Toward Open Integrated Access and Backhaul with O-RAN

Eugenio Moro*, Gabriele Gemmi†, Michele Polese‡, Leonardo Maccari†, Antonio Capone*, Tommaso Melodia‡

*Department of Electronics, Information and Bioengineering,
Polytechnic University of Milan, Italy
{name.surname}@polimi.it
†Department of Environmental Sciences, Informatics and Statistics,
Ca' Foscari University of Venice, Italy.
{name.surname}@unive.it
‡Institute for the Wireless Internet of Things,
Northeastern University, Boston, MA, U.S.A.
{n.surname}@northeastern.edu

***Abstract*—Millimeter wave (mmWave) communications has been recently standardized for use in the fifth generation (5G) of cellular networks, fulfilling the promise of multi-gigabit mobile throughput of current and future mobile radio network generations. In this context, the network densification required to overcome the difficult mmWave propagation will result in increased deployment costs. Integrated Access and Backhaul (IAB) has been proposed as an effective mean of reducing densification costs by deploying a wireless mesh network of base stations, where backhaul and access transmissions share the same radio technology. However, IAB requires sophisticated control mechanisms to operate efficiently and address the increased complexity. The Open Radio Access Network (RAN) paradigm represents the ideal enabler of RAN intelligent control, but its current specifications are not compatible with IAB. In this work, we discuss the challenges of integrating IAB into the Open RAN ecosystem, detailing the required architectural extensions that will enable dynamic control of 5G IAB networks. We implement the proposed integrated architecture into the first publicly-available Open-RAN-enabled experimental framework, which allows prototyping and testing Open-RAN-based solutions over end-to-end 5G IAB networks. Finally, we validate the framework with both ideal and realistic deployment scenarios