Received 24 July 2024; accepted 14 August 2024. Date of publication 22 August 2024; date of current version 10 September 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3447472

Colosseum: The Open RAN Digital Twin

MICHELE POLESE (Member, IEEE), LEONARDO BONATI (Member, IEEE), SALVATORE D'ORO (Member, IEEE), PEDRAM JOHARI (Member, IEEE), DAVIDE VILLA (Graduate Student Member, IEEE), SAKTHIVEL VELUMANI, RAJEEV GANGULA, MARIA TSAMPAZI (Graduate Student Member, IEEE), CLIFTON PAUL ROBINSON (Graduate Student Member, IEEE), GABRIELE GEMMI (Member, IEEE), ANDREA LACAVA (Graduate Student Member, IEEE), STEFANO MAXENTI (Graduate Student Member, IEEE), HAI CHENG, AND TOMMASO MELODIA (Fellow, IEEE)

Institute for the Wireless Internet of Things, Northeastern University, Boston, MA 02115, USA

CORRESPONDING AUTHOR: M. POLESE (e-mail: m.polese@northeastern.edu)

This work was supported in part by the O-RAN ALLIANCE; in part by the U.S. National Science Foundation under Grant CNS-1925601 and Grant CNS-2112471; in part by the National Telecommunications and Information Administration (NTIA)'s Public Wireless Supply Chain Innovation Fund (PWSCIF) under Award 25-60-IF011 and Award 25-60-IF054; and in part by the OUSD(R&E) through Army Research Laboratory Cooperative Agreement under Grant W911NF-19-2-0221. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ABSTRACT Recent years have witnessed the Open Radio Access Network (RAN) paradigm transforming the fundamental ways cellular systems are deployed, managed, and optimized. This shift is led by concepts such as openness, softwarization, programmability, interoperability, and intelligence of the network, which have emerged in wired networks through Software-defined Networking (SDN) but lag behind in cellular systems. The realization of the Open RAN vision into practical architectures, intelligent data-driven control loops, and efficient software implementations, however, is a multifaceted challenge, which requires (i) datasets to train Artificial Intelligence (AI) and Machine Learning (ML) models; (ii) facilities to test models without disrupting production networks; (iii) continuous and automated validation of the RAN software; and (iv) significant testing and integration efforts. This paper is a tutorial on how Colosseum—the world's largest wireless network emulator with hardware in the loop—can provide the research infrastructure and tools to fill the gap between the Open RAN vision, and the deployment and commercialization of open and programmable networks. We describe how Colosseum implements an Open RAN digital twin through a high-fidelity Radio Frequency (RF) channel emulator and end-to-end softwarized O-RAN and 5G-compliant protocol stacks, thus allowing users to reproduce and experiment upon topologies representative of real-world cellular deployments. Then, we detail the twinning infrastructure of Colosseum, as well as the automation pipelines for RF and protocol stack twinning. Finally, we showcase a broad range of Open RAN use cases implemented on Colosseum, including the real-time connection between the digital twin and real-world networks, and the development, prototyping, and testing of AI/ML solutions for Open RAN.

INDEX TERMS O-RAN, Open RAN, wireless network emulation, 5G, 6G.

I. INTRODUCTION

WIRELESS networks and systems are evolving into more programmable, open, and virtualized solutions. The transition to software-defined networking [1], which has taken over the management operations of the wired networking industry in the last decade, is now moving to cellular networks. While introducing software and programmability in radio systems is more challenging compared to switching and routing, it also opens opportunities to deploy and configure bespoke networks that can be dynamically tuned to optimally support a variety of different verticals. The Open RAN paradigm, its embodiment in the technical specifications of the O-RAN ALLIANCE, and evolution in the 3GPP RAN and core network designs are converging toward a network architecture which is open, disaggregated, programmable, and intelligent [2], [3]. Openness, embedded in interfaces that expose RAN telemetry and control, enables the programmability and optimization of RAN functions through intelligent closed-loop control driven from the so-called RAN Intelligent Controllers (RICs) [4], [5]. Disaggregation, implemented by splitting RAN functionalities into self-contained units across user and control planes, fosters innovation and introduces multi-vendor interoperability. Finally, programmability, realized through open Application Programming Interfaces (APIs) of the software solutions, enables swift reconfiguration of the white-box components of the cellular networks to adapt to ever-changing network conditions and demand. Overall, these ingredients have the potential to transform how we deploy, manage, and optimize cellular networks.

However, advancing the Open RAN vision toward a fully developed architecture with robust and reliable algorithmic components and seamless multi-vendor integration requires addressing several challenges at the architectural, algorithmic, and system-level design, which include the following.

Need for Datasets. To develop robust and scalable AI and ML solutions, which generalize well across a variety of real-world deployment scenarios, it is necessary to leverage rich datasets of RAN telemetry, data, and performance indicators [6]. While network operators are in a unique position to collect such datasets, it is often impractical or impossible to use them for research and development due to privacy and security concerns. Wireless testbeds represent a feasible path to overcome this limitation [7], [8], [9], [10], [11]. However, they are often limited to the RF characteristics and topology of their deployment area.

End-to-end AI and ML Testing. Once trained, AI/ML-based control solutions need to be validated and tested in controlled environments to avoid disruption in production networks. At the same time, the testing conditions need to be realistic to obtain meaningful results that consider, for instance, the user load, traffic patterns, and RF characteristics of real-world deployments the models will be used to control.

Continuous Software Validation. While softwarization introduces flexibility and programmability of the stack, it also comes with concerns around software quality, reliability, security, and performance [12]. Therefore, integrating, validating, testing, and profiling software for wireless in a continuous fashion is key to the Open RAN vision. Additionally, this validation needs to consider various compute platforms and hardware acceleration solutions for physical layer processing.

Automated Integration and Testing of Disaggregated Components. Disaggregation comes with a more robust supply chain, but also a need for the validation of interoperability across vendors and devices. This is a labor-intensive and often manual process that calls for the development of automated techniques [13].

In this paper, we describe how Colosseum—the world's largest wireless network emulator with hardware in the loop [14]—can be leveraged as a digital twin platform to address these challenges and to develop end-to-end, fully integrated, and reliable solutions for Open RAN, as shown in Figure 1. Through its channel and traffic emulation capabilities, Colosseum can replicate countless real-world scenarios representative of real-world cellular deployments, and generate datasets that can be used to train AI/ML models robust to network changes [15]. A combination of generic compute nodes, Software-defined Radios (SDRs), and the possibility of integrating Commercial Off-the-Shelf (COTS) devices, allows for the digital replica of Open RAN and 5G-and-beyond protocol stacks [16], [17], which we manage through automation and Continuous Integration (CI)/Continuous Deployment (CD) pipelines [18]. This enables repeatable experiments, where different network configurations and protocol stacks can be tested against the same channel and traffic conditions, as well as a safe playground for testing of AI/ML solutions. The CI/CD and automation also enable continuous validation, as the same software can be tested over time for regression in a realistic environment, and against various other stacks for integration.

Compared to other papers on digital twinning [19], [20], [21], [22], [23], [24], [25], [26], this work discusses a hardware-in-the-loop, end-to-end prototype of a digital twin, its twinning process, and its connection and integration with real-world environments. With respect to prior work on Open RAN and Colosseum [7], [15], [27], [28], [29], this paper is the first providing a comprehensive view, with a tutorial style, on how Colosseum's capabilities can be leveraged to twin a full-stack Open RAN deployment, as shown in Fig. 1.

The remaining of this paper—which is meant as a tutorial that introduces Colosseum as an Open RAN digital twin—is organized as follows. Section II reviews the state of the art on digital twins and experimental frameworks for Open RAN. Section III introduces Colosseum, its infrastructure, and capabilities, including recent extension that introduced Graphics Processing Unit (GPU) servers for AI and ML training. Section IV reviews the open-source protocol stacks that can be twinned in Colosseum to replicate 5G and Open RAN systems. These include OpenAirInterface (OAI) [16], srsRAN [17], and the OpenRAN Gym [27] framework. Section V reviews methods and techniques to replicate real-world RF scenarios with high fidelity [7], [28], and discusses the planned evolution of the Colosseum channel emulator toward a dynamic RF twinning. Section VI details the fully automated framework that we developed and that leverages CI/CD techniques to continuously update and test the 5G and Open RAN protocol stacks in Colosseum. This section also presents a novel software broker to enable twinning of scenarios across the real and digital domains. Section VII discusses how Colosseum experiments can

