017.

[12] P. A. de Souza, A. S. Abdallah, E. F. Bueno, and K. V. Cardoso, "Virtualized radio access networks: Centralization, allocation, and positioning of resources," *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018.

[13] P. Rost *et al.*, "Benefits and challenges of virtualization in 5G radio access networks," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 75–82, Dec. 2015.

[14] J. Kempf and P. Yegani, "OpenRAN: A new architecture for mobile wireless internet radio access networks," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 118-123, May 2002.

[15] *Everything You Need to Know About Open RAN*. Parallel Wireless, 2020. [Online].

[16] P. Marsch *et al.*, "5G radio access network architecture: Design guidelines and key considerations," *IEEE Communication Magazine*, vol. 54, no. 11, pp. 24-32, Nov. 2016.

[17] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng, "OpenRAN: A software-defined RAN architecture via virtualization," *2013 ACM SIGCOMM Conference*, Aug. 2013.

[18] B. P. Lin, "Toward an AI-enabled O-RAN-based and SDN/NFV-driven 5G & IoT network era," *Network and Communication Technologies*, vol. 6, no. 1, pp. 6-15, Jun. 2021.

[19] P. Hammond, "Open architecture vs. open source in defence systems," *BAE Systems Insyte*, pp. 1 - 7, Dec. 2007.

[20] "Audit report: First openRAN network Japan in international comparison," Umlaut, 2021. [Online].

[21] "Transition toward open & interoperable networks," 5G Americas, White Paper, 2020. [Online].

[22] C. Coletti *et al.*, "O-RAN: Towards an open and smart RAN," O-RAN Alliance, White Paper, 2018. [Online].

[23] "Open RAN integration: Run with it," iGillott Research Inc., White Paper, 2020. [Online].

[24] G. Brown, "TIP openRAN: Toward disaggregated mobile networking," Heavy Reading, White Paper, 2020. [Online].

[25] A. Akman *et al.*, "O-RAN minimum viable plan and acceleration towards commercialization," O-RAN Alliance, White Paper, 2021. [Online].

[26] J. Hodges, "Disaggregation and the virtual network," Heavy Reading, White Paper, 2019. [Online].

[27] O. Sunay *et al.*, "SD-RAN, ONF's software-defined RAN platform consistent with the O-RAN architecture," Open Networking Foundation, 2020. [Online].

[28] "Software-Defined Networking (SDN) Definition." opennetworking.org. https://opennetworking.org/sdn-definition/ (accessed Dec. 21, 2021).

[29] M. Fakoorrad, "Application layer of software defined networking: Pros and cons in terms of security," M.S. thesis, Dept. of Computer Science, Tallinn University of Technology, Tallinn, Estonia, 2016.

[30] D. Harutyunyan and R. Riggio, "Flexible functional split in 5G networks," *2017 13th International Conference on Network and Service Management (CNSM)*, Nov. 2017.

[31] A. Maeder *et al.*, "Towards a flexible functional split for cloud-RAN networks," *2014 European Conference on Networks and Communications (EuCNC)*, June 2014.

[32] M. Condolucia, and T. Mahmood, "Softwarization and virtualization in 5G mobile networks: Benefits, trends and challenges," *Computer Networks*, vol. 146, pp. 65-84, Dec. 2018.

[33] S. Zhang, "An overview of network slicing for 5G," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 80-87, June 2019.

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

[35] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, Oct. 2021.

[36] V. Krishnan and S. Hefty. (2019). Enabling Applications to Exploit SmartNICs, FPGAs, and Accelerators [PowerPoint slides].

[37] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322-2358, Fourthquarter 2017.

[38] Global System for Mobile Communications Association (GSMA). "5G-Era mobile network cost evolution," Global System for Mobile Communications Association (GSMA), 2019. [Online].

[39] "About O-RAN Alliance." o-ran.org. https://www.o-ran.org/about (accessed Oct. 19, 2021).

