

OpenRAN Gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms[☆]

Leonardo Bonati^{*}, Michele Polese, Salvatore D'Oro, Stefano Basagni, Tommaso Melodia

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

ARTICLE INFO

Keywords:
5G/6G
O-RAN
AI/ML

ABSTRACT

Open Radio Access Network (RAN) architectures will enable interoperability, openness and programmable data-driven control in next generation cellular networks. However, developing and testing efficient solutions that generalize across heterogeneous cellular deployments and scales, and that optimize network performance in such diverse environments is a complex task that is still largely unexplored. In this paper, we present OpenRAN Gym, a unified, open, and O-RAN-compliant experimental toolbox for data collection, design, prototyping and testing of end-to-end data-driven control solutions for next generation Open RAN systems. OpenRAN Gym extends and combines into a unique solution several software frameworks for data collection of RAN statistics and RAN control, and a lightweight O-RAN near-real-time RAN Intelligent Controller (RIC) tailored to run on experimental wireless platforms. We first provide an overview of the various architectural components of OpenRAN Gym and describe how it is used to collect data and design, train and test artificial intelligence and machine learning O-RAN-compliant applications (xApps) at scale. We then describe in detail how to test the developed xApps on software-defined RANs and provide an example of two xApps developed with OpenRAN Gym that are used to control a network with 7 base stations and 42 users deployed on the Colosseum testbed. Finally, we show how solutions developed with OpenRAN Gym on Colosseum can be exported to real-world, heterogeneous wireless platforms, such as the Arena testbed and the POWDER and COSMOS platforms of the PAWR program. OpenRAN Gym and its software components are open-source and publicly-available to the research community. By guiding the readers from instantiating the components of OpenRAN Gym, to running experiments in a software-defined RAN with an O-RAN-compliant near-RT RIC and xApps, we aim at providing a key reference for researchers and practitioners working on experimental Open RAN systems.

1. Introduction

Once seen as monolithic and mostly immutable “black-box” systems, cellular networks are converging toward the more flexible, software-based open architectures based on the Open Radio Access Network (RAN) paradigm. This new approach to cellular communications promotes openness, virtualization, and programmability of RAN functionalities and components, and enables data-driven intelligent control loops for cellular systems [2]. As such, the Open RAN enables network operators to support new bespoke services on shared physical infrastructures, and to dynamically reconfigure them based on network conditions and user demand. The resulting increased efficiency will also decrease the operational costs of the network.

In this context, standardization bodies and other organizations are releasing a number of specifications to regulate the operations of the Open RAN, and to define its capabilities, constraints, and use cases. The most notable is the O-RAN Alliance, which is developing specifications—collected under the O-RAN umbrella—to apply Open RAN principles to prevailing radio access technologies, including 3rd Generation Partnership Project (3GPP) LTE and NR networks [3].

O-RAN introduces two network RAN Intelligent Controllers (RICs), operating at different timescales, enabling programmatic closed-loop control of the RAN elements. It also defines a set of open interfaces to connect the controllers to key elements of the RAN, such as the NR Central Units (CUs), Distributed Units (DUs), Radio Units (RUs), and the LTE O-RAN-compliant evolved Node Bases (eNBs) [4]. In details,

[☆] This is a revised and substantially extended version of Bonati et al. (2022) [1], which appeared in the Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC) 2022 Workshops.

This work was partially supported by the U.S. National Science Foundation under Grants CNS-1925601, CNS-2120447, and CNS-2112471.

^{*} Corresponding author.

E-mail addresses: l.bonati@northeastern.edu (L. Bonati), m.polese@northeastern.edu (M. Polese), s.doro@northeastern.edu (S. D'Oro), s.basagni@northeastern.edu (S. Basagni), melodia@northeastern.edu (T. Melodia).

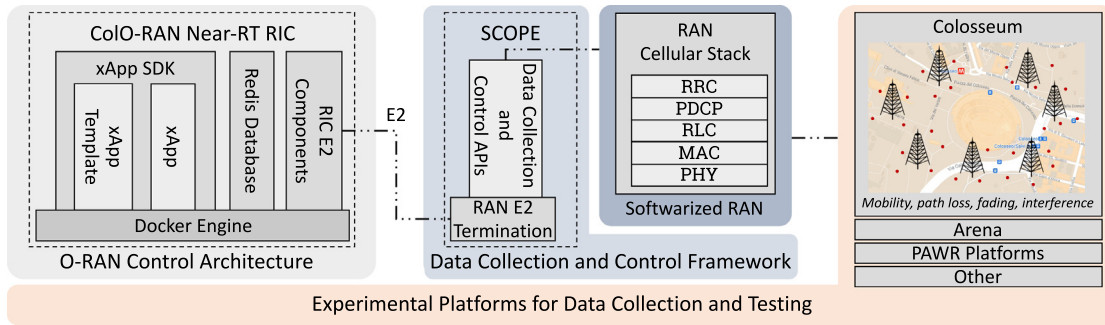


Fig. 1. OpenRAN Gym architecture.

the near-real-time (or near-RT) RIC connects to the RAN elements (i.e., the CUs and DUs) through the E2 interface, and enables control loops operating at timescales ranging between 10 ms and 1 s [5]. Instead, the non-real-time (or non-RT) RIC is included as part of Service Management and Orchestration (SMO) frameworks, and operates at timescales larger than 1 s [6]. This component also interacts with one or multiple near-RT RICs via the A1 interface, which is used to disseminate policies and information external to the network. The non-RT RIC also manages the Artificial Intelligence (AI) and Machine Learning (ML) models, which are instantiated on the RICs in the form of standalone applications, namely xApps (on the near-RT RIC) and rApps (on the non-RT RIC). Finally, the SMO connects to the RAN through the O1 interface, used for management and orchestration routines, and to the O-RAN virtualization platform (the O-Cloud) via the O2 interface.

Thanks to its RICs, open interfaces and disaggregated architecture, O-RAN ultimately enables the practical deployment and execution of AI/ML solutions at scale, which can be used to infer and forecast network traffic, or to reconfigure the nodes of the RAN at run time based on real-time conditions and user demand. Typical workflows for the design and testing of such AI/ML algorithms encompass a number of different steps such as [7,8]: (i) *data collection*, to create practical datasets representative of the different environments (e.g., the wireless channel) where the AI/ML models will be deployed, as well as of various performance indicators of the network; (ii) *AI/ML model design*, selecting the inputs and outputs of the models, and *training and testing*, to evaluate the effectiveness and limits of such models; (iii) *model deployment* as applications deployed on the RICs, i.e., xApps/rApps or—as recently proposed in [9]—directly on the CUs/DUs via dApps; (iv) *model fine-tuning* with run-time data from the RAN, to adapt the models to different production environments, and (v) the actual *control, inference and/or forecasting* on the RAN.

In this paper, we build upon and extend our prior work [1], and present OpenRAN Gym, an open-source toolbox to develop AI/ML O-RAN-compliant inference and control algorithms, to deploy them as xApps on the near-RT RIC, and to test them on a large-scale softwarized RAN controlled by the RIC. Throughout this work, we guide readers on how to start using the OpenRAN Gym framework to run experiments in a softwarized Open RAN managed by an O-RAN-compliant near-RT RIC, how to implement custom AI/ML solutions of their design as xApps on the Colosseum wireless network emulator, and how to transition such solution on over-the-air open testbeds for wireless research. First, we give a high-level overview of the various components of OpenRAN Gym, and discuss how they enable development and testing workflows of data-driven xApps. We showcase an example of two xApps designed with OpenRAN Gym and used to control a large-scale RAN instantiated on the Colosseum wireless network emulator [10] through the SCOPE framework [11], and controlled by the CoIo-RAN near-RT RIC [8]. We also show how OpenRAN Gym can be seamlessly ported from an emulator such as Colosseum to over-the-air real-world platforms, such as the Arena testbed [12], and the platforms of the U.S. National Science Foundation-sponsored Platforms for Advanced

Wireless Research (PAWR) program [13] including the Platform for Open Wireless Data-driven Experimental Research (POWDER) [14] and the Cloud Enhanced Open Software Defined Mobile Wireless Testbed for City-Scale Deployment (COSMOS) [15] platforms. To the best of our knowledge, OpenRAN Gym is the first open, portable toolset for end-to-end design, prototyping, testing, and experimentation of AI/ML O-RAN xApps on heterogeneous wireless platforms. As such, we hope that this work can be an important reference for researchers and practitioners working on—or starting to work on—experimental Open RAN systems.