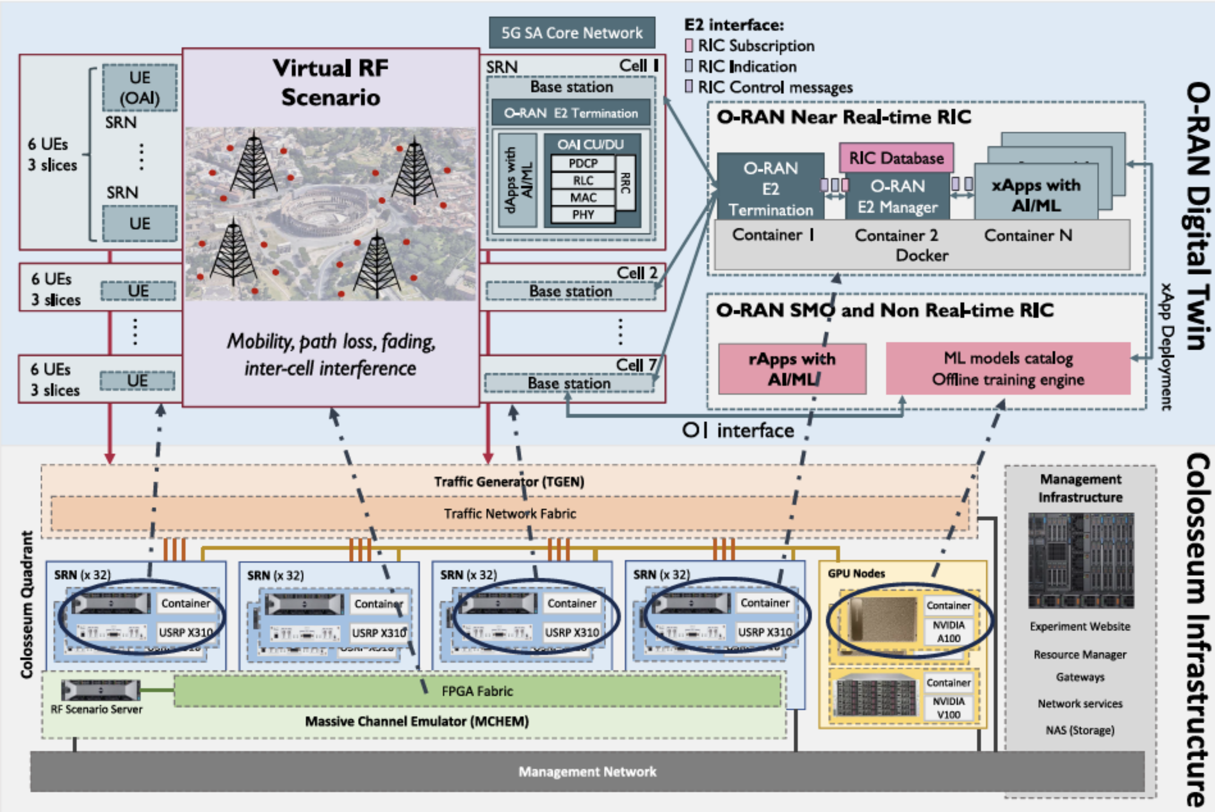

FIGURE 1. Open RAN twinning capabilities in Colosseum.

extend to external platforms, including over-the-air-testbeds at Northeastern University [30], [31] and the Platforms for Advanced Wireless Research (PAWR) platforms [9], [32]. Section VIII reviews Open RAN applications and use cases that have been demonstrated on Colosseum. These include intelligent network slicing [15], [33], real-time spectrum sharing enabled by Open RAN interfaces and controllers [29], [34], explainable AI for O-RAN [35], Integrated Access and Backhaul (IAB) optimization [36], and jamming and fingerprinting with O-RAN primitives. This is to provide the researchers in the wireless space with an idea of what research applications can be developed on top of Colosseum. Section IX instead provides a brief summary of the steps required to run an experiment on Colosseum. Section X discusses the role of the Colosseum in testing and integration activities for interoperable Open RAN systems. Finally, Section XI concludes our paper. Overall, we believe that this tutorial paper can serve the research community as a comprehensive resource to approach, understand, and leverage the resources that Colosseum makes available in the context of Open RAN systems.

II. RELATED WORK

The development of digital twins and testbeds for cellular systems and, specifically, for Open RAN is a topic that has recently received significant attention, owing to the importance of addressing the challenges described in Section I. The O-RAN ALLIANCE itself has dedicated a study area to digital twins for Open RAN within its next Generation Research Group (nGRG).

Digital twin frameworks for cellular networks are discussed in [19], [20], [21], [22], [23], [24], [25], [26], with an analysis of the challenges and opportunities of using digital twins for Open RAN and beyond-5G design. Paper [37] proposes a network digital twin framework for 5G and develops a use case on RAN control with reinforcement learning, with a simulation-based twin that replicates a real-world scenario in Barcelona, Spain. The authors of [38] discuss the opportunity of integrating digital twins in automated operations for the development of reinforcement learning algorithms for Open RAN. Reference [39] performs a simulation study and frames it in the context of digital twins to optimize fixed wireless access use cases for cellular networks. The studies in [40], [41] leverage a digital twin to create a virtual environment for the testing of network updates driven by AI/ML components. These papers, however, are primarily based on simulation-driven twins, which fails to capture the complexity of real-world RF equipment. In addition, there is a limited evaluation of the digital twin in the real world, e.g., through experiments in real networks or testbeds. The Colosseum digital twin described in this paper is based on real-time channel emulation and end-to-end Open RAN protocol stacks that can be either automatically deployed and tested in the real world or in the digital twin as-is, thus achieving a high level of fidelity as testified by the experiments in [7], [27].

Implementations of digital twins or components of digital twins for networked systems are presented in [42], [43], [44], focusing on middleware for digital twins for aerial networks and on optimizing network service for the replica of a real-world setup (e.g., for Industry 4.0) into a digital twin. These papers, however, do not focus on end-to-end Open RAN systems and on the development of intelligent control solutions. Testbeds and frameworks for the evaluation of Open RAN systems are discussed in [30], [45], [46], [47], [48], and Open RAN evaluation is also supported in the PAWR platforms [8], [9], [10], [11]. The PAWR platforms are the closest in terms of capabilities and programmability to Colosseum, being mostly based on SDRs, but lack the emulated environment that Colosseum supports. Nonetheless, these systems can be (and have been) used in combination with the twinning infrastructure of Colosseum to replicate Open RAN studies across twinned and real-world domains.

III. COLOSSEUM WIRELESS NETWORK EMULATOR

The architecture of Colosseum is shown in the bottom part of Figure 1. Its main components are: (i) 128 pairs of generic compute servers and SDRs, named Standard Radio Nodes (SRNs); (ii) a channel emulation system; and (iii) a state-of-the-art AI/ML infrastructure.

Each SDR is paired with a Dell server connected to an NI/Ettus Universal Software Radio Peripheral (USRP) X310 SDR through a dedicated 10 Gbps link.[1] The users of Colosseum can reserve these SRNs and deploy either custom or pre-configured protocol stacks through Linux Container (LXC). Container images are stored in a dedicated storage infrastructure, which is part of a larger management infrastructure that includes services such as websites, Virtual Private Network (VPN) for external connectivity, and networking functionalities.

The Colosseum channel emulator, namely Massive Channel Emulator (MCHEM), is in charge of reproducing digital representations of real-world RF propagation scenarios, which represent the state of the RF channel at given time instants. Specifically, each pair of transceivers, i.e., those of the SDRs, is modeled through a Tapped Delay Line (TDL) with up to four non-zero taps that represent the multipath components of the Channel Impulse Response (CIR) for each millisecond of the captured scenario. At its core, MCHEM includes four quadrants with 64 Field Programmable Gate Arrays (FPGAs) and 128 SDRs that up-/down-convert the signal between RF and baseband. The FPGAs implement real-time convolutional operations to the digital baseband signal, modeling RF taps of the wireless channel effects of the real-world environments. These taps, which have a maximum propagation delay of 5.12 μs [49], are streamed to the FPGAs from a dedicated RF scenario server. We refer readers to [14] for a comprehensive overview of the Colosseum architecture.

The AI/ML infrastructure of Colosseum, shown in yellow at the bottom of Figure 1, consists of two NVIDIA DGX A100 stations with 8 GPUs each, providing 10 petaFLOPS of compute power and one Supermicro Superserver 8049U-E1CR4T with 6 NVIDIA V100 GPUs and 3 TB of RAM. Compared to the virtualization system of the SRNs, the GPU nodes leverage Docker containers. Resources on these nodes are managed through HashiCorp frameworks [50], such as Nomad (for workload orchestration), Consul (for service discovery, configuration, and connectivity), and Traefik (for reverse proxy and load balancing), deployed in a redundant manner for fault tolerance.

[1]We are in the process of updating our compute solutions within Colosseum, transitioning from Dell PowerEdge R730 servers to Dell PowerEdge R750.

IV. END-TO-END OPEN RAN TWINNING IN COLOSSEUM

The emulation capabilities of Colosseum can be leveraged to deploy, study, and profile wireless protocol stacks in controlled and repeatable environments. As an example, Figure 1 showcases the deployment of an end-to-end, fully programmable Open RAN cellular system in Colosseum. The top-left portion of the figure depicts the RAN, which consists of core network and cellular base stations—either monolithic or disaggregated in Central Units (CUs) and Distributed Units (DUs)—deployed on an emulated RF deployment of interest and serving mobile User Equipments (UEs). The top-right portion of the figure, instead, shows the Service Management and Orchestration (SMO) and O-RAN RICs that interface with the RAN base stations through open interfaces and augment their operations through AI/ML applications deployed therein, namely xApps and rApps. We refer the readers to [2] for a detailed overview of the O-RAN architecture and functionalities.

In Colosseum, these O-RAN-aligned solutions are provided to the users through ready-to-use LXC images that implement the functionalities made available through the OpenRAN Gym framework, which will be described next.

A. OPEN-SOURCE CELLULAR PROTOCOL STACKS

Colosseum enables experimentation with different open-source softwarized protocol stacks, such as OAI [16] for 5G RANs and srsRAN [17] for 4G ones. These protocol stacks—which are provided as ready-to-use images [51]—are the same that can be used for over-the-air experiments with minor variations in the configuration parameters, discussed next.