[40] S. Abeta, T. Kawahara, A. Umesh, dan R. Matsukawa, "O-RAN Alliance standardization trends," NTT DOCOMO Technical Journal, vol. 21, pp. 38–45, July 2019.

[41] C. L. I. (2018). From "Green & Soft" to "Open & Smart" [Powerpoint slides].

[42] "O-RAN Software Community." o-ran.org. https://www.o-ran.org/software (accessed Sept. 30, 2021).

[43] "About O-RAN Software Community (SC)." o-ran-sc.org. https://o-ran-sc.org/about/ (accessed Oct. 31, 2021).

[44] R. Jana, D. Kinsey and PTLs. (2021). O-RAN Software Community (OSC), D & E release epics [PowerPoint slides].

[45] O-RAN Software Community. (2022). OSC Release "F" Plan [PowerPoint slides].

[46] "Testing & Integration." o-ran.org. https://www.o-ran.org/testing-integration (accessed Dec. 29, 2021).

[47] IEEE Standards Association. (2021). Advancing open-RAN: Research, standards, and deployment. [Online].

[48] "OAI 5G RAN PROJECT GROUP." openairinterface.org. https://openairinterface.org/oai-5g-ran-project/ (accessed May 20, 2022).

[49] N. Nikaein *et al.*, "Demo: OpenAirInterface: An open LTE network in a PC," *2014 20th Annual International Conference on Mobile Computing and Networking*, Sep. 2014.

[50] N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "OpenAirInterface: A flexible platform for 5G research," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 5, pp. 33–38, Oct. 2014.

[51] "The srsLTE project is evolving." srslte.com. https://www.srslte.com/srslte-srsran (accessed Apr. 07, 2022).

[52] "ONAP 5G blueprint overview," Open Network Automation Platform (ONAP), 2018. [Online].

[53] "Our Mission." opennetworking.org. https://opennetworking.org/mission/ (accessed Dec. 30, 2021).

[54] "SD-RAN." opennetworking.org. https://opennetworking.org/sd-ran/ (accessed Nov. 21, 2021).

[55] "COLOSSEUM." northeastern.edu. https://www.northeastern.edu/colosseum/ (accessed Dec. 30, 2021).

[56] "About 3GPP." 3gpp.org. https://www.3gpp.org/about-3gpp (accessed Mar. 30, 2022).

[57] A. Minokuchi, S. Isobe, H. Takahashi, and S. Nagata, "5G standardization trends at 3GPP," *NTT DOCOMO Technical Journal*, vol. 19, no. 3, pp. 5-12, Jan. 2018.

[58] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639-127651, Sep. 2019.

[59] A. Umesh, W. A. Hapsari, T. Uchino, T. Toeda, and H. Takahashi, "5G radio access network standardization trends," *NTT DOCOMO Technical Journal*, vol. 19, no. 3, pp. 36-47, Jan. 2018.

[60] O-RAN Alliance, "O-RAN-WG1-O-RAN Architecture Description — v06.00.00," Tech. Spec., unpublished.

[61] O-RAN Alliance, "O-RAN Operations and Maintenance Architecture (O-RAN.WG1.OAM-Architecture-v04.00)," Tech. Spec., Nov. 2020.

[62] O-RAN Alliance, "O-RAN Working Group 3, Near-Real-time RAN Intelligent Controller, Near-RT RIC Architecture (O-RAN.WG3.RICARCH-v02.00)," Tech. Spec., Mar. 2021.

[63] O-RAN Alliance, "O-RAN Working Group 8, Base Station O-DU and O-CU Software Architecture and APIs (O-RAN.WG8.AAD.0-v05.00)," Tech. Spec., unpublished.

[64] O-RAN Alliance, "O-RAN Fronthaul Working Group, Control, User and Synchronization Plane Specification (O-RAN.WG4.CUS.0-v07.00)," Tech. Spec., Jul. 2021.