Previous experimental work has focused on the development of data-driven solutions and xApps for specific use cases [16,17], on the description of the AI/ML capabilities of O-RAN [18,19], on interoperability testing [20], and on orchestration [21]. Compared to the state of the art, OpenRAN Gym enables an end-to-end workflow for the design and testing of AI/ML solutions as xApps in the O-RAN ecosystem. By doing so, it empowers users with a first-of-its-kind open and publicly-available O-RAN-compliant toolbox that will unleash the potential of data-driven applications for next generation cellular networks. OpenRAN Gym aims at contributing to the thriving community of wireless researchers and developers by providing open-source software components for experimental O-RAN-enabled data-driven research. We actively maintain an up-to-date resource on the OpenRAN Gym project¹ that can be used to review the functionalities of our framework, and as a reference to repositories, documentation, tutorials, and containers. This also includes publications that describe in details each component, the different use cases in which they were used, as well as public datasets collected through OpenRAN Gym.

The remainder of this paper is organized as follows. We give an overview of the various components of OpenRAN Gym in Section 2. Practical descriptions of the OpenRAN Gym data collection and control framework, and of the O-RAN control architecture, are given in Sections 3 and 4, respectively. The xApp design and testing workflow is presented in Section 5, along with an example of large-scale RAN control using xApps developed with OpenRAN Gym on Colosseum. Section 6 discusses how OpenRAN Gym components and experiments can be ported from Colosseum to heterogeneous real-world testbeds. Section 7 showcases exemplary results obtained on the different platforms considered in this work. Finally, conclusions are drawn in Section 8.

2. OpenRAN Gym

The OpenRAN Gym architecture is shown in Fig. 1. Its main components are: (i) publicly- and remotely-accessible *experimental wireless platforms* for collecting data, prototyping, and testing solutions in heterogeneous environments. Example of these are the Colosseum wireless network emulator [10], the Arena testbed [12], and the platforms of the PAWR program [13]; (ii) a *softwarized RAN* implemented through open protocol stacks for cellular networks, such as srsRAN [22] and

¹ <https://openrangym.com>

OpenAirInterface [23]; (iii) a *data collection and control framework*, such as SCOPE [11], that exposes Application Programming Interfaces (APIs) to extract relevant Key Performance Measurements (KPMs) from the RAN, and dynamically control it at run-time, and (iv) an *O-RAN control architecture*, such as CoLo-RAN [8], able to connect to the RAN through open and standardized interfaces (e.g., the O-RAN E2 interface), receive the run-time KPMs from the RAN, and control it through AI/ML solutions running, for instance, as xApps/rApps. As we will show in Sections 6 and 7, OpenRAN Gym is platform-independent, and it allows users to perform data collection campaigns, prototype, and evaluate solutions in a set of heterogeneous wireless environments and deployments before transitioning them to production networks. As such, OpenRAN Gym can be used to first prototype and validate solutions on the Colosseum wireless network emulator, and then seamlessly transfer such solutions to heterogeneous platforms, such as the Arena testbed, and the POWDER and COSMOS platforms from the PAWR program [13]. The procedures to port the various components of OpenRAN Gym on these platforms will be described in Section 6.

Arena is an indoor wireless testbed equipped with a grid of 64 antennas and 24 Software-defined Radios (SDRs) (among USRPs X310 and N210) controlled by high-performance compute servers [12]. Its deployment is representative of a live office environment.

Colosseum is the world's largest wireless network emulator [10]. It allows researchers and practitioners to experiment at scale, and in different channel conditions and virtual environments through a set of 128 SDRs (USRPs X310) controlled through dedicated servers—namely, Standard Radio Nodes (SRNs)—interconnected through a Massive Channel Emulator (MCHEM). The latter, is capable of reproducing conditions of the wireless channel (e.g., path loss, fading, user mobility, signal interference and superimposition) by means of Finite Impulse Response (FIR) filters implemented through Field Programmable Gate Arrays (FPGAs). The channel emulation is performed by the FIR filters, which apply the channel impulse response of the desired wireless channel to the signals transmitted by the SRNs. Sets of channel impulse responses for different environments (e.g., urban, rural, etc.)—referred to as *Radio Frequency (RF) scenarios* in Colosseum—are modeled through mathematical equations, or captured through ray-tracing software.

POWDER is a city-scale wireless testbed deployed in Salt Lake City, UT [14]. The testbed includes a number of SDRs deployed across an outdoor area, an over-the-air indoor laboratory setup, and a wired attenuator matrix. The objective of this testbed is to foster experimental research in heterogeneous technology, such as 5G cellular technologies and network orchestration.

COSMOS is a city-scale testbed deployed in New York City, NY, which mainly focuses on mmWave communications with edge-computing capabilities [15]. This testbed absorbed the Open-Access Research Testbed for Next-Generation Wireless Networks (ORBIT) [24], an indoor over-the-air wireless platform with remotely-accessible SDR devices and compute servers.

At the time of this writing, OpenRAN Gym softwarized RAN leverages the cellular implementation provided by srsRAN [22], which allows users to instantiate protocol stacks of 3GPP base stations and User Equipments (UEs) using SDRs as front-end interfaces. This cellular protocol stack is augmented by the SCOPE framework, which adds a number of networking and control functionalities to srsRAN including network slicing capabilities, support for additional scheduling algorithms, data collection pipelines, and open APIs to control such functionalities at run time. As we will discuss in Section 3, SCOPE can facilitate data collection campaigns by automating the collection of relevant RAN KPM in the heterogeneous testbed where it is instantiated [8,21,25].

Finally, CoLo-RAN implements the O-RAN control architecture of OpenRAN Gym. This framework adapts the near-RT RIC provided by the O-RAN Software Community (OSC) to run in a lightweight containerized environment, and extends it to swiftly interface with, and control, the SCOPE base stations through the E2 interface standardized

by O-RAN. As discussed in Section 4, CoLo-RAN allows users to prototype AI/ML-based O-RAN applications through an *xApp Software Development Kit (SDK)*, to instantiate them on an OSC-compliant near-RT RIC, and to leverage them to control a softwarized RAN (Fig. 1).

3. Data collection and control framework

The data collection and control framework of OpenRAN Gym is based on SCOPE [11]. This framework provides a programmable environment for prototyping and testing solutions for softwarized RANs, and data collection capabilities of relevant KPMs (e.g., throughput, Transport Blocks (TBs), buffer occupancy). Concerning the cellular protocol stack for base stations and UEs, SCOPE leverages srsRAN [22], and extends it with novel network slicing and a set of additional scheduling algorithms. Open APIs to fine-tune the configuration of the RAN at run time, and to perform data collection campaigns are also provided by SCOPE. Coupled with different testbeds—such as Colosseum and the platform of the PAWR program—SCOPE can facilitate the collection of RAN KPMs in a set of heterogeneous scenarios and environments by automatically collecting such statistics from the running experiments [8,11,25]. Finally, SCOPE connects to the O-RAN near-RT RIC through a RAN-side O-RAN E2 termination, which is based on the OSC DU [26]. This allows user-defined xApps running on the near-RT RIC to swiftly interface with the RAN base stations, and to dynamically control their functionalities at run time (e.g., modify the scheduling policy and set the amount of resources allocated to each network slice). In the remainder of this section, we will give a high-level overview of the main configuration options and parameters of the SCOPE-enabled base stations, and show how to instantiate a cellular network with it. SCOPE has been open-sourced to the research community,² and also provided to the Colosseum users in the form of a ready-to-use Linux Container (LXC) (namely `scope/scope-with-e2`). In Section 6, we will show how the publicly-available SCOPE container can be ported to different testbeds (e.g., the Arena testbed, and the POWDER and COSMOS testbeds of the PAWR program) with minor modifications. In this way, SCOPE truly enables the process of cellular-network-as-a-service, in which the solutions are first prototyped in a controlled environment (e.g., Colosseum), and then ported in the wild on real-world testbeds.

3.1. Starting SCOPE

SCOPE provides Command-line Interface (CLI) tools to start the cellular base stations and configure them through parameters passed via configuration files. The main parameters of interest to OpenRAN Gym are described as follows.³

- `network-slicing`: enables/disables the network slicing functionalities of the base station.
- `slice-allocation`: if network slicing has been enabled, this parameter sets the Resource Block Groups (RBGs) allocated by the base station to each slice. The input of this configuration option is passed as `{slice: [first_rbg, last_rbg], . . . }`. As an example, `{0: [0, 5], 1: [6, 10]}` allocates RBGs 0–5 to slice 0 and 6–10 to slice 1.
- `slice-scheduling-policy`: sets the scheduling policy used for each network slice of the base station. As an example, `[1, 2]` assigns slicing policy 1 to slice 0 and policy 2 to slice 1. The possible numerical values for this field match the scheduling policies supported by SCOPE (i.e., 0: round-robin, 1: waterfilling, 2: proportionally fair).

² The SCOPE source code is available at <https://github.com/wineslab/colosseum-scope> and <https://github.com/wineslab/colosseum-scope-e2>.