OpenAirInterface. OAI is an open-source project that provides a reference implementation of 3GPP cellular protocol stacks, including that of 5G NR. The OAI project adopts a modified version of the Apache 2 license which grants royalty-free patent privileges solely for study, testing and research purposes, while allowing the patent holder organization that contributed to specific functionalities to charge royalties for other uses according to 3GPP industry-standard Fair, Reasonable and Non-Discriminatory (FRAND) terms. The OAI-based NR system consists of 5G RAN and core network that run on Linux-based general purpose compute platforms and can control SDRs, e.g., the NI/Ettus USRPs deployed in Colosseum. The OAI RAN applications support the instantiation of Next Generation Node Bases (gNBs) and UEs. The gNB can operate in monolithic or disaggregated manner that leverages the CU/DU split option. Both modes are available in Colosseum. With the disaggregated base station, CU and DU exchange data with control and user plane via the F1-C and F1-U interfaces, respectively, through the F1AP protocol. The CU can further be disaggregated into Central Unit Control Plane (CU-CP) and Central Unit User Plane (CU-UP). The control information between control plane and user plane nodes follows the E1AP protocol.

Users of Colosseum can deploy 5G NR cellular networks on Colosseum through ready-to-use container images. Compared to over-the-air environments, however, the extra components in Colosseum transmit and receive chains (e.g., those part of MCHEM) require adjusting the timing advance during the establishment of the initial connection between gNB and UEs. This can be done through the `ta` input argument when starting the base station application.

srsRAN. For legacy 4G networks, Colosseum provides container images for the 4G version of srsRAN. Similarly to OAI, this open-source software allows users to instantiate experiments with softwarized evolved Node Bases (eNBs), UEs, and core network. 4G systems are primarily based on Frequency Division Duplexing (FDD), with the downlink and uplink in two separate frequency bands. In Colosseum, these need to be accommodated in two portions of the 80 MHz usable bandwidth through a custom configuration of the srsLTE eNB, which is already pre-configured in the LXC images available to Colosseum users.

B. OPENRAN GYM

OpenRAN Gym is an open-source framework for O-RAN experimentation that lets users instantiate softwarized RANs based on the OAI and srsRAN protocol stacks, and control them through xApps deployed on an O-RAN Near-Real-time (RT) RIC [27]. Its components, shown in Figure 2, include (i) SCOPE [52], a wrapper for RAN software that simplifies the experiment workflow of users and enables data-collection at scale, and (ii) ColO-RAN [15], a simplified version of the O-RAN Software Community Near-RT RIC that provides an Software Development Kit (SDK) for xApp design, as well as pipelines for data collection and training of AI/ML models for RAN inference and control to be embedded into xApps. Overall, OpenRAN Gym lets users instantiate a fully compliant Near-RT RIC using Docker containers, as well as xApps that can interact with the RAN (e.g., to receive Key Performance Indicators (KPIs) and enforce control policies) via the E2 interface. Examples of developed xApps include Deep Reinforcement Learning (DRL) agents to optimize the performance of the RAN in real-time by controlling the scheduling and slicing configuration [15], [33], [53]. Tutorials on how to start using OpenRAN Gym on Colosseum are publicly available on the OpenRAN Gym website.[2]

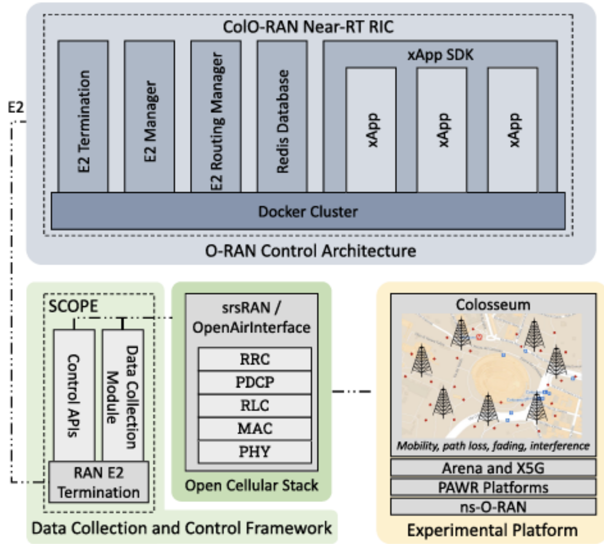

FIGURE 2. Intelligent Control Loops with Colosseum, based on the OpenRAN Gym framework [27].

Finally, this framework has also been used to demonstrate the portability of experiments from Colosseum to external platforms, e.g., the Arena testbed [31] and the platforms from the PAWR program [54] funded by the U.S. National Science Foundation, as discussed in Section VII, and to interface with RANs simulated through the ns-3 discrete-event network simulator using ns-O-RAN [55]. This highlights the extensibility of the OpenRAN Gym framework and the flexibility of the digital twinning capabilities of Colosseum.

V. RADIO-FREQUENCY TWINNING

The channel and traffic emulation and the cellular protocol stacks that can be deployed on Colosseum provide key components toward the creation of real-time, high-fidelity digital replicas of Open RAN systems with RF hardware in the loop. This section describes a set of toolchains that we designed and built to twin real-world RF scenarios in Colosseum. Section V-A describes our Channel Emulation Scenario Generator and Sounder Toolchain (CaST), which enables RF scenario twinning by pre-generating offline scenarios, while Section V-B discusses our roadmap toward real-time RF twinning on Colosseum.

[2]OpenRAN Gym website: https://openrangym.com.

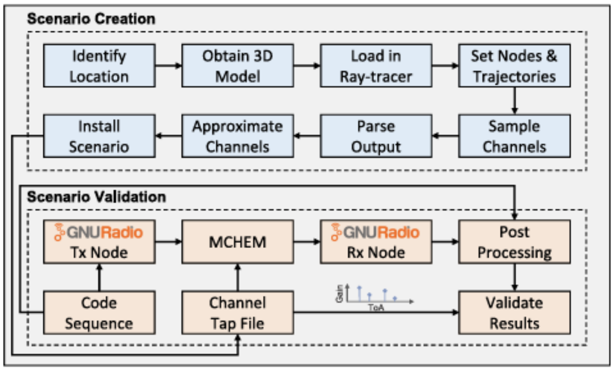

FIGURE 3. CaST scenario creation (top) and scenario validation (bottom) workflows.

A. RF TWINNING WITH CAST

Because of the computational and time complexity of running ray-tracing operations to derive the taps of the wireless channels to emulate, RF scenarios are currently generated offline, and then replayed in real-time by the scenario server. To do so, we leverage the RF twinning procedure of the CaST framework [28]. This toolchain, which is publicly available for the research community,[3] enables users to characterize real-world RF environments and to turn them into digital twin representations to be used in channel emulators such as Colosseum. CaST workflow and procedures, depicted in Figure 3, comprise two main components: (i) a streamlined framework to create realistic mobile wireless scenarios (top part of the figure); and (ii) an SDR-based channel sounder for characterizing emulated RF channels (bottom part). The combination of these two components allows for the generation of digital twin RF scenarios, and for their validation, ensuring that the digital replicas of the wireless environments align with their real-world counterpart.

The main steps of CaST scenario creation workflow are as follows: (i) identify the wireless environment to capture, i.e., the physical location to twin, which can vary in size and type, e.g., indoor, outdoor, urban, rural [7], [29]; (ii) obtain the 3D model of the environment by leveraging online databases such as Open Street Map (OSM) or modeling software such as SketchUp; (iii) load the model in a ray-tracer, e.g., MATLAB ray-tracer, Wireless InSite (WI), Opal, or Sionna; (iv) set the location and trajectories of nodes; (v) sample the channel with the ray-tracer, capturing effects such as mobility of the nodes; (vi) parse the ray-tracer output to extract the channels for each pair of nodes for each millisecond of the emulated scenario; (vii) approximate the resulting channel information in a format suitable for the Colosseum channel emulator, i.e., 512 channel taps with at most 4 non-zero, spaced at 10 ns intervals and a maximum delay spread of 5.12 μs; (viii) install the scenario in Colosseum, converting it into an FPGA-based format. This scenario creation toolchain is modular, and it allows users to provide their inputs at any of the above-mentioned steps.

After a scenario is installed in Colosseum, it undergoes validation through the second part of the CaST framework to ensure that it closely follows the expected real-world behavior. This involves a channel sounder implemented through the GNU Radio software [56], and consists of the following steps: (i) transmission of a known code sequence used as a reference signal via the GNU Radio transmitter; (ii) reception via the GNU Radio receiver of the signal processed by MCHEM, on top of which the channel characteristics were applied; (iii) post-processing of the received data with known code sequence with added channel effects to extract CIR and Path Loss (PL) of the waveform; (iv) validation of the results with the original modeled channel taps. As demonstrated in [28], the CaST sounder is able to achieve an accuracy of 20 ns for CIR delays and 0.5 dB for tap gains. More details on the sounding sequences and processing algorithms can be found in [28].

Finally, since Colosseum SDRs are equipped with two transmit-receive chains, Colosseum and MCHEM support 2-by-2 Multiple Input, Multiple Output (MIMO) operations.[4] Thus, CaST can be used to generate simplified MIMO channels leveraging MATLAB or Sionna's NR 3rd Generation Partnership Project (3GPP) channel models. A comprehensive list of created and validated scenarios available to Colosseum users can be found in [57].

B. ROADMAP TOWARD REAL-TIME TWINNING

At the time of this writing, Colosseum allows users to replay high-fidelity RF scenarios that have been previously modeled offline and installed on the system. While this process is extremely flexible and allows modeling RF scenarios with different kind of tools (e.g., ray-tracers, mathematical models, etc.), it is not suitable for those cases in which the channel between nodes need to dynamically change in real-time. This is the case, for instance, in which the digital twin needs to adapt to the state of the physical environment.