[65] A. Umesh, T. Yajima, T. Uchino, and S. Okuyama, "Overview of O-RAN fronthaul specifications," *NTT DOCOMO Technical Journal*, vol. 21, no. 1, pp. 46-59. July 2019.

[66] L. M. P. Larsen, A. Checko and H. L. Christiansen, "A survey of the functional splits proposed for 5G mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146-172, Firstquarter 2019.

[67] M. Mohsin, J. M. Batalla, E. Pallis, G. Mastorakis, E. K. Markakis, and C. X. Mavromoustakis, "On analyzing beamforming implementation in O-RAN 5G," *Electronics*, vol. 10, no. 2162, pp. 1-17, Sep. 2021.

[68] C. Li *et al.*, "O-RAN use cases and deployment scenarios," O-RAN Alliance, White Paper, 2020. [Online].

[69] H. Kumar, V. Sapru, and S. K. Jaisawal, "O-RAN based proactive ANR optimization," *2020 IEEE Globecom Workshops*, pp. 1 - 4, Mar. 2021.

[70] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network Automation through Orchestrated Intelligence in the Open RAN," 2022, arXiv:2201.05632.

[71] S. Niknam *et al.*, "Intelligent O-RAN for beyond 5G and 6G wireless networks," 2020, arxiv:2005.08374.

[72] H. Lee, J. Cha, D. Kwon, M. Jeong and I. Park, "Hosting AI/ML workflows on O-RAN RIC platform," *2020 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2020.

[73] R. Schmidt, M. Irazabal and N. Nikaein, "FlexRIC: An SDK for next-generation SD-RANs," *2021 17th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '21)*, Dec. 2021.

[74] A. Huff, M. Hiltunen, and E. P. Duarte, "RFT: Scalable and fault-tolerant microservices for the O-RAN control plane," *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, June 2021.

[75] D. Johnson, D. Maas, and J. Van Der Merwe, "NexRAN: Closed-loop RAN slicing in POWDER - A top-to-bottom open-source open-RAN use case," *2022 15th ACM Workshop on Wireless Network Testbeds Experimental evaluation & CHaracterization (WiNTECH '21)*, Jan.-Feb. 2022.