³ A comprehensive description of the SCOPE APIs and configuration parameters can be found at <https://github.com/wineslab/colosseum-scope>.

- `slice-users`: associates UEs to a specific network slice. The input of this configuration option is passed as `{slice:[ue1,ue2], ...}`. As an example, `{0:[4,5],1:[2,3]}` assigns UEs 4, 5 to slice 0, and UEs 2, 3 to slice 1.
- `generic-testbed`: specifies whether SCOPE is running on a testbed other than Colosseum. In this case, the parameters `node-is-bs` and `ue-id` can also be passed to specify whether the node should act as a base station or a UE, and the identifier of the UE in the latter case.⁴

After the SCOPE configuration has been written in a JSON-formatted file,⁵ (named `radio.conf` in the code snippet below), the cellular base station, core network, and UE applications can be started through the commands of Listing 1.

```

1 #!/bin/bash
2 cd radio_api/
3 python3 scope_start.py --config-file
  radio.conf

```

Listing 1: Commands to start the SCOPE applications.

At run-time, the SCOPE APIs can be leveraged to fine-tune the configuration of the base station, e.g., to modify the scheduling policy of each slice, or to set the amount of RBGs of each slice (see [11, Section 3.3]). Relevant KPMs from the RAN are automatically logged by the SCOPE base stations while traffic is exchanged among base stations and UEs. These KPMs are saved in CSV-formatted files that can be either used on-the-fly (e.g., for online AI/ML model training or inference), or retrieved for offline processing after the experiment ends (e.g., to perform offline AI/ML model training).

4. O-RAN control architecture

The O-RAN control architecture leveraged by OpenRAN Gym is based on CoO-RAN, an open-source framework to develop, design, prototype, and test O-RAN-ready solutions at scale [8]. This framework provides a lightweight implementation of the OSC near-RT RIC—which has been adapted to run on the Colosseum system as a set of standalone Docker containers—as well as automated pipelines for the deployment of the various services of the RIC. The main components implemented by CoO-RAN are shown in Fig. 1, left. They are services in charge of overseeing the interactions with the RAN (e.g., the *E2 termination*, *E2 manager*, and *E2 routing manager*), a *Redis database* that keeps records of the connected RAN nodes (e.g., the base stations), and an *xApp SDK* with tools to prototype and test AI/ML-based xApp for run-time RAN inference and/or control.

A high-level diagram of a CoO-RAN xApp is shown in Fig. 2. This is formed of two main parts: (i) the *Service Model (SM) connector*, in charge of handling the messages between the xApp and the near-RT RIC (e.g., the control messages for the RAN), and (ii) the *data-driven logic unit*, which processes KPMs received from the RAN base stations, and performs tasks based on AI/ML models (e.g., traffic prediction and/or control of the functionalities of the base stations). The data-driven logic unit hosts two sub-components, namely the *AI/ML model* and the *data processing module*. The former consists of the specific data-driven model (e.g., Deep Reinforcement Learning (DRL) agent, Deep Neural Network (DNN), Long Short Term Memory (LSTM), to name a

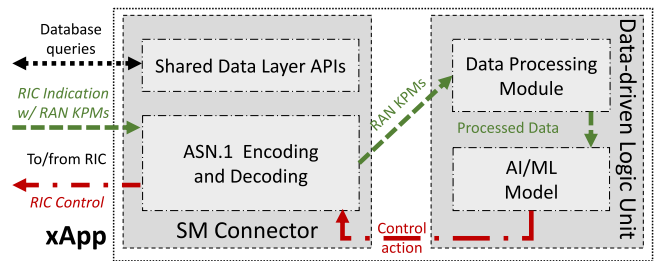


Fig. 2. CoO-RAN xApp. Source: Adapted from [8].

few) used to perform inference and/or control tasks. The latter, instead, executes data processing functionalities to convert the input KPMs into data that can be fed to the AI/ML model. For instance, the majority of AI/ML models are designed to receive inputs with a fixed size and format (e.g., a two-dimensional array of a specific length, an image, or a time-series). However, the KPMs received over the E2 termination might have a different format, or might contain more data than what is required by the AI/ML model. In this case, the data processing module performs the necessary operations to convert the input data in the correct format. In some cases, the data processing module can also host some AI/ML models that execute advanced data processing operations. Examples of these are autoencoders to extract latent data representation and to perform dimensionality reduction [8,25].

In the remainder of this section, we detail how to instantiate the CoO-RAN near-RT RIC (Section 4.1), how to interface it with the SCOPE base station through the O-RAN E2 termination (Section 4.2), and how to start a sample xApp that controls the base station (Section 4.3). CoO-RAN has been open-sourced and made available to the research community,⁶ and also provided to the Colosseum users in the form of a ready-to-use LXC container (namely `coloran-near-rt-ric`). In Section 6, we will show how CoO-RAN can be seamlessly ported to different testbeds (e.g., the Arena testbed, and the POWDER and COSMOS platforms of the PAWR program).

4.1. Starting the CoO-RAN near-RT RIC

The CoO-RAN near-RT RIC can be built and instantiated as a set of Docker containers by running the `setup-ric.sh` script and the commands of Listing 2.⁷ This script, adapted from [27], takes as input the network interface the RIC uses to receive and exchange messages with the RAN (e.g., the `col0` interface in Colosseum). As a first step, the base Docker images that will be used to build the RIC are imported. Then, the actual Docker images of the CoO-RAN near-RT RIC are built. These images include: (i) the `e2term`, which is the endpoint of the RIC E2 messages; (ii) the `e2mgr`, which is in charge of managing the messages to/from the E2 interface; (iii) the `e2rtmansim`, which uses the RIC Message Router (RMR) protocol to route the E2 messages within the RIC; and (iv) the `db`, which implements a Redis database with records of the RAN nodes connected to the RIC (e.g., the base stations). During this step, the IP addresses and ports that will be used by the Docker containers are also configured as set up in the `setup-lib.sh` file. After the Docker images have been built, the RIC containers, listening for incoming connections from the RAN through the E2 termination endpoint, are spawned. The logs of the various containers can be accessed through the `docker logs` command, e.g., `docker logs -f e2term` shows the run-time logs of the E2 termination (`e2term`) container.

⁴ When running on Colosseum, SCOPE automatically derives the role of the node, and the UE identifier based on the allocated SRNs.

⁵ An example of configuration file can be found at <https://tinyurl.com/2s3pvw83> (for Colosseum), and at <https://tinyurl.com/35t7s97a> (for testbeds other than Colosseum).

⁶ The CoO-RAN source code is available at <https://github.com/wineslab/colosseum-near-rt-ric>.

⁷ A pre-built version of CoO-RAN, named `coloran-near-rt-ric-prebuilt`, is also provided to the Colosseum users. The pre-built Docker images are also hosted on Docker Hub at <https://hub.docker.com/u/wineslab>.

```

1 #!/bin/bash
2 cd setup-scripts/
3 ./setup-ric.sh col0

```

Listing 2: Commands to set up the CoLo-RAN near-RT RIC.

4.2. Connecting the SCOPE base station to CoLo-RAN

After setting up and starting CoLo-RAN through the steps described in Section 4.1, the cellular base station—provided by SCOPE and set up in Section 3.1—can be connected to it through the O-RAN E2 termination, which has been adapted from the OSC DU implementation [26]. To this aim, the RAN-side E2 termination can be used to: (i) receive *RIC Subscription* messages from the xApps; (ii) transmit periodic KPM reports to the xApps through *RIC Indication* messages, and receive control actions from them through *RIC Control* messages, and (iii) interact with the APIs provided by SCOPE to modify the configuration of the base station at run time (e.g., the scheduling and slicing policies) based on the control messages received from the xApps. The steps to initialize the E2 termination at the SCOPE base station are shown in Listing 3. First, the E2 termination is built through the `build_odu.sh` script (line 3). This script also specifies the IP address and port of the near-RT RIC to connect to, as well as the local network interface used for the connection. Then, the E2 termination process is started through the `run_odu.sh` script (line 4), which establishes the initial connection between the base station and the near-RT RIC. The successful outcome of this connection can be verified in the logs of the `e2term` container (via the `docker logs -f e2term` command, see Section 4.1), which reports the identifier of the connected base station (e.g., `gnb:311-048-01000501`).

```

1 #!/bin/bash
2 cd colosseum-scope-e2/
3 ./build_odu.sh clean
4 ./run_odu.sh

```

Listing 3: Commands initialize the SCOPE E2 termination process.