To overcome this limitation, we are updating the infrastructure of Colosseum to enable real-time RF twinning. Our phased approach combines an update in the RF scenario server with a more flexible MCHEM implementation, to stream channel taps generated in real-time, with real-time ray-tracing solutions. While the feature is currently under development and not generally available, we have demonstrated dynamic real-time twinning capabilities in [58], primarily around the possibility of updating the channel representation in real-time. Specifically, here we use input from a vehicular mobility simulator, SUMO, to represent the position of the nodes. The channel is then updated based on a path loss model and the real-time position of the vehicles, with a 100 Hz refresh rate.

[3] CaST is available at https://github.com/wineslab/cast.

[4] Larger MIMO channels are possible, in principle, by connecting multiple SDRs to the same compute server.

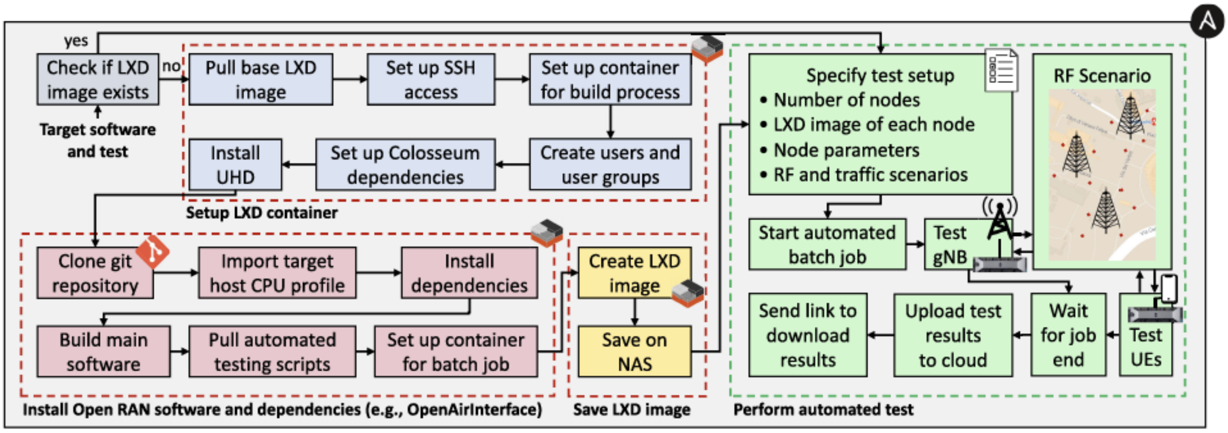

FIGURE 4. Ansible pipeline to perform an automated test on Colosseum.

In the following we discuss possible approaches we are considering toward enabling real-time twinning as a general feature available to all Colosseum users, as well as some challenges.

Real-time Channel Modeling. Using this approach, a detailed model of the RAN can be recreated in the digital twin using scenario and channel modeling techniques, e.g., 3D models and ray-tracing. This model includes static and dynamic components. The former is pre-configured in the system, and includes, for instance, the location of the base stations and buildings. The latter is transmitted in real-time from the real-world RAN to reflect changes such as location of the UEs, the transmission power. Finally, once the digital component has been reconstructed, techniques such as ray-tracing or stochastic propagation models can be used to characterize the RF propagation environment, then used to update MCHEM in real-time. The challenge here is associated to generating the channel representation with a high refresh rate, while maintaining high realism and taking into account high-order channel phenomena such as diffraction and diffusion.

Real-time Channel Sensing. The previous approach can be extended with channel sensing techniques (e.g., the CaST sounding component) that continuously estimate channels across devices or test locations in the real-world environment. This channel estimation can be continuously sent from the RAN to the digital twin, and then converted into an input that can be processed by MCHEM. Additionally, techniques could be used to derive an estimation of the channel taps from the Channel State Information Reference Signal (CSI-RS) or uplink Sounding Reference Signal (SRS) of the 5G NR physical layer [59]. The challenge here is how to manage and orchestrate the estimation process so that the twinned system receives fresh, updated, and relevant channel representations when required.

VI. AUTOMATION FOR THE COLOSSEUM DIGITAL TWIN

Besides the RF and application twinning, a second component in the Colosseum digital twin platform allows for the automated replica of end-to-end, full-stack Open RAN systems across real-world testbeds and Colosseum. Colosseum implements CI, CD, and Continuous Testing (CT) pipelines for automating the twinning and testing of protocol stacks, which can be used to address challenges associated to continuous software validation and automated integration and testing of disaggregated RAN components. The CI and CD pipelines rely on Red Hat's Ansible automation framework, while the CT process relies on Colosseum batch jobs.[5] A diagram of this workflow is shown in Figure 4.

The CI/CD is formed of two logical steps: (i) *build a new image* (enclosed with dashed red lines), and (ii) *perform an automated test* with the built image (dashed green line). Upon triggering the pipeline with a test to run, a check is made to ensure that an LXC image of the software to test exists. In case of positive outcome from this check, the automated tests are performed; otherwise, a new LXC image is built. At a high level, the steps to build a new LXC image involve three main tasks: (i) *setup a new LXC container*, (ii) *install the Open RAN software and dependencies*, and (iii) *save the LXC image*. Setting up a new LXC container involves a series of steps that start with pulling a base LXC image, such as a base Ubuntu 22.04 image, from the repository with the public Colosseum LXC images. The setup process continues with the configuration of SSH access to allow users to log into the container once deployed on Colosseum. The container is then prepared for the build process by installing necessary compilers and tools, creating users and groups, and installing various dependencies and tools required by Colosseum. Additionally, the UHD software suite, essential for communication with the SDRs, is installed. Then, the task to install the Open RAN software and dependencies clones the repository of the software to test (e.g., the OAI software); imports the profile

[5]Colosseum experiments can be run in interactive mode, where users are provided with a shell to the LXC containers to run their code and experiments, or in batch mode, where the infrastructure takes care of executing the experiments whenever resources are available in the system [14].

of the target node to build the software binary for (e.g., in case the compilation process is optimized for specific CPU instruction sets); installs the required dependencies, e.g., the ASN.1 definitions; builds the main software binaries; pulls some scripts to automated the testing process; and sets up the container for batch jobs execution on Colosseum. Finally, the task to save the resulting LXC image exports the container built so far as an LXC image; and saves it on the Colosseum storage.

After building the LXC image of the software to test, the step to perform automated tests through Colosseum batch jobs is run. First, the test setup, which details the batch job parameters, is specified. These parameters include details on the number of nodes involved in the test, the LXC image to load on each node (e.g., different nodes may act as gNB, UE, or core network, thus requiring different images), the RAN configuration (e.g., frequency and band to use), and the Colosseum RF and traffic scenarios to run the test with. A new batch job is then triggered on Colosseum. This involves, for instance, deploying a gNB providing service to cellular UEs—both implemented via the OAI protocol stack—through heterogeneous Colosseum RF scenarios. After the batch job ends, results are uploaded to the cloud, and a link to download them is returned by the CI/CD/CT pipeline so that users can easily retrieve them.

VII. CONNECTING REAL WORLD AND DIGITAL TWIN

The Colosseum digital twin can be integrated with external platforms and tools thanks to dedicated infrastructure or processes. Having real-world platforms integrated with a digital twin environment makes it possible to streamline the support for two main use cases, which can also be combined for closed-loop control: (i) update the twinned representation based on updates in the real-world setup; and (ii) update the real-world setup based on configurations tested and optimized within the digital twin platform.

The integration is performed either by connecting them directly (as described in Section VII-A for the infrastructure part, and in Section VII-B for the software enablers), or by transitioning experiments among platforms (Section VII-C) using dedicated software pipelines. In addition, it is possible to deploy and extend emulated experiments using discrete-event network simulators such as ns-3 and its O-RAN extensions [60].

A. INTEGRATION WITH ARENA AND X5G

The Arena testbed [31] is an over-the-air indoor programmable testbed with 24 SDRs connected to a grid of 64 antennas, enabling experimentation in the sub-7.2 GHz frequency spectrum. It is deployed in the Boston campus of Northeastern University, and co-located with X5G, a 5G- and O-RAN-compliant testbed based on NVIDIA Aerial, which implements the NR physical layer on GPU, OAI, and commercial O-RAN Radio Units (RUs) [30]. We

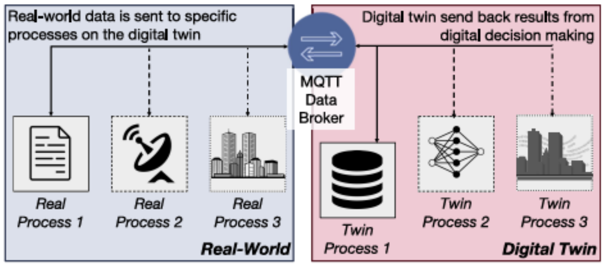

FIGURE 5. An overview depicting how the MQTT communication broker facilitates bidirectional data distribution to specific processes, enabling scalable implementation of digital twins.

leveraged the 10 Gbps shared metro link (round-trip time of ~3 ms) between the Northeastern Boston and Burlington, MA campuses (where Colosseum is located) to set up a site-to-site VPN between Colosseum and Arena, and to virtually merge the two networks through dedicated firewalls. This integration will allow Colosseum users to execute experiments on the Arena testbed through unified Web portal and LXC, thus enabling swiftly transition of experiments between real and emulated environments.