[76] F. Mungari, "An RL approach for radio resource management in the O-RAN architecture," *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, July 2021.
[77] Xilinx. (2020). Xilinx 5G Telco Accelerator Cards. [PowerPoint slides].
[78] Light Reading. (2021). Accelerating Open RAN Platforms Operator Survey. [Online].
[79] "The total economic impact Of Arm Neoverse," Forrester, 2020. [Online].
[80] X. Wang, J. D. Thomas, R. J. Piechocki, S. Kapoor, R. Santos-Rodriguez, and A. Parekh, "Self-Play learning strategies for resource assignment in Open-RAN networks," 2021, arxiv:2103.02649.
[81] J. S. Lee, J. Park, J. Choi, and M. S. Lee, "Design of a management plane for 5G open fronthaul interface," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2016.
[82] O-RAN Alliance, "O-RAN Alliance Working Group 4, Management Plane Specification (O-RAN.WG4.MP.0-v07.00)," Tech. Spec., Jul. 2021.
[83] O-RAN Alliance, "O-RAN Alliance Working Group 4, O-RAN Management Plane Specification - YANG Models (O-RAN.WG4.MP-YANGs-v07.00)," Tech. Spec., Jul. 2021.
[84] P. Ringqvist, "5G challenge notice of inquiry: Ericsson response," Ericsson, 2021. [Online].
[85] J. Y. Cho and A. Sergeev, "Secure open fronthaul interface for 5G networks," *2021 16th International Conference on Availability, Reliability and Security (ARES)*, Aug. 2021.
[86] E. Sarikaya and E. Onur, "Placement of 5G RAN slices in multi-tier O-RAN 5G networks with flexible functional splits," *2021 17th International Conference on Network and Service Management (CNSM)*, Oct. 2021.
[87] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: Conventional and new approaches," *Sensors*, vol. 20, no. 9, p. 2708, May 2020.
[88] I. Tamim, A. Saci, M. Jammal and A. Shami, "Downtime-aware O-RAN VNF deployment strategy for optimized self-healing in the O-Cloud," *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021.
[89] C. Gabriel, "Migration choices and dilemmas in vRAN and Open RAN," Rethink Technology Research, 2021. [Online].
[90] "Open and virtualised radio access networks: An explanatory guide for policymakers," GSMA, 2021. [Online].
[91] R. Curran, "Who's winning in open RAN? " Appledore Research, 2021. [Online].
[92] 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," 2021, arxiv:2112.09559.
[93] S. Ravindran, S. Chaudhuri, J. Bapat, and D. Das, "Predictive dynamic scaling multi-slice-in-slice-connected users for 5G system resource scheduling," 2021, arxiv:2106.15229.
[94] J. F. Santos *et al.*, "Breaking Down network slicing: Hierarchical orchestration of end-to-end networks," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 16-22, Oct. 2020.
[95] L. Bertizzolo *et al.*, "Streaming from the air: Enabling drone-sourced video streaming applications on 5G Open-RAN architectures," *IEEE Transactions on Mobile Computing*, Nov. 2021
[96] A. Mughees, M. Tahir, M. A. Sheikh, and A. Ahad, "Towards energy efficient 5G networks using machine learning: Taxonomy, research challenges, and future research directions," *IEEE Access*, vol. 8, pp. 187498 - 187522, Oct. 2020.
[97] T. Pamuklu, M. Erol-Kantarci, and C. Ersoy, "Reinforcement Learning Based Dynamic Function Splitting in Disaggregated Green Open RANs," *ICC 2021 - IEEE International Conference on Communications*, June 2021.
[98] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-Computing-Based radio access networks: Issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46-53, July 2016.
[99] N. Cheng, W. Xu, W. Shi, Y. Zhou, N. Lu, H. Zhou, and X. Shen, "Air-Ground integrated mobile edge networks: Architecture, challenges, and opportunities," *IEEE Communications Magazine*, V=vol. 56, no. 8, pp. 26-32, Aug. 2018.
[100] F. Z. Morais, G. M. F. de Almeida, L. Pinto, K. V. Cardoso, L. M. Contreras, R. R. Righi, and C. B. Both, "PlaceRAN: Optimal placement of virtualized network functions in the next-generation radio access networks," 2021, arxiv:2102.13192.
[101] S. P. Jason and S. Boswell, "Security considerations of Cloud RAN," Ericsson, White Paper, 2021. [Online].
[102] C. Suraci, G. Araniti, A. Abrardo, G. Bianchi, and A. Iera, "A stakeholder-oriented security analysis in virtualized 5G cellular networks," *Computer Networks*, vol. 184, p. 107604, Jan. 2021.