4.3. Initializing a sample xApp

After the SCOPE base station has been connected to the near-RT RIC, the xApps can be started. To facilitate the design of novel xApps, we provide a ready-to-use sample xApp template in which researchers and practitioners can plug-in their custom AI/ML models. This sample xApp can be started through the `setup-sample-xapp.sh` script and the commands shown in Listing 4. This script takes as input the identifier of the base station the xApp should subscribe to (see Section 4.2), and builds the Docker image of the sample xApp. Then, the script starts the xApp Docker container—dubbed `sample-xapp`—on the near-RT RIC.

```

1 #!/bin/bash
2 cd setup-scripts/
3 ./setup-sample-xapp.sh gnb:311-048-01000501

```

Listing 4: Commands to build the CoLo-RAN sample xApp Docker image, and to start and configure the xApp container.

After the container has started, the xApp processes can be run through the commands of Listing 5. These commands trigger the xApp subscription to the targeted RAN nodes (e.g., one or multiple base stations connected to the RIC) through *RIC Subscription* messages, and the periodic reports of RAN KPMs from such nodes. Starting from

the provided template, OpenRAN Gym users can build xApps running custom solutions (e.g., with custom AI/ML agents).

```

1 #!/bin/bash
2 docker exec -it sample-xapp
   /home/sample-xapp/run_xapp.sh

```

Listing 5: Commands to run the CoLo-RAN sample xApp process.

5. xApp development workflow on Colosseum

The main steps to develop a data-driven xApp using OpenRAN Gym on Colosseum are shown in Fig. 3.

(1) **Data collection.** This step involves collecting the data that will be used to train and test the AI/ML model to embed in the xApp. In Colosseum, this can be done by combining the data-collection capabilities of SCOPE with the automated experiments of Colosseum. This allows to automatically run experiments with several base stations and users in a set of heterogeneous scenarios, and to collect the RAN KPMs—saved in CSV-formatted files—from Colosseum Network Attached Storage (NAS) once the experiment ends [11].

(2) **Model design, training and testing.** After data collection campaigns have been performed in heterogeneous wireless environments and scenarios, the AI/ML model can be designed. This step includes the selection of the AI/ML algorithm that the model will use, along with the data used as input, the reward function, and the set of output actions (e.g., to perform inference or control of the RAN). After this design phase, the model is first trained offline using the data collected in step 1, and then tested at scale.⁸ Being computationally-intensive, this step may benefit from Graphics Processing Unit (GPU)-enabled environments. As such, these operations can be carried out either locally (i.e., on the user’s own GPU-enabled machines), or on Colosseum’s GPU-enabled SRNs or NVIDIA A100 DGXs.

(3) **Deploy the model as an xApp.** After the model has been tested (step 2), it can be deployed as an xApp on the CoLo-RAN near-RT RIC by following the procedures of Section 4.3. Specifically, the AI/ML model is included in the *data-driven logic unit* of CoLo-RAN xApp (see Fig. 2) by modifying the provided xApp template. Finally, the modified xApp is built and instantiated on the near-RT RIC through the commands of Listings 4 and 5.

(4) **Online model fine-tuning.** At run-time, the xApp communicates with the SCOPE base station through the near-RT RIC and the E2 termination. To this aim, the xApp first subscribes to the base station by sending it a *RIC Subscription* message. Then, it triggers periodic KPMs reports—with periodicity tunable based on the needs of the users [8]—from the base station. These reports are sent through *RIC Indication* messages, and they may be used by the xApp to fine-tune the model online, allowing it to adapt to varying wireless conditions and traffic demand. Once the model has been fine-tuned online, the Docker image of the xApp can be updated with the trained weights.

(5) **Perform RAN control/inference.** At this stage, the xApp can be used in the live infrastructure to perform inference and/or control of the RAN. This entails the xApp transmitting the actions computed by the model to the SCOPE base station through *RIC Control* messages. Example of these actions to modify the parameters and configuration of the base station, e.g., to modify the resources allocated to the slices of the network, or their scheduling policies. At the base station, these *RIC Control* messages—received through the O-RAN E2 interface—trigger the SCOPE control APIs of Fig. 3, which apply the new policies to the configuration of the base station at run time. At this point, the xApps

⁸ It is worth mentioning that the O-RAN specifications forbid the deployment of models that have not been trained offline beforehand. This is to shield the RAN from poor performance or outages [7].

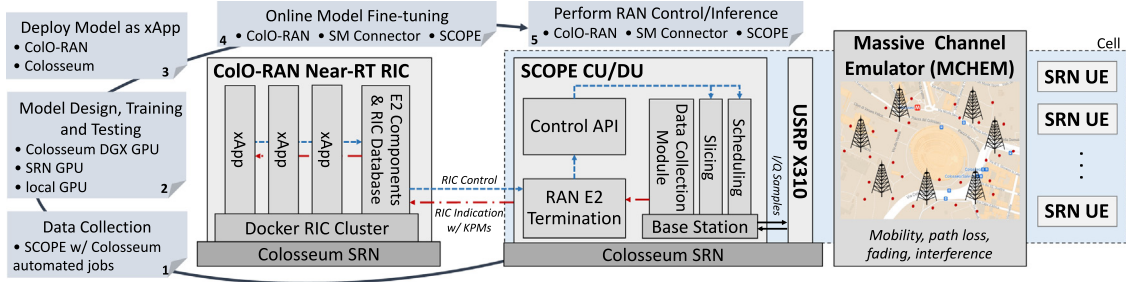


Fig. 3. OpenRAN Gym xApp design and testing workflow on Colosseum.

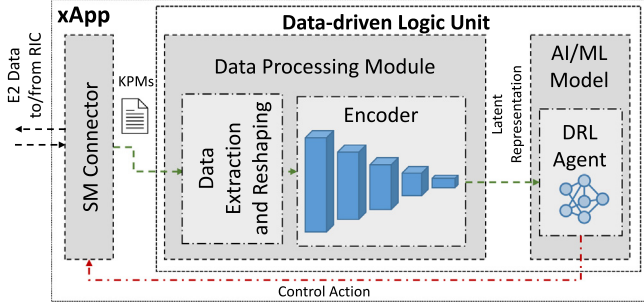


Fig. 4. The architecture of the xApps used in Section 5.1. Source: Adapted from [28].

can be tested and validated on Colosseum. In Sections 6 and 7, we will show how these newly developed xApps can be ported and instantiated on external wireless testbeds.

5.1. Example of xApps

For the sake of completeness, we now provide an example of two xApps designed, trained and tested with OpenRAN Gym on Colosseum. These xApps are used to control a cellular network with 7 base stations and 42 UEs instantiated on the Colosseum network emulator. Base stations adopt a Frequency Division Duplexing (FDD) configuration with 50 Physical Resource Blocks (PRBs) (corresponding to 10 MHz of bandwidth). Each base station is implemented through SCOPE and serves 6 UEs with different traffic requirements. The UEs are divided into two classes of traffic, allocated to different slices of the network: time-sensitive (e.g., Ultra Reliable and Low Latency Communications (URLLC)) and broadband (e.g., Enhanced Mobile Broadband (eMBB) and Machine-type Communications (MTC)). Further examples of xApps developed with OpenRAN Gym and used to optimize the network performance of a softwareized RAN instantiated on a general-purpose infrastructure are discussed in details in [8,21,25,28].

The xApp structure used in this section is shown in Fig. 4. Although the general architecture stems from that depicted in Fig. 2, in this specific example, the data-driven logic unit consists of two elements: (i) an encoder for data dimensionality reduction, which is in charge of converting KPM reports received on the E2 interface into a latent (and low dimension) representation of the data, and (ii) a DRL agent that converts the latent representation into a state of the network and computes the optimal control action that maximizes an agent-specific reward according to the current state.

Agent Design. DRL agents are trained on a dataset with 3.4 GB of RAN traces (and more than 73 h of experiments) collected on Colosseum—that make control decisions on the configuration of the base station based on the received RAN KPMs (see [8]). The DRL agents considered in this paper implement a Proximal Policy Optimization (PPO) architecture that leverages an actor-critic structure. The actor network is trained to take actions according to the current state of

the system, while the critic network is used during the training phase to evaluate the reward obtained by selecting a specific action in a certain state. Then, the critic network instructs the actor network on how valuable the action was, in this way steering the actor network toward actions that bring the highest reward for each state. Both actor and critic networks consist of fully-connected neural networks with 5 layers, where each layer has 30 neurons.

Actions. To showcase the impact of different design choices on the overall performance of the network, we trained two xApps with different action spaces. One xApp (named *sched*) controls the scheduling policies that a base station uses for specific classes of traffic. Another xApp (*sched-slicing*) is instead operating over a larger action space as it controls both scheduling policies, and the resource allocated to each slice (i.e., the number of RBGs assigned to each class of traffic).