B. CROSS-TWIN COMMUNICATION BROKER FOR SCENARIO TWINNING

The Colosseum-Arena integration has also been used to prototype and test a software framework for the real-time twinning of a real-world scenario on Arena, and its digital counterpart on Colosseum. We leveraged the MQTT protocol to implement the real-time communication between the two endpoints [61]. Because of its publish/subscribe mechanism, this protocol facilitates smooth data interchange among numerous devices, guaranteeing minimal bandwidth consumption, low power usage, and steadfast connectivity, which are key to our real-time twinning, at the cost of some additional latency with respect to the link Round Trip Time (RTT).

The communication broker facilitates communication between devices and applications by managing the exchange of messages in a point-to-point manner. It allows for a single entity to control the flow of data in both directions, making management of the system streamlined, centralized, and responsive to changing data requirements. We placed the broker in the digital twin environment, as shown in Figure 5, ensuring that resource constraints are not added to the real-world system, which acts as data distributor.

Table 1 shows the communication latency between the two systems, illustrating how packet size impacts the latency. Overall, this has a limited contribution to the overall latency, with a difference of 0.2 ms between the smallest packet (e.g., sensor readings or simple command exchanges) and the largest ones (e.g., small datasets or spectrum information).

The largest contribution to latency comes from the usage of the MQTT broker, which, as discussed above, has benefits in terms of managing complexity at the cost of

TABLE 1. **Packet arrival times between the digital twin and the real-world counterpart in milliseconds (ms).** ***(Average latency of 100 samples with 95% confidence intervals).***

Packet Size	Latency Between Real-World Environment and Digital Twin (CI = 95%)
1 Byte	9.471 ms (9.411 ms, 9.531 ms)
100 Bytes	9.476 ms (9.411 ms, 9.540 ms)
1 Kbyte	9.569 ms (9.512 ms, 9.625 ms)
10 Kbytes	9.571 ms (9.520 ms, 9.623 ms)
100 Kbytes	9.615 ms (9.563 ms, 9.666 ms)
1 Mbyte	9.632 ms (9.565 ms, 9.698 ms)

some additional steps in the end-to-end message exchange (e.g., going through the message broker) that lead to a slight increase in latency. To show the direct impacts using MQTT has on the latency, we measure the time it takes to complete each step from start to finish of transmitting a packet. The total added latency, combining the broker processing (2.612 ms), MQTT protocol overhead (1.370 ms), RX packet process (0.002 ms), and network congestion (4.298 ms), amounts to 7.282 milliseconds. This helps show the slightly longer transmission times between the two entities under the MQTT implementation over the shared metro link.

C. FROM COLOSSEUM TO THE PAWR PLATFORMS

The capabilities of moving experiments between Colosseum and external testbeds have also been demonstrated in [27], where an experimental campaign was first executed on Colosseum, and then transitioned to the Arena testbed [31] and to the platforms of the PAWR program [54]. Specifically, the OpenRAN Gym toolbox was leveraged on Colosseum to run a data collection campaign, train ML xApps, and deploy them on a Near-RT RIC to perform data-driven control of a softwarized RAN implemented through the SCOPE framework and instantiated on an emulated urban RF scenario [15], [52]. Then, the same xApps and LXC containers used on Colosseum were transferred to the PAWR platforms and used to perform data-driven control of RANs in both outdoor and indoor environments, achieving comparable trends in real-world and emulated environments. This demonstrated the feasibility of transferring experiments between Colosseum and the external testbed.

VIII. USE CASES OF COLOSSEUM AS OPEN RAN DIGITAL TWIN

Colosseum has been extensively used to study open problems around the algorithmic, architectural, and system-level design of Open RAN systems, leveraging the realism in the RF scenario emulation and the possibility of running state-of-the-art software-defined protocol stacks that twin real-world scenarios. In this section, we review recent results as examples of Open RAN studies that can be performed on Colosseum.

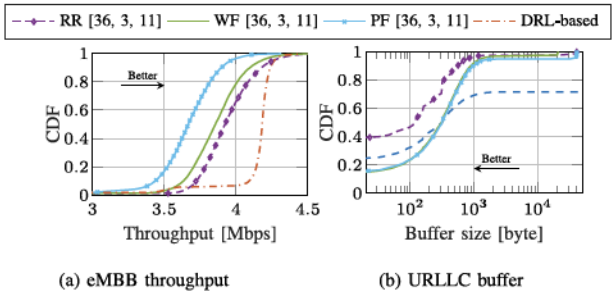

FIGURE 6. **Performance of a DRL agent controlling slicing and scheduling policies, from [15]. We compare the data-driven xApp and baselines without DRL-based adaptation. For the latter, the performance is based on the slicing configuration chosen with the highest probability by the best-performing DRL agent, and the three scheduler policies.**

A. SLICING AND SCHEDULING OPTIMIZATION

RAN slicing allows operators to dynamically partition the available spectrum bandwidth to accommodate UEs and traffic with heterogeneous Quality of Service (QoS), performance, and Service Level Agreement (SLA) requirements. A common example of UE and type of service in 5G is that of Enhanced Mobile Broadband (eMBB), Machine-type Communications (MTC), and Ultra Reliable and Low Latency Communications (URLLC). Another control configuration that we considered is the selection of a different scheduling policies for each of the above-mentioned RAN slices [15], [62].

To achieve flexibility in RAN control, AI/ML agents can be designed and embedded in xApps, then deployed on the Near-RT RIC and used to adapt the RAN to different network conditions and demand. Specifically, DRL can be leveraged in the design of such control solutions for the Open RAN as they can be trained a-priori and do not require prior knowledge of the run-time network dynamics [33]. Since these solutions require large datasets to be trained before they are deployed on a production RAN (to avoid disruption to its operations), the scale and controlled environment of digital twin offer fertile ground for their development.

Within this context, examples of Colosseum as the platform for the development of slicing and scheduling solutions for O-RAN systems are introduced in [15], [27], [33], [36], [62], [63], [64]. Through the OpenRAN Gym framework (described in Section IV-B), Colosseum provides the tools and infrastructure for data collection, design, training, testing, and evaluation of AI/ML xApps before their deployment on over-the-air deployments.

As an example of this, in [15] we developed, trained, and tested DRL-based xApps for slicing and scheduling control of a softwarized RAN on the Colosseum digital twin. The effectiveness of one of these xApps in controlling the scheduling (to be chosen among Round-robin (RR), Waterfilling (WF), and Proportionally Fair (PF)) and slicing (in terms of allocated Physical Resource Blocks (PRBs)) is shown in Figure 6(c). Specifically, Figure 6 shows the throughput performance of an eMBB slice, and Figure 6(b)

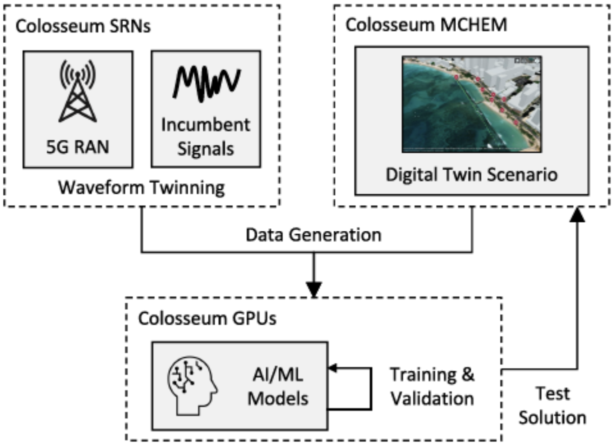

FIGURE 7. **Colosseum workflow for data-driven spectrum-sharing solutions development and testing.**

the buffer occupancy of an URLLC slice, with our adaptive DRL-based xApp and static baselines. We notice that the adapting the RAN configuration in near-real time allows the DRL-based xApp to outperform the baselines for the eMBB slice, while matching the best static configuration for URLLC one. The xApps trained in this way were then transitioned to the Arena testbed, where they adapted to over-the-air setup therein with only a few rounds of online learning, and to the PAWR platforms, where they exhibited similar trends to those achieved in Colosseum [27].

Finally, in [33], we further expanded on this analysis by benchmarking 12 xApps trained with different reward functions, action spaces, and hierarchical control. These xApps demonstrated how different design choices might deliver higher performance while others might result in competitive behavior between different classes of traffic with similar objectives.

B. SPECTRUM SHARING

Colosseum has also been used to evaluate spectrum-sharing solutions for Open RAN. These involve increased awareness of spectrum incumbents external to the RAN, as well as additional control actions (e.g., vacate a portion of the spectrum or null resource blocks). The digital nature of the systems allows for a safe evaluation of sharing and coexistence schemes, without risks of harming incumbents, before the solutions are transitioned into the real world.

Figure 7 shows an example of workflow that Colosseum enables for spectrum-sharing studies. This includes steps such as waveform and scenario twinning, testing, and validation of the designed sharing solutions.

The papers [29], [65] leverage this workflow to study AI-based sharing solutions, with the goal of improving RAN performance over static schemes or model-based approaches. Specifically, [65] leveraged Colosseum to collect a dataset with Wi-Fi and cellular users coexisting in the same frequency band, and to train a detector to identify spectrum usage patterns based on recurrent neural networks. Similarly, [29] demonstrated the use of Colosseum to carry out safe experiments with a radar incumbent operating in the Citizen Broadband Radio Service (CBRS) band [66], and to vacate cellular operations from such band if incumbent are present.

C. EXPLAINABLE AI FOR OPEN RAN

AI will be integral to the Open RAN, but ensuring the reliability of AI solutions within xApps and rApps for autonomous decision-making is still an open issue. While the RICs can leverage these applications to make decisions through data-driven logic, the effectiveness of the closed-loop control can be compromised if the AI algorithms fail (i) to adapt to changing conditions or (ii) to compute robust network policies in noisy and stochastic environments. This could lead to outages and sub-optimal performance.