[103] O-RAN Alliance, "O-RAN ALLIANCE Security Focus Group, O-RAN Security Threat Modeling and Remediation Analysis 2.0," Tech. Spec., July 2021.
[104] S. Sullivan, A. Brighente, S. Kumar, and M. Conti, "5G security challenges and solutions: A review by OSI layers," *IEEE Access*, vol. 9, pp. 116294-116314, Aug. 2021.
[105] T. F. Rahman, A. S. Abdalla, K. Powell, W. AlQwider, and V. Marojevic, "Network and physical layer attacks and countermeasures to AI-Enabled 6G O-RAN," 2021, arxiv:2106.02494.
[106] C. Cowan, "Software security for open-source systems," *IEEE Security & Privacy*, vol. 1, no. 1, pp. 38-45, Jan.-Feb. 2003.
[107] O-RAN Alliance, "O-RAN ALLIANCE Security Focus Group, O-RAN Security Protocols Specifications 2.0," Technical Specification, July 2021.
[108] O-RAN Alliance, "O-RAN ALLIANCE Security Focus Group, O-RAN Security Requirements Specifications 1.0," Technical Specification, July 2021.
[109] S. Zhang, Y. Wang, and W. Zhou, "Towards secure 5G networks: A survey," *Computer Networks*, vol. 162, p. 106871, Oct. 2019.
[110] A. Mathew, "Network slicing in 5G and the security concerns," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Apr. 2020.
[111] D. Mimran *et al.*, "Evaluating the security of open radio access networks," 2022, arxiv:2201.06080.
[112] O-RAN Alliance, "O-RAN ALLIANCE Test and Integration Focus Group, O-RAN End-to-End Test Specification 2.0," Tech. Spec., Jul. 2021.
[113] S. Khan, S. Parkinson, and Y. Qin, "Fog computing security: a review of current applications and security solutions," *Cloud Computing: Advances, Systems and Applications*, vol. 6, no. 1, p. 90, Dec. 2017.
[114] "Security in Open RAN," Altiostar, White Paper, 2021. [Online].
[115] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5G/6G tactical networks: Principles, challenges, and the role of machine learning," 2022, arxiv:2105.01478.
[116] X. Ling, J. Wang, T. Bouchoucha, B. C. Levy, and Z. Ding, "Blockchain Radio Access Network (B-RAN): Towards decentralized secure radio access paradigm," *IEEE Access*, vol. 7, pp. 9714-9723, Jan. 2019.
[117] L. Giupponi and F. Wilhelmi, "Blockchain-Enabled network sharing for O-RAN," 2021, arxiv:2107.02005.
[118] "Open Mobile Evolved Core." opennetworking.org. https://opennetworking.org/omec/ (accessed Aug. 01, 2022).
[119] "Home." magmacore.org. https://magmacore.org/ (accessed Aug. 01, 2022).
[120] "Home." free5gc.org. https://www.free5gc.org/ (accessed Aug. 01, 2022).
[121] "Home." open5gs.org. https://open5gs.org/ (accessed Aug. 01, 2022).
[122] "Open RAN (radio access network) initiative industry connections activity initiation document (ICAID)," IEEE Standards Association, 2021. [Online].
[123] N. Kazemifard and V. Shah-Mansouri, "Minimum delay function placement and resource allocation for Open RAN (O-RAN) 5G Networks," *Computer Networks*, vol. 188, p. 107809, Apr. 2021.
[124] P. E. I. Rivera, S. Mollahasani, and M. Erol-Kantarci, "Multi-Agent team learning in virtualized open radio access networks (O-RAN)," 2021, arXiv:2012.04861.
[125] "OAI MOSAIC5G (M5G) PROJECT GROUP – CHARTER," OpenAirInterface, 2021. [Online].
[126] S. Pongratz, "The market upside with open RAN," Dell'Oro Group, White Paper, 2021. [Online].
[127] S. Pongratz, "Advanced research reports on open RAN," Dell'Oro Group, 2021. [Online].
[128] J. C. Plantin, "The geopolitical hijacking of open networking: the case of Open RAN," *European Journal of Communication*, vol. 36, pp. 404-417, July 2020.
[129] 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 Communications*, vol. 27, no. 1, pp. 16-23, Feb. 2020.

[130] F. Rinaldi, H. L. Maattanen, J. Torsner, S. Pizzi, S. Andreev, A. Iera, Y. Koucheryavy, and G. Araniti, "Non-Terrestrial networks in 5G & beyond: A survey," *IEEE Access*, vol. 8, pp. 165178-165200, Sep. 2020.
[131] W. Liu, J. Ding, J. Zheng, X. Chen, and C. L. I, "Relay-Assisted technology in optical wireless communications: A survey," *IEEE Access*, vol. 8, pp. 194384-194409, Oct. 2020.
[132] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 197-216, Firstquarter 2019.
[133] 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," *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2022.
[134] "About." bubbleran.com. https://bubbleran.com/about/ (accessed Jul. 25, 2022).