Reward. In this example, both xApps aim at maximizing the amount of transmitted data belonging to the broadband traffic class (in this case measured by the number of downlink TBs transmitted successfully by the base station to the UEs), and minimizing the end-to-end latency of the time-sensitive traffic class. As the protocol stack of the base station does not have a direct measurement of the end-to-end system latency, we use the buffer occupancy metric as our proxy for latency, which reflects how much time packets spend in the transmission buffer queue.

To capture these objectives, the reward of the DRL agents consists of a reward function that jointly maximizes the throughput for the broadband slice and minimizes the downlink buffer size for the time-sensitive slice. These two elements are combined with a weighted sum, whose details can be found in [8,28].

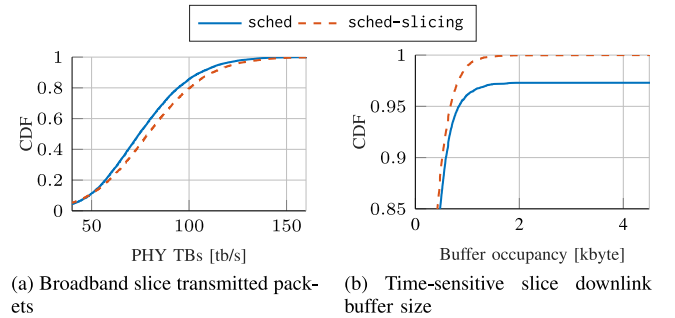


Fig. 5. Comparison of xApps developed with OpenRAN Gym.

Fig. 5 shows the Cumulative Distribution Function (CDF) of some RAN metrics measured at the base stations when the two xApps are instantiated on the CoIO-RAN near-RT RIC and used to control the RAN. Specifically, Fig. 5(a) shows the transmitted TBs for the broadband slice, while Fig. 5(b) displays the downlink buffer occupancy of the time-sensitive slice. By acting on a large action set (i.e., the slice resource allocation), the *sched-slicing* xApp achieves superior performance by delivering a higher number of transmitted packets and reducing the occupancy of the downlink buffer.⁹

⁹ A detailed evaluation of OpenRAN Gym xApps, including their orchestration, and control of large-scale experimental networks can be found in [8, 21].

Table 1
Compute node and radio setups used across the different testbeds.

Testbed	Compute node	Processor	CPU cores	RAM [GB]	Software-defined Radio
Base Station (BS)/UE					
Arena	Dell PowerEdge R340	Intel Xeon E-2146G	6	32	NI USRP X310
Colosseum	Dell PowerEdge R730	Intel Xeon E5-2650	48	128	NI USRP X310
COSMOS	Asus server (model undisclosed)	Intel i7-4790	4	16	NI USRP B210
POWDER (BS)	Dell PowerEdge R740	Intel Xeon Gold 6126	24	98	NI USRP X310
POWDER (UE)	Intel NUC 8559	Intel 7-8559U	4	32	NI USRP B210
Near-RT RIC					
Arena	Dell PowerEdge R340	Intel Xeon E-2146G	6	32	N/A
Colosseum	Dell PowerEdge R730	Intel Xeon E5-2650	48	128	N/A
COSMOS	Supermicro 1028U-TRT+	Intel Xeon E5-2698	16	251	N/A
POWDER	Dell PowerEdge R740	Intel Xeon Gold 6126	24	98	N/A

6. Traveling containers

In this section, we illustrate how the OpenRAN Gym containerized applications, including the xApps developed and pre-trained on Colosseum, can be transferred to other testbeds, and describe the necessary adjustments (if any) to run these applications on each experimental platform. Although the above procedure may seem trivial in the case of self-contained applications, in our case this is challenging due to the fact that our *traveling OpenRAN Gym containers* need to interact with the underlying network resources and be able to properly control the potentially diverse set of SDRs available in the different testbeds. In some cases, the firmware of the SDRs requires specific tools only available on certain operating systems versions or distributions. In such cases, the containers may need to be updated or rebuilt.

```
1 #!/bin/bash
2 lxc image import scope-with-e2.tar.gz --alias
  scope-e2
```

Listing 6: Commands to import the SCOPE LXC image with the E2 termination module.

To facilitate these tasks, we developed some tools to automatically start the OpenRAN Gym LXC containers on the different platforms considered in this work, and to properly interface them with the available radio resources. After the LXC images have been transferred (e.g., through the `scp` or `rsync` utilities) in a running instance of the testbed of interest, the image can be imported with the commands shown in Listing 6. These commands import the `scope-with-e2.tar.gz` LXC image transferred from Colosseum (i.e., the SCOPE image with the module for the E2 termination) to the compute machine of the remote testbed. After the above operation completes successfully, the new image, named `scope`, is visible by running the following command: `lxc image list`.

The LXC container can be, then, created from the imported image by running the commands shown in Listing 7. After creating the

```
1 #!/bin/bash
2 lxc init local:scope-e2 scope
```

Listing 7: Commands to create the SCOPE LXC container with the E2 termination module from the image imported in Listing 6.

container, additional operations may be required based on the specific OpenRAN Gym image, and SDR available in the remote testbed (e.g., USRP B210 or X310). As an example, if running the SCOPE container with an USRP B210, it is necessary to perform an *USB passthrough* operation to allow the container to use the USB interfaces and devices connected to the physical host (e.g., the USB interface to control the USRP). If using an USRP X310, instead, the container needs access to

the network interface the host machine uses to communicate with the SDR, to set the right Maximum Transmission Unit (MTU) for it, and possibly to flash the FPGA of the USRP with the appropriate image. In both these cases, the container may require some additional permissions to be able to use the passed devices and interfaces (e.g., read/write permissions on the USB devices).

In the case of CoLo-RAN—which can be imported and started with commands analogous to those shown in Listings 6 and 7—it is necessary to configure the Network Address Translation (NAT) rules of the host machine for it to forward the messages directed to the RIC to the CoLo-RAN container (e.g., the *E2 Setup Request* message used by the base station to subscribe to the RIC, and the *RIC Indication* messages used to send the KPMs to the xApps). Similarly to the previous case, the LXC container may require some additional permissions (e.g., to run the Docker containers of the CoLo-RAN near-RT RIC in a *nested* manner inside the CoLo-RAN LXC container).

```
1 #!/bin/bash
2 lxc start scope
```

Listing 8: Commands to start the SCOPE LXC container created in Listing 7.

After these operations have been executed, the LXC container (e.g., the SCOPE LXC container created in Listing 7) can be started with the commands of Listing 8. Now, the OpenRAN Gym applications can be executed by following the procedures detailed in Sections 3 and 4.

To simplify and automate the above setup operations, and to allow OpenRAN Gym users to swiftly configure and run the transferred containers, we developed and open-sourced a set of scripts that take care of (i) passing the right radio interface to the containers; (ii) giving them the required permissions; (iii) setting up the NAT rules of the host machine, and, finally, (iv) starting the OpenRAN Gym LXC containers from the imported images.¹⁰ These scripts, which are supposed to be run after the commands of Listing 6, i.e., after the LXC image has been imported, are described in Listings 9 and 10.

```
1 #!/bin/bash
2 ./start-lxc-scope.sh testbed usrp_type [flash]
```

Listing 9: Commands to start the SCOPE LXC container.

Specifically, Listing 9 creates, sets up, and starts the SCOPE LXC container starting from the image imported in Listing 6. After creating the container on the testbed of interest (i.e., `arena`, `powder`, or `cosmos`), the script configures the USRP specified through the

¹⁰ <https://github.com/wineslab/openrangym-pawr>

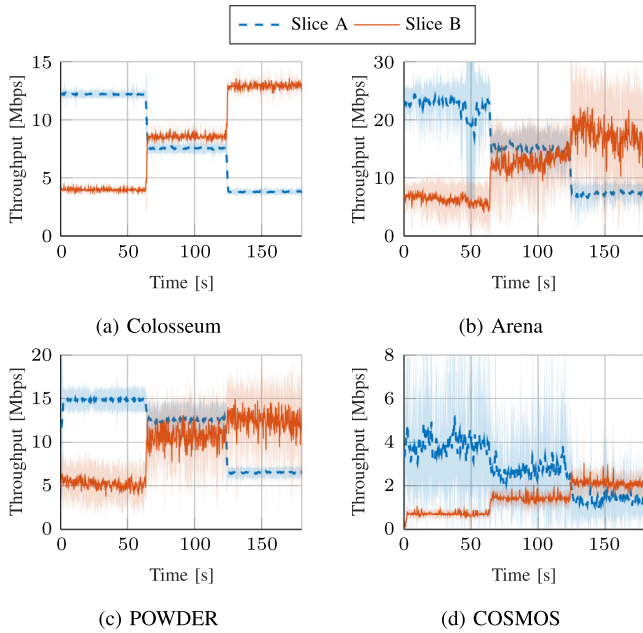


Fig. 6. Overall slice throughput varying the percentage of RBGs allocated to each slice over time according to the configuration reported in Table 2.

usrp_type parameter (i.e., b210 or x310) following the procedures described above (i.e., passing to the container the devices to interface with the USRP, and assigning the appropriate permissions to the container). The optional flash parameter also allows to flash the FPGA of the USRP X310 with the UHD image used by the container.¹¹

```
1 #!/bin/bash
2 ./start-lxc-ric.sh
```

Listing 10: Commands to start the CoLo-RAN LXC container.