Explainable AI (XAI) has emerged as a solution to this challenge by providing a platform to understand how AI agents take decisions [67]. XAI allows for the exploration of relationships between AI algorithm inputs and outputs, such as control actions, predictions, and forecasts. By studying such interactions, XAI solutions can validate that AI algorithms in xApps and rApps behave as expected, or explain why they do not. This can be particularly beneficial in achieving performance parity with legacy RAN systems, ensuring robustness against noise and attacks, and identifying areas for design improvement. How to design XAI solutions for Open RAN, and more generally, for closed-loop control based on complex DRL systems is an open challenge.

Colosseum can be used to design, develop, and test such XAI techniques for O-RAN applications. Specifically, Colosseum is unique in that it can facilitate XAI development as it: (i) supports testing of xApps and rApps across a variety of topology, mobility, channel, and traffic conditions to analyze their behavior across a variety of scenarios; (ii) offers a ready-to-use development framework to integrate XAI algorithms into xApps and rApps; and (iii) provides a realistic Open RAN digital twin that enables safe XAI analysis of xApps and rApps prior to their actual deployment on the production network, so as to avoid potential performance degradation and/or outages. For example, the EXPLORA framework, presented in [35], is an XAI tool for DRL-based xApps and rApps developed on Colosseum via OpenRAN Gym. EXPLORA generates high-level interpretable explanations that describe why certain actions have been selected by the DRL agent, as well as operational conditions (e.g., traffic conditions) that led to a certain choice. Moreover, the XAI insights generated by EXPLORA can be used to identify sub-optimal action-state pairs and prevent the enforcement of such actions by replacing them with other actions that deliver higher performance.

D. INTEGRATED ACCESS AND BACKHAUL OPTIMIZATION WITH OPEN RAN

IAB is a flexible wireless backhaul technology for cellular systems that has proved effective in reducing the deployment costs for ultra-dense RANs [68], [69]. It was introduced in the 5G NR standard specifications as part of 3GPP Release 16, and comes with a new kind of RAN entity, called IAB-node. The IAB node includes a Mobile Termination (MT) for wireless backhaul communications with upstream nodes, as well as the lower part of the cellular protocol stack (i.e., the layers in the DU and RU). IAB nodes are not connected through wired paths to the core network, instead, they interconnect to each other creating a multi-hop graph that leverages the same waveform, protocol stack, and potentially the same spectrum. This wireless graph terminates in base stations with wired connections to the core network, called IAB donors.

Colosseum has been used as a digital playground to prototype and test in-band IAB networks based on OAI. While not fully standard compliant, the solution developed in [36] allows for the deployment of IAB nodes and donors with minimal modifications to OAI. This enables to study relevant problems and use cases in the IAB domain, such as backhaul and access traffic partitioning, latency in multi-hop IAB, and optimal routing and scheduling. Since the wireless nature of the backhaul link in IAB increases the space and parameters for network control, the Open RAN architecture and the RICs can be leveraged to orchestrate and optimize IAB deployments with a data-driven approach.

E. SECURITY IN OPEN RAN

Security is another key concern and research and development area within Open RAN systems. Virtualization, softwarization, and open interfaces extend the attack surface of cellular systems, calling for additional testing and scrutiny into network management procedures, protocols, and implementations [70], [71], [72], [73], [74]. At the same time, open interfaces allow for increased visibility into the dynamics and telemetry of these systems, opening the door to robust data-driven anomaly- and intrusion-detection schemes [75], [76]. Finally, as the 5G cellular technology becomes more prevalent, the risk of more complex jamming attacks escalates, threatening network availability, reliability, and security [77], [78]. This makes the case for continuing research on jamming to safeguard future communications from malicious interference. As a digital twin based on real-time realistic RF emulation, Colosseum can be used as a sandbox to assess threats, test countermeasures, and validate protection against jammers without the risk of harming real-world over-the-air systems.

Colosseum for Evaluating Open RAN Security. The digital twinning capabilities of Colosseum have been used to evaluate the security of O-RAN interfaces and implementations in [79]. This assessment has been performed on the twinned 5G and Open RAN end-to-end protocol stacks deployed on the Colosseum emulation infrastructure

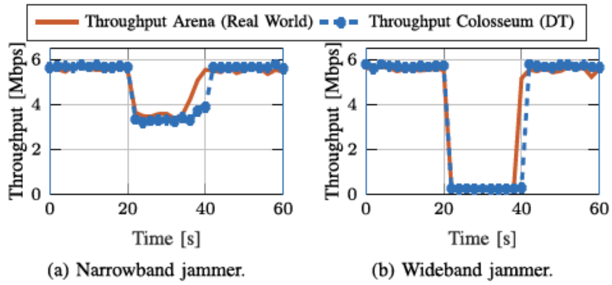

FIGURE 8. Throughput results from jamming experiments on the Arena and Colosseum testbeds, showcasing narrowband and wideband cases, adapted from [7].

in [80], focusing on the E2 interface and on numerically and experimentally evaluating the cost associated with supporting modern encryption techniques for the E2 SCTP stream.

Beside securing the interfaces, there is wide interest in evaluating how the openness of the O-RAN architecture makes it possible to design and implement more advanced anomaly- and intrusion-detection mechanisms. In [81], [82], the authors investigate traffic classification mechanisms for Open RAN systems, implemented on the Near-RT RIC for the analysis of specific KPIs associated with end-to-end application patterns. The paper [76] takes a step further and combines these insights with deep learning on the RAN I/Q samples (through a dApp [34]) to detect anomalies related to, for example, cloned user secure identity modules.

Colosseum for Evaluating Jamming Attacks. Understanding evolving jamming techniques is crucial to developing effective countermeasures and ensuring uninterrupted connectivity for critical services relying on 5G networks. However, often times laws and regulations prohibit causing intentional interference to authorized radio communication systems, which can significantly hinder large-scale jamming research. This opens the door for digital twins to offer improved jamming research due to the accurate, real-world emulation.

Large-scale jamming experimentation offers a unique look at how jamming attacks impact the spectrum. These signals have specific attack goals, but there is also incidental interference that can occur in wireless communications. By emulating real-world environments through digital twins, it is possible to show both the intentional and accidental consequences of jamming attacks and gain crucial insight on them. For example, in [83], Colosseum has been used to perform large-scale controlled jamming experiments with hardware in the loop and a commercial jamming system. The experimental findings reveal that the digital twin achieves an accuracy of up to 98% in replicating throughput, Signal to Interference plus Noise Ratio (SINR), and link status patterns when compared to real-world jamming experiments. Additionally, Figure 8(c) compares results for two types of jamming attacks, narrowband and wideband, against an OFDM system. The narrowband signal occupies ~ 156 kHz of the 20 MHz spectrum used for communications, the wideband signal occupies 10 MHz. By looking at the figure, we notice that, in the real-world experiments, narrowband jamming causes throughput decreases of around 37-43%, while wideband jamming of around 94-96%. Similar trends are observed in digital twin experiments.

IX. COLOSSEUM EXPERIMENTS: A PRIMER

Besides the digital twin capabilities that enable the use cases described above, Colosseum also comes with a framework that supports the execution of experiments and guides researchers through a workflow that involves the reservation of resources (i.e., SRNs, GPU nodes), the deployment of LXC or Docker containers on these resources, and the configuration of the RF and traffic scenarios. Access to Colosseum can be requested through a form on its public website [84]. The following paragraphs describe the main steps involved with executing an experiment in Colosseum. We refer the interested reader to further resources in [84], [85] for a detailed breakdown and guides on using the platform.

Colosseum users can interface with Colosseum shared resources through its experiment Web portal. On this webpage, users can check resource availability, view lists of scenarios, and manage images and reservations. Reservations can be made on the "Create New Reservation" page, where users can specify: (i) the name of the reservation; (ii) the time window when the resources are requested; (iii) the resources to be reserved, namely, SRNs and GPU nodes; (iv) the images to deploy to each node; and (v) additional flags and options for the current reservation. Specifically, for the LXC and Docker images, the user can leverage default ones provided through the Colosseum portal (e.g., those mentioned in Section IV), or prepare their own and upload them in the Colosseum Network Attached Storage (NAS).

Once the SRN reservation is completed and resources become available, users can access their reserved SRNs via SSH and interact with the resources available within each SRN as they would do in a local setup. During the experiment, users can also access the NAS and use the `colosseum-cli` utility to manage RF, traffic scenarios and snapshot the SRN LXC images. Indeed, at the end of the reservation, it is up to the user to save the container state and any generated data, or to copy files to the team network storage. Logs from the experiment will be accessible in the team's shared directory through the a file proxy server after the reservation ends.

Note that, during the reservation process, the choices on the scenario and configuration of the experiment are made manually by the users and, especially in cases of multiple reservations or numerous nodes, can require significant time and effort. For this reason, we have recently introduced the support for *blueprints* for Colosseum users. Blueprints are predefined sets of reservation information that can be used to pre-fill details on the "Create New Reservation" page of the experimenter portal. This feature not only simplifies the process by eliminating the need to manually select the number of nodes and their corresponding images each time but also aids in categorizing and providing a broad overview of the available types of experiments and setups in Colosseum.

X. COLOSSEUM FOR O-RAN TESTING AND INTEGRATION

The Colosseum Open RAN digital twin is also at the center of the recently established Open Testing and Integration Center (OTIC) in the Northeastern University Open6G center. An OTIC is a center officially recognized by the O-RAN ALLIANCE for the testing and integration of Open RAN components from multiple vendors. Currently, the main testing and integration activities carried out by OTICs worldwide relate to achieving interoperability in the disaggregated Open RAN architecture. For instance, OTICs can assess the conformance of an O-RAN interface or piece of equipment with respect to the O-RAN specifications, or certify the interoperability of pairs of components and end-to-end systems.