The script of Listing 10 can be used to create, setup, and start the CoLo-RAN near-RT RIC container starting from the image imported with commands analogous to the ones shown in Listing 6. This script creates the CoLo-RAN near-RT RIC container, assigns it the required permissions (e.g., to run the nested Docker containers), and starts it. Then, it sets the NAT rules of the host machine (where the container is running) for it to forward the messages intended for the RIC. Finally, it builds and starts the Docker containers of the CoLo-RAN near-RT RIC (see Section 4) inside the created LXC container.

7. Experimental results

In this section, we showcase some experimental results obtained from running OpenRAN Gym and its components across a set of heterogeneous testbeds. We ported the SCOPE and CoLo-RAN near-RT RIC containers from Colosseum to the Arena, POWDER, and COSMOS testbeds (see Sections 2 and 6). A description of the setups used in these testbeds (also summarized in Table 1) follows. Since the capabilities offered by the different testbeds can be substantially different (e.g., number of available over-the-air nodes), for the sake of consistency, and to fairly compare results, we run experiments with

¹¹ Please note that after the FPGA has been flashed with a new image, the USRP may need to be rebooted. We refer to the documentation of the various testbeds for the instructions on how to achieve this.

Table 2

Slicing configuration, expressed as percentage of RBGs, used in Figs. 6 and 7.

Figure	Slice	First minute	Second minute	Third minute
Fig. 6	Slice A	75% RBGs	50% RBGs	25% RBGs
	Slice B	25% RBGs	50% RBGs	75% RBGs
Fig. 7	Slice A	75% RBGs	25% RBGs	75% RBGs
	Slice B	25% RBGs	75% RBGs	25% RBGs

one cellular base station and up to three UEs, and one near-RT RIC node. In all cases, we use a FDD configuration with 50 PRBs, corresponding to 10 MHz of bandwidth. We divide the spectrum of the base stations into up to three network slices, and statically assign the UEs to them (e.g., based on the Service Level Agreement (SLA) between users and their network operator). Downlink User Datagram Protocol (UDP) traffic generated through the iPerf3 tool is leveraged to evaluate the network performance. Finally, the base stations—implemented through SCOPE—connect to CoLo-RAN near-RT RIC through the E2 interface standardized by O-RAN.

POWDER. We instantiated both the CoLo-RAN near-RT RIC and the SCOPE base station on Dell PowerEdge R740 compute nodes, while the UEs were instantiated on Intel NUC 8559 nodes. The radio front-end of the base station was implemented through a USRP X310, while USRP B210 were used for the UEs. As this testbed does not natively support the LXC virtualization technology, the OpenRAN Gym container images were transferred from Colosseum to the compute nodes through the scp utility, instantiated on Ubuntu Linux images loaded on the bare-metal servers of the testbed.

COSMOS. In this case, the near-RT RIC was instantiated on a Supermicro 1028U-TRT+ server. Base station and UE, instead, were virtualized on Asus servers driving USRP B210 SDRs. Similarly to what done for POWDER, as the LXC virtualization technology is not directly supported by this testbed, the container images were transferred from Colosseum through the scp utility, and instantiated on Ubuntu Linux images loaded on the bare-metal nodes available on the testbed.

Arena. All applications were run on Dell PowerEdge R340 servers. In this case, the OpenRAN Gym LXC containers are instantiated directly on the bare-metal nodes of the testbed, which leverage USRP X310 SDRs as radio front-ends. On this testbed, the UEs are implemented through commercial smartphones.

Colosseum. To mimic the same deployment scenario used in the other testbeds, in Colosseum we considered cellular nodes deployed in a static RF scenario without user mobility. In this case the LXC containers of RIC, base station, and UEs directly run on Colosseum bare-metal Dell PowerEdge R730 servers. All the cellular nodes leverage USRP X310 SDRs as radio-front ends.

7.1. Results

To showcase the flexibility of OpenRAN Gym in dynamically reconfiguring the spectrum allocated to the network slices across different testbeds, Figs. 6 and 7 show the overall throughput of each network slice varying the resources allocated to them, in terms of RBGs. 95% confidence intervals are also represented by the shaded areas in the figures. For both figures, the percentage of RBGs allocated to each slice of the base station—which uses a 10 MHz configuration—is dynamically changed through the SCOPE APIs according to the following configuration (also summarized in Table 2).

In Fig. 6, the two network slices, i.e., slice A and B in the figure, are allocated the following RBGs percentage: (i) 75% to slice A and 25% to slice B in the first minute; (ii) 50% to each slice in the second minute, and (iii) 25% to slice A and 75% to slice B in the third minute. In Fig. 7, instead they are allocated the following RBGs percentage: (i) 75% to slice A and 25% to slice B in the first minute; (ii) 25% to slice A and 75% to slice B in the second minute, and (iii) 75% to slice A and 25% to

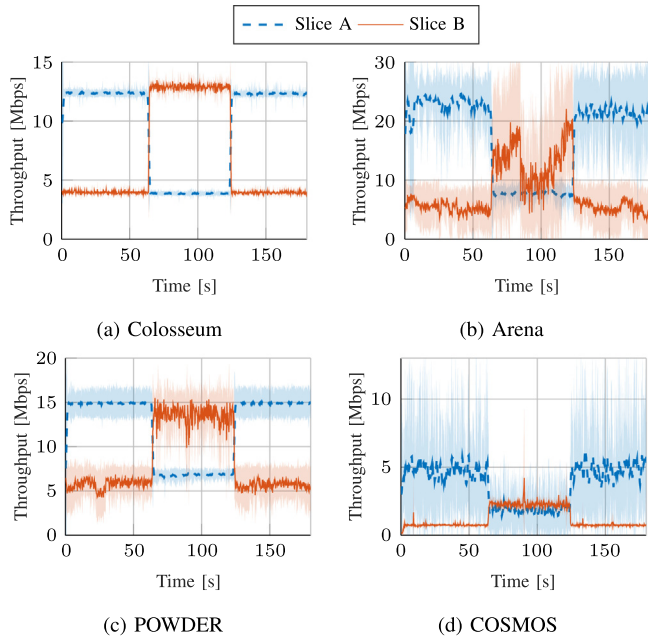


Fig. 7. Overall slice throughput varying the percentage of RBGs allocated to each slice over time according to the configuration reported in Table 2.

slice B in the third minute. In both these figures, the throughput varies proportionally to the specific allocation of slice resources, in which slices with more RBGs achieve higher throughput values. These values then change during the experiment as RBGs are dynamically reallocated to the slices. We notice that even if the throughput differs across the various testbeds because of the different capabilities and environments they offer—with Arena achieving the highest performance due to the use of commercial smartphones as the UEs—the overall trends are consistent across the different setups.

We now showcase an instance in which the CoLo-RAN near-RT RIC is leveraged to control a softwarized RAN implemented through SCOPE. LXC containers for both applications are deployed on the testbeds mentioned above, whose specifications are summarized in Table 1. Fig. 8 shows the evolution in time of the throughput of the three network slices (namely, slice A, B, and C) implemented by the SCOPE base station. Initially, we consider the baseline configuration where no control is performed by the RIC. Then, at around second 150, an xApp that prioritizes one of the network slices (slice A in the figure) is instantiated on the near-RT RIC. Following the O-RAN RAN slicing SM, this is done by reconfiguring at run time the number of RBGs that are assigned exclusively to each slice, and can be used to schedule transmissions for users belonging to that slice [29]. This, for example, allows slice A to transmit across many RBGs, as shown in Fig. 8. As a result, the xApp dynamically reallocates the amount of RBGs of each slice, which reflects on the performance of the slices of the RAN. Similar to the previous case, slices with a larger amount of RBGs allocated to them achieve higher throughput values. Overall, even in this case results are consistent across the different testbeds.

Now we show some timing statistics on the average amount of time taken to transfer the SCOPE and CoLo-RAN LXC images from Colosseum to the Arena, COSMOS, and POWDER platforms. All the image transfers were performed through the `scp` utility, while the LXC containers were created following the procedures detailed in Section 6. In both cases, these timing statistics were derived using the hardware of Table 1. We used the compute nodes listed in the “base station (BS)/UE” section of the table for the SCOPE LXC image/container (in the case of the POWDER platform, in which different compute nodes were listed

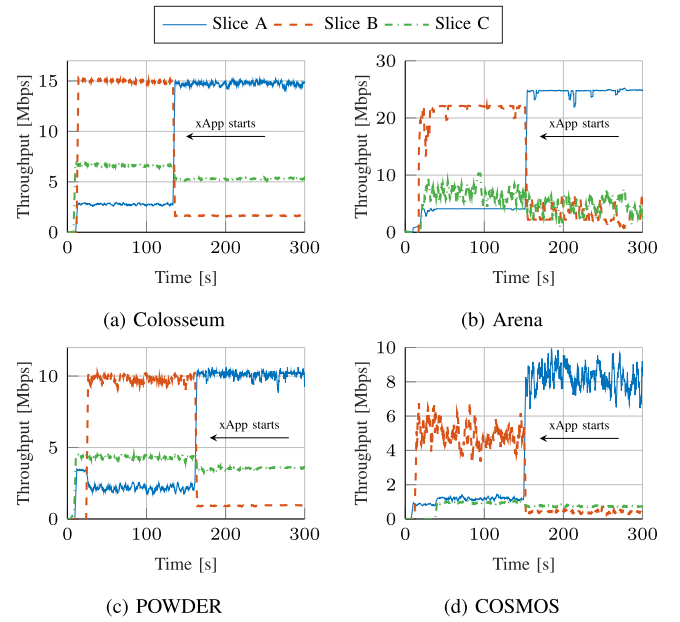


Fig. 8. Slice throughput when the SCOPE RAN is controlled by CoLo-RAN near-RT RIC. At around second 150, xApp to prioritize the amount of resources (i.e., RBGs) allocated to slice A is instantiated on the near-RT RIC.

Table 3

Average time to transfer the LXC images from Colosseum to specific testbeds. The size of each image is listed in brackets.

Testbed	SCOPE w/ E2 (1.7 GB)	CoLo-RAN near-RT RIC, prebuilt (6.5 GB)	CoLo-RAN near-RT RIC, to build (1.6 GB)
Arena	1 m 27.413 s	5 m 41.487 s	1 m 25.002 s
COSMOS	1 m 28.631 s	5 m 39.704 s	1 m 27.352 s
POWDER	1 m 30.787 s	5 m 43.704 s	1 m 28.546 s

Table 4

Average time to start as a container the LXC image exported from Colosseum on specific testbeds. The size of each image is listed in brackets.

Testbed	SCOPE w/ E2 (1.7 GB)	CoLo-RAN near-RT RIC, prebuilt (6.5 GB)	CoLo-RAN near-RT RIC, to build (1.6 GB)
Arena	0.887 s	1 m 11.483 s	46 m 18.110 s
COSMOS	25.463 s	2 m 34.905 s	26 m 4.410 s
POWDER	30.139 s	2 m 55.654 s	21 m 11.220 s

for base station and UE, the base station node was used), and the compute nodes in “near-RT RIC” section of the table for CoLo-RAN. In the tables that will be described next, we consider the following LXC images:

- *SCOPE w/ E2*: this is the SCOPE LXC image with the O-RAN E2 termination to interface with the near-RT RIC.
- *CoLo-RAN near-RT RIC, prebuilt*: this is the CoLo-RAN LXC image in which the Docker containers of the RIC, and sample xApp (described in Section 4) have been built a priori.
- *CoLo-RAN near-RT RIC, to build*: this is the CoLo-RAN LXC image with the scripts to build the Docker containers of the RIC and sample xApp from scratch.

Table 3 shows the average time required to transfer the LXC images from Colosseum to the other platforms. Times span from as low as ~ 1.5 min to as high as almost 6 min, depending on the size of each image—also listed in the table—and capabilities of the testbeds. However, transfer times are consistent across the different testbeds.

Finally, Table 4 shows the times taken to instantiate LXC containers from the images transferred from Colosseum. In this case, we notice

some difference among the times achieved on the different testbeds. For instance, Arena is significantly faster than COSMOS and POWDER in instantiating the SCOPE container—completing the instantiation in less than 1 s—and the prebuilt CoLO-RAN container (instantiation in approximately 1 minute). This is mainly due to the fact that Arena allows users to instantiate applications on the bare-metal nodes directly. This removes the latency of the extra virtualization layer of the other two testbeds, in which the LXC containers are nested inside the virtualized architecture the users are given access to. When it comes to building the Docker containers of the CoLO-RAN near-RT RIC (see Section 4) from scratch, instead, POWDER and COSMOS are significantly faster than Arena, taking approximately half the time to complete the same operations. This is mainly due to the superior compute capabilities of the nodes of these two testbeds (24-core CPU server on POWDER, and 16-core server on COSMOS vs. 6-core CPU server on Arena). Nonetheless, this building operation needs to be completed only once, as the compiled CoLO-RAN LXC image can be saved to be used in subsequent experiments, with instantiation times sensibly lower (slightly above 1 minute for Arena, and below 3 min for POWDER and COSMOS).

8. Conclusions

We presented OpenRAN Gym, the first publicly-available research platform for data-driven O-RAN experimentation at scale on heterogeneous wireless testbeds. Building on, and extending, frameworks for data collection and RAN control, OpenRAN Gym enables the end-to-end design and testing of data-driven xApps instantiated on the O-RAN infrastructure. We described the core components of OpenRAN Gym—including frameworks and experimental platforms—and detailed procedures and configuration options for experimenting at scale on a softwarized RAN instantiated on Colosseum. Then, we gave an overview of the xApp design and testing workflow enabled by OpenRAN Gym, also showcasing an example of two xApps used to control a large-scale O-RAN managed softwarized RAN deployed on Colosseum. Finally, we demonstrated how OpenRAN Gym solutions and experiments can be transitioned from Colosseum to heterogeneous real-world platforms, such as the Arena testbed, and the POWDER and COSMOS platforms of the PAWR program. OpenRAN Gym is publicly-available to the research community, and opened up for community contributions and additions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

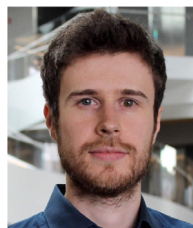
Data availability

Data is available at <https://openrangym.com>.

References

- [1] L. Bonati, M. Polese, S. D'Oro, S. Basagni, T. Melodia, OpenRAN Gym: An open toolbox for data collection and experimentation with AI in O-RAN, in: *Proceedings of IEEE WCNC Workshop on Open RAN Architecture for 5G Evolution and 6G*, Austin, TX, USA, 2022.
- [2] L. Bonati, M. Polese, S. D'Oro, S. Basagni, T. Melodia, Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead, *Comput. Netw.* 182 (2020) 1–28.
- [3] O-RAN Working Group 1, O-RAN Architecture Description 5.00, 2021, O-RAN.WG1.O-RAN-Architecture-Description-v05.00 Technical Specification.
- [4] M. Polese, L. Bonati, S. D'Oro, S. Basagni, T. Melodia, Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges, 2022, [arXiv:2202.01032](https://arxiv.org/abs/2202.01032) [cs.NI].
- [5] O-RAN Working Group 3, O-RAN near-RT RAN intelligent controller near-RT RIC architecture 2.00, 2021, O-RAN.WG3.RICARCH-v02.00.