While the Colosseum data center is currently equipped with solutions to perform such testing, we also consider its twinning capabilities—together with the controlled and repeatable experimental environment and its AI/ML infrastructure—as a key component in driving the testing and, eventually, certification of AI solutions for the Open RAN. As of today, there are indeed discussions around interoperability certification and conformance of interfaces associated with the RICs, but testing and evaluating AI is an open challenge that requires further analysis and development.

AI solutions may indeed be non-deterministic, and exhibit different behaviors when developed in different scenarios and under dynamic conditions and network load. This makes defining requirements for the certification of AI solution a non-trivial problem. A possible strategy is leveraging benchmarking scenarios, as typically done with datasets in the computer vision community. In this sense, Colosseum, with its OpenRAN Gym framework, can implement and provide such reference scenarios to the community, leveraging a variety of wireless deployment scenarios, open-source protocol stacks, and either open-source or commercial RIC platforms. Testing can involve minimum accuracy or efficacy scores to achieve in the reference scenarios using xApps or rApps, adopting explainable AI frameworks, and providing confidence levels around AI agents misbehaving under unknown conditions.

Finally, Colosseum is evolving to accommodate custom and COTS radios connected to the digital twin RF emulation infrastructure. This makes it possible to onboard RF equipment that allows for evaluations beyond the capabilities of USRP SDRs, e.g., O-RAN-compliant RUs or commercial radars. Examples in the literature include studies with programmable Wi-Fi boards in [65], and jamming experiments in [7], [83].

XI. CONCLUSION AND FUTURE WORK

This paper reviewed how Colosseum, the world's largest wireless network emulator with hardware in the loop, can be leveraged as a digital twin for end-to-end Open RAN systems. We introduced the Colosseum architecture and focused on how it enables a high-fidelity digital replica of real-world RF environments, as well as the deployment of real-world protocol stacks, both through automated pipelines. We then discussed how the digital twin integrates with real-world testbeds, and introduced several use cases where Colosseum digital twin capabilities have been used to advance algorithms, architecture, and testing of Open RAN systems.

We showed how Colosseum, through its emulation capabilities and support for softwarized stacks, allows for the development of effective solutions for the control and optimization of the RAN, from slicing and radio resource management to spectrum sharing and explainable AI techniques.

Colosseum is constantly evolving to further improve its support for the design and evaluation of wireless systems. As future work, beside a refresh of the hardware infrastructure, we plan to introduce more flexible channel emulation options and to tighten and streamline the connectivity across the digital and real-world testbeds discussed in this work.

REFERENCES

[1] N. McKeown et al., "OpenFlow: Enabling innovation in campus networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, Mar. 2008.

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

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

[4] O-RAN Working Group 3, "O-RAN near-RT RAN intelligent controller near-RT RIC architecture 2.00," O-RAN Alliance, Alfter, Germany, document O-RAN.WG3.RICARCH-v02.00, Mar. 2021.

[5] B. Balasubramanian et al., "RIC: A RAN intelligent controller platform for AI-enabled cellular networks," *IEEE Internet Comput.*, vol. 25, no. 2, pp. 7–17, Mar./Apr. 2021.

[6] U. Challita, H. Ryden, and H. Tullberg, "When machine learning meets wireless cellular networks: Deployment, challenges, and applications," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 12–18, Jun. 2020.

[7] D. Villa et al., "Colosseum as a digital twin: Bridging real-world experimentation and wireless network emulation," *IEEE Trans. Mobile Comput.*, early access, Jan. 29, 2024, doi: 10.1109/TMC.2024.3359596.

[8] J. Breen et al., "POWDER: Platform for open wireless data-driven experimental research," in *Proc. ACM WiNTECH*, 2020, pp. 17–24.

[9] D. Raychaudhuri et al., "Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless," in *Proc. ACM MobiCom*, 2020, pp. 1–13.

[10] V. Marojevic, I. Guvenc, R. Dutta, M. L. Sichitiu, and B. A. Floyd, "Advanced wireless for unmanned aerial systems: 5G standardization, research challenges, and AERPAW architecture," *IEEE Veh. Technol. Mag.*, vol. 15, no. 2, pp. 22–30, Jun. 2020.

[11] H. Zhang et al., "ARA: A wireless living lab vision for smart and connected rural communities," in *Proc. ACM WiNTECH*, New Orleans, LA, USA, 2021, pp. 9–16.

[12] C.-F. Hung, Y.-R. Chen, C.-H. Tseng, and S.-M. Cheng, "Security threats to xApps access control and E2 interface in O-RAN," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1197–1203, 2024.

[13] S. Pongratz (Dell'Oro Group, Redwood City, CA, USA). *Open RAN Market Opportunity and Risks*, 2021. [Online]. Available: https://www.delloro.com/wp-content/uploads/2021/11/DellOro-Group-Article-Open-RAN-Market-Opportunity-and-Risks.pdf

[14] L. Bonati et al., "Colosseum: Large-scale wireless experimentation through hardware-in-the-loop network emulation," in *Proc. IEEE DySPAN*, 2021, pp. 105–113.

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

[16] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, and T.-T. Nguyen, "OpenAirInterface: Democratizing innovation in the 5G era," *Comput. Netw.*, vol. 176, May 2020, Art. no. 107284.

[17] I. Gomez-Miguelez, A. Garcia-Saavedra, P. Sutton, P. Serrano, C. Cano, and D. Leith, "srsLTE: An open-source platform for LTE evolution and experimentation," in *Proc. ACM WiNTECH*, 2016, pp. 25–32.

[18] L. Bonati, M. Polese, S. D'Oro, P. B. del Prever, and T. Melodia, "5G-CT: Automated deployment and over-the-air testing of end-to-end open radio access networks," 2023, *arXiv:2311.03206*.

[19] J. Jagannath, K. Ramezanpour, and A. Jagannath, "Digital twin Virtualization with machine learning for IoT and beyond 5G networks: Research directions for security and optimal control," in *Proc. ACM Workshop Wireless Secur. Mach. Learn.*, 2022, pp. 81–86. [Online]. Available: https://doi.org/10.1145/3522783.3529519

[20] A. Masaracchia, V. Sharma, M. Fahim, O. A. Dobre, and T. Q. Duong, "Digital twin for open RAN: Toward intelligent and resilient 6G radio access networks," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 112–118, Nov. 2023.

[21] M. Corici and T. Magedanz, "Digital twin for 5G networks," in *The Digital Twin*. Cham, Switzerland: Springer, 2023, pp. 433–446.

[22] H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, "Digital twin for 5G and beyond," *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 10–15, Feb. 2021.

[23] X. Lin, L. Kundu, C. Dick, E. Obiodu, T. Mostak, and M. Flaxman, "6G digital twin networks: From theory to practice," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 72–78, Nov. 2023.

[24] A. Alkhateeb, S. Jiang, and G. Charan, "Real-time digital twins: Vision and research directions for 6G and beyond," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 128–134, Nov. 2023.

[25] M. McManus et al., "Digital twin-enabled domain adaptation for zero-touch UAV networks: Survey and challenges," *Comput. Netw.*, vol. 236, Nov. 2023, Art. no. 110000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128623004450

[26] J. Mirzaei, I. Abualhaol, and G. Poitau, "Network digital twin for open RAN: The key enablers, standardization, and use cases," 2023, *arXiv:2308.02644*.

[27] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "OpenRAN gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms," *Comput. Netw.*, vol. 220, Jan. 2023, Art. no. 109502.

[28] D. Villa, M. Tehrani-Moayyed, P. Johari, S. Basagni, and T. Melodia, "CaST: A toolchain for creating and characterizing realistic wireless network emulation scenarios," in *Proc. ACM Workshop Wireless Netw. Testbeds, Exp. Eval. Charact. (WiNTECH)*, 2022, pp. 45–52.

[29] D. Villa, D. Uvaydov, L. Bonati, P. Johari, J. M. Jornet, and T. Melodia, "Twinning commercial radio waveforms in the colosseum wireless network emulator," in *Proc. 17th ACM Workshop Wireless Netw. Testbeds, Exp. Eval. Charact.*, 2023, pp. 33–40.

[30] D. Villa et al., "An open, programmable, multi-vendor 5G O-RAN Testbed with NVIDIA ARC and OpenAirInterface," in *Proc. 2nd Workshop Next-Gener. Open Program. Radio Access Netw. (NG-OPERA)*, 2024, pp. 1–6.

[31] L. Bertizzolo et al., "Arena: A 64-antenna SDR-based ceiling grid testing platform for sub-6 GHz 5G-and-beyond radio spectrum research," *Comput. Netw.*, vol. 181, pp. 1–17, Nov. 2020.

[32] J. Breen et al., "POWDER: Platform for open wireless data-driven experimental research," *Comput. Netw.*, vol. 197, pp. 1–18, Oct. 2021.

[33] M. Tsampazi et al., "A comparative analysis of deep reinforcement learning-based xApps in O-RAN," in *Proc. IEEE GLOBECOM*, 2023, pp. 1638–1643.

[34] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "dApps: Distributed applications for real-time inference and control in O-RAN," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 52–58, Nov. 2022.