- [6] O-RAN Working Group 2, O-RAN Non-RT RIC Architecture 1.0, 2021, O-RAN.WG2.Non-RT-RIC-ARCH-TS-v01.00 Technical Specification.
- [7] O-RAN Working Group 2, O-RAN AI/ML workflow description and requirements 1.03, 2021, O-RAN.WG2.AI/ML-v01.03 Technical Specification.
- [8] M. Polese, L. Bonati, S. D'Oro, S. Basagni, T. Melodia, CoLO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms, *IEEE Trans. Mob. Comput.* (2022) 1–14.
- [9] S. D'Oro, M. Polese, L. Bonati, H. Cheng, T. Melodia, dApps: Distributed applications for real-time inference and control in O-RAN, *IEEE Commun. Mag.* (2022) 1–7, in print; preprint available at <https://arxiv.org/pdf/2203.02370.pdf>.
- [10] L. Bonati, P. Johari, M. Polese, S. D'Oro, S. Mohanti, M. Tehrani-Moayyed, D. Villa, S. Shrivastava, C. Tassie, K. Yoder, A. Bagga, P. Patel, V. Petkov, M. Seltser, F. Restuccia, A. Gosain, K.R. Chowdhury, S. Basagni, T. Melodia, Colosseum: Large-scale wireless experimentation through hardware-in-the-loop network emulation, in: *Proceedings of IEEE DySPAN*, 2021.
- [11] L. Bonati, S. D'Oro, S. Basagni, T. Melodia, SCOPE: An open and softwarized prototyping platform for NextG systems, in: *Proceedings of ACM MobiSys*, 2021.
- [12] L. Bertizzolo, L. Bonati, E. Demirors, A. Al-Shawabka, S. D'Oro, F. Restuccia, T. Melodia, Arena: A 64-antenna SDR-based ceiling grid testing platform for sub-6 GHz 5G-and-beyond radio spectrum research, *Comput. Netw.* 181 (2020) 1–17.
- [13] Platforms for Advanced Wireless Research (PAWR). <https://www.advancedwireless.org>. (Accessed December 2021).
- [14] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, A. Ghosh, M. Hibler, D. Johnson, S.K. Kasera, E. Lewis, et al., POWDER: Platform for open wireless data-driven experimental research, *Comput. Netw.* 197 (2021) 1–18.
- [15] D. Raychaudhuri, I. Seskar, G. Zussman, T. Korakis, D. Kilper, T. Chen, J. Kolodziejki, M. Sherman, Z. Kostic, X. Gu, H. Krishnaswamy, S. Maheshwari, P. Skrimponis, C. Gutterman, Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless, in: *Proceedings of ACM MobiCom*, London, United Kingdom, 2020.
- [16] M. Dryjanski, L. Kulacz, A. Kliks, Toward modular and flexible open RAN implementations in 6G networks: Traffic steering use case and O-RAN xApps, *Sensors* 21 (24) (2021) 1–14.
- [17] D. Johnson, D. Maas, J. Van Der Merwe, Open source RAN slicing on POWDER: A top-to-bottom O-RAN use case, in: *Proceedings of ACM MobiSys*, 2021.
- [18] H. Lee, J. Cha, D. Kwon, M. Jeong, I. Park, Hosting AI/ML workflows on O-RAN RIC platform, in: *Proceedings of IEEE GLOBECOM Workshops*, 2020.
- [19] A.S. Abdalla, P.S. Upadhyaya, V.K. Shah, V. Marojevic, Toward next generation open radio access network—what O-RAN can and cannot do!, 2021, [arXiv:2111.13754](https://arxiv.org/abs/2111.13754) [cs.NI].
- [20] O-RAN Alliance Conducts First Global Plugfest to Foster Adoption of Open and Interoperable 5G Radio Access Networks. <https://tinyurl.com/f48auynf>. 2019.
- [21] S. D'Oro, L. Bonati, M. Polese, T. Melodia, Orchestran: Network automation through orchestrated intelligence in the open RAN, in: *Proceedings of IEEE INFOCOM*, 2022, [arXiv:2201.05632](https://arxiv.org/abs/2201.05632) [cs.NI].
- [22] I. Gomez-Miguel, A. Garcia-Saavedra, P. Sutton, P. Serrano, C. Cano, D. Leith, srsLTE: An open-source platform for LTE evolution and experimentation, in: *Proceedings of ACM WINTeCH*, 2016.
- [23] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, T.-T. Nguyen, OpenAirInterface: Democratizing innovation in the 5G era, *Comput. Netw.* (107284) (2020).
- [24] M. Kohli, T. Chen, M.B. Dastjerdi, J. Welles, I. Seskar, H. Krishnaswamy, G. Zussman, Open-access full-duplex wireless in the ORBIT and COSMOS testbeds, *Comput. Netw.* (2021).
- [25] L. Bonati, S. D'Oro, M. Polese, S. Basagni, T. Melodia, Intelligence and learning in O-RAN for data-driven NextG cellular networks, *IEEE Commun. Mag.* 59 (10) (2021) 21–27.
- [26] O-RAN Software Community. O-DU L2 Repository. <https://github.com/o-ran-sc/o-du-l2>. (Accessed December 2021).
- [27] D. Johnson, POWDER RIC Profile Repository. <https://gitlab.flux.utah.edu/johnson/ric-profile>. (Accessed December 2021).
- [28] L. Bonati, M. Polese, S. D'Oro, S. Basagni, T. Melodia, Intelligent Closed-loop RAN Control with xApps in OpenRAN Gym, in: *Proceedings of European Wireless 2022*, Dresden, Germany, 2022.
- [29] O-RAN Working Group 3, O-RAN near-real-time RAN intelligent controller E2 service model 2.00, 2021, O-RAN.WG3.E2SM-v02.00 Technical Specification.



Leonardo Bonati is an Associate Research Scientist at the Institute for the Wireless Internet of Things, Northeastern University, Boston, MA. He received the Ph.D. degree in Computer Engineering from Northeastern University in 2022. His main research focuses on softwarized approaches for the Open Radio Access Network (RAN) of the next generation of cellular networks, on O-RAN-managed networks, and on network automation and orchestration. He served multiple times on the technical program committee of the ACM Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization, and as guest editor of the special issue of Elsevier's *Computer Networks* journal on *Advances in Experimental Wireless Platforms and Systems*.



Michele Polese is a Principal Research Scientist at the Institute for the Wireless Internet of Things, Northeastern University, Boston, since March 2020. He received his Ph.D. at the Department of Information Engineering of the University of Padova in 2020. He also was an adjunct professor and postdoctoral researcher in 2019/2020 at the University of Padova, and a part-time lecturer in Fall 2020 and 2021 at Northeastern University. During his Ph.D., he visited New York University (NYU), AT&T Labs in Bedminster, NJ, and Northeastern University. His research interests are in the analysis and development of protocols and architectures for future generations of cellular networks (5G and beyond), in particular for millimeter-wave and terahertz networks, spectrum sharing and passive/active user coexistence, open RAN development, and the performance evaluation of end-to-end, complex networks. He has contributed to O-RAN technical specifications and submitted responses to multiple FCC and NTIA notice of inquiry and requests for comments, and is a member of the Committee on Radio Frequency Allocations of the American Meteorological Society (2022–2024). He collaborates and has collaborated with several academic and industrial research partners, including AT&T, Mavenir, NVIDIA, InterDigital, NYU, University of Aalborg, King's College, and NIST. He was awarded with several best paper awards, is serving as TPC co-chair for WNS3 2021–2022, as an Associate Technical Editor for the IEEE Communications Magazine, and has organized the Open 5G Forum in Fall 2021. He is a Member of the IEEE.



Salvatore D'Oro is a Research Assistant Professor at Northeastern University. He received his Ph.D. degree from the University of Catania in 2015. Salvatore is an area editor of Elsevier Computer Communications journal and serves on the Technical Program Committee (TPC) of multiple conferences and workshops such as IEEE INFOCOM, IEEE CCNC, IEEE ICC and IFIP Networking. Dr. D'Oro's research interests include optimization, artificial intelligence, security, network slicing and their applications to 5G networks and beyond. He is a Member of the IEEE.



Stefano Basagni is with the Institute for the Wireless Internet of Things and a professor at the ECE Department at Northeastern University, in Boston, MA. He holds a Ph.D. in electrical engineering from the University of Texas at Dallas (2001) and a Ph.D. in computer science from the University of Milano, Italy (1998). Dr. Basagni's current interests concern research and implementation aspects of mobile networks and wireless communications systems, wireless sensor networking for IoT (underwater, aerial and terrestrial), and definition and performance evaluation of network protocols. Dr. Basagni has published over twelve dozen of highly cited, refereed technical papers and book chapters. His h-index is currently 49 (November 2022). He is also co-editor of three books. Dr. Basagni served as a guest editor of multiple international ACM/IEEE, Wiley and Elsevier journals. He has been the TPC co-chair of international conferences. He is a distinguished scientist of the ACM, a senior member of the IEEE, and a member of CUR (Council for Undergraduate Education).



Tommaso Melodia is the William Lincoln Smith Chair Professor with the Department of Electrical and Computer Engineering at Northeastern University in Boston. He is also the Founding Director of the Institute for the Wireless Internet of Things and the Director of Research for the PAWR Project Office. He received his Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2007. He is a recipient of the National Science Foundation CAREER award. Prof. Melodia has served as Associate Editor of IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, Elsevier Computer Networks, among others. He has served as Technical Program Committee Chair for IEEE Infocom 2018, General Chair for IEEE SECON 2019, ACM Nanocom 2019, and ACM WUWnet 2014. Prof. Melodia is the Director of Research for the Platforms for Advanced Wireless Research (PAWR) Project Office, a \$100M public-private partnership to establish 4 city-scale platforms for wireless research to advance the US wireless ecosystem in years to come. Prof. Melodia's research on modeling, optimization, and experimental evaluation of Internet-of-Things and wireless networked systems has been funded by the National Science Foundation, the Air Force Research Laboratory the Office of Naval Research, DARPA, and the Army Research Laboratory. Prof. Melodia is a Fellow of the IEEE and a Senior Member of the ACM.