[35] C. Fiandrino, L. Bonati, S. D'Oro, M. Polese, T. Melodia, and J. Widmer, "EXPLORA: AI/ML EXPLainability for the open

RAN," *Proc. ACM Netw.*, vol. 1, pp. 1–26, Nov. 2023. [Online]. Available: https://doi.org/10.1145/3629141
[36] E. Moro, G. Gemmi, M. Polese, L. Maccari, A. Capone, and T. Melodia, "Toward open integrated access and backhaul with O-RAN," in *Proc. 21st Mediterr. Commun. Comput. Netw. Conf. (MedComNet)*, 2023, pp. 61–69.
[37] I. Vilà, O. Sallent, and J. Pérez-Romero, "On the design of a network digital twin for the radio access network in 5G and beyond," *Sensors*, vol. 23, no. 3, p. 1197, 2023.
[38] P. Li et al., "RLOps: Development life-cycle of reinforcement learning aided open RAN," *IEEE Access*, vol. 10, pp. 113808–113826, 2022.
[39] A. Ndikumana, K. K. Nguyen, and M. Cheriet, "Digital twin assisted closed-loops for energy-efficient open RAN-based fixed wireless access provisioning in rural areas," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2023, pp. 6285–6290.
[40] J. Deng et al., "A digital twin approach for self-optimization of mobile networks," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, 2021, pp. 1–6.
[41] Y. Ren, S. Guo, B. Cao, and X. Qiu, "End-to-end network SLA quality assurance for C-RAN: A closed-loop management method based on digital twin network," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4405–4422, May 2024.
[42] S. K. Moorthy et al., "A Middleware for digital twin-enabled flying network simulations using UBSim and UB-ANC," in *Proc. 18th Int. Conf. Distrib. Comput. Sens. Syst. (DCOSS)*, 2022, pp. 322–327.
[43] J. Hu et al., "NeXT: Architecture, prototyping and measurement of a software-defined testing framework for integrated RF network simulation, experimentation and optimization," *Comput. Commun.*, vol. 210, pp. 342–355, Oct. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014036642300302X
[44] J. Baranda et al., "Demo: AIML-as-a-service for SLA management of a digital twin virtual network service," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–2.
[45] P. Bahl et al., "Accelerating open RAN research through an enterprise-scale 5G testbed," in *Proc. 29th Annu. Int. Conf. Mobile Comput. Netw.*, 2023, pp. 1–3.
[46] P. S. Upadhyaya, N. Tripathi, J. Gaeddert, and J. H. Reed, "Open AI cellular (OAIC): An open source 5G O-RAN testbed for design and testing of AI-based RAN management algorithms," *IEEE Netw.*, vol. 37, no. 5, pp. 7–15, Sep. 2023.
[47] P. S. Upadhyaya, A. S. Abdalla, V. Marojevic, J. H. Reed, and V. K. Shah, "Prototyping next-generation O-RAN research testbeds with SDRs," 2022, *arXiv:2205.13178*.
[48] H. Koumaras et al., "5GENESIS: The genesis of a flexible 5G facility," in *Proc. IEEE CAMAD*, Barcelona, Spain, 2018, pp. 1–6.
[49] A. Chaudhari and M. Braun, "A scalable FPGA architecture for flexible, large-scale, real-time RF channel emulation," in *Proc. 13th Int. Symp. Reconfig. Commun.-Centric Syst. Chip (ReCoSoC)*, 2018, pp. 1–8.
[50] "HashiCorp." Accessed: Feb. 2023. [Online]. Available: https://www.hashicorp.com/
[51] "Colosseum software." Colosseum. Accessed: Mar. 12, 2024. [Online]. Available: https://www.northeastern.edu/colosseum/colosseum-software/
[52] L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "SCOPE: An open and softwarized prototyping platform for NextG systems," in *Proc. ACM MobiSys*, 2021, pp. 415–426.
[53] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network automation through orchestrated intelligence in the open RAN," in *Proc. IEEE INFOCOM*, 2022, pp. 270–279.
[54] "PAWR." Accessed: Feb. 2024. [Online]. Available: https://www.advancedwireless.org
[55] A. Lacava et al., "ns-O-RAN: Simulating O-RAN 5G systems in ns-3," in *Proc. Workshop Ns-3*, 2023, pp. 35–44. [Online]. Available: https://doi.org/10.1145/3592149.3592161
[56] "Eric blossom and GNU radio project." GNU radio. Accessed: Feb. 2023. [Online]. Available: https://www.gnuradio.org.
[57] Colosseum Team, "Colosseum scenario summary list." Accessed: Feb. 2023. [Online]. Available: https://colosseumneu.freshdesk.com/a/solutions/articles/61000306089
[58] G. Gemmi, P. Johari, P. Casari, M. Polese, T. Melodia, and M. Segata, "ColosSUMO: Evaluating cooperative driving applications with colosseum," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Kobe, Japan, 2024, pp. 1–4.
[59] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Commun. Stand. Mag.*, vol. 1, no. 4, pp. 24–30, Dec. 2017.
[60] A. Lacava et al., "Programmable and customized intelligence for traffic steering in 5G networks using open RAN architectures," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2882–2897, Apr. 2024.
[61] (OASIS, Woburn, MA, USA). *MQTT Version 3.1.1*. 2014. [Online]. Available: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/
[62] 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 Commun. Mag.*, vol. 59, no. 10, pp. 21–27, Oct. 2021.
[63] R. Raftopoulos, S. D'Oro, T. Melodia, and G. Schembra, "DRL-based latency-aware network slicing in O-RAN with time-varying SLAs," 2024, *arXiv:2401.05042*.
[64] C. Puligheddu, J. Ashdown, C. F. Chiasserini, and F. Restuccia, "SEM-O-RAN: Semantic and flexible O-RAN slicing for NextG edge-assisted mobile systems," in *Proc. IEEE INFOCOM*, New York, NJ, USA, 2023, pp. 1–10.
[65] L. Baldesi, F. Restuccia, and T. Melodia, "ChARM: NextG spectrum sharing through data-driven real-time O-RAN dynamic control," in *Proc. IEEE INFOCOM*, 2022, pp. 240–249.
[66] R. Caromi, M. Souryal, and W. Yang, "Detection of incumbent radar in the 3.5 GHz CBRS band," in *Proc. IEEE GlobalSIP*, 2018, pp. 241–245.
[67] E. Puiutta and E. M. S. P. Veith, "Explainable reinforcement learning: A survey," in *Machine Learning and Knowledge Extraction*, A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds. Cham, Switzerland: Springer Int. Publ., 2020, pp. 77–95.
[68] C. Madapatha et al., "On integrated access and backhaul networks: Current status and potentials," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1374–1389, 2020.
[69] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G: Performance analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2669–2684, Dec. 2019.
[70] C. Shen et al., "Security threat analysis and treatment strategy for ORAN," in *Proc. 24th Int. Conf. Adv. Commun. Technol. (ICACT)*, 2022, pp. 417–422.
[71] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN," *Comput. Netw.*, vol. 217, Nov. 2022, Art. no. 109358.
[72] D. Dik and M. S. Berger, "Transport security considerations for the open-RAN fronthaul," in *Proc. IEEE 4th 5G World Forum (5GWF)*, 2021, pp. 253–258.
[73] S.-H. Liao, C.-W. Lin, F. A. Bimo, and R.-G. Cheng, "Development of C-plane DoS attacker for O-RAN FHI," in *Proc. 28th Annu. Int. Conf. Mobile Comput. Netw.*, 2022, pp. 850–852.
[74] A. Chiejina, B. Kim, K. Chowhdury, and V. K. Shah, "System-level analysis of adversarial attacks and defenses on intelligence in O-RAN based cellular networks," 2024, *arXiv:2402.06846*.
[75] H. Wen, P. Porras, V. Yegneswaran, A. Gehani, and Z. Lin, "5G-SPECTOR: An O-RAN compliant layer-3 cellular attack detection service," in *Proc. Netw. Distrib. Syst. Secur. (NDSS) Symp.*, 2024, pp. 1–20.
[76] A. Scalingi, S. D'Oro, F. Restuccia, T. Melodia, and D. Giustiniano, "Det-RAN: Data-driven cross-layer real-time attack detection in 5G open RANs," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2024, pp. 1–10.
[77] M. Lichtman, R. Rao, V. Marojevic, J. Reed, and R. P. Jover, "5G NR jamming, spoofing, and sniffing: Threat assessment and mitigation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2018, pp. 1–6.
[78] Z. A. E. Houda, H. Moudoud, and B. Brik, "Federated deep reinforcement learning for efficient jamming attack mitigation in O-RAN," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 9334–9343, Jul. 2024.
[79] J. Groen et al., "Implementing and evaluating security in O-RAN: Interfaces, intelligence, and platforms," Apr. 2023, *arXiv:2304.11125*.
[80] J. Groen, B. Kim, and K. Chowdhury, "The cost of securing O-RAN," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 5444–5449.

[81] J. Groen, M. Belgiovine, U. Demir, B. Kim, and K. Chowdhury, "TRACTOR: Traffic analysis and classification tool for open RAN," 2023, *arXiv:2312.07896*.
[82] M. Belgiovine, J. Gu, J. Groen, M. Sirera, U. Demir, and K. Chowdhury, "MEGATRON: Machine learning in 5G with analysis of traffic in open radio access networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, 2024, pp. 1054–1058.
[83] C. P. Robinson et al., "eSWORD: Implementation of wireless jamming attacks in a real-world emulated network," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2023, pp. 1–6.
[84] "Colosseum." Accessed: Dec. 2021. [Online]. Available: https://www.colosseum.net
[85] Colosseum Team, "Colosseum young gladiators 2024." 2024. [Online]. Available: https://github.com/colosseum-wiot/colosseum-school-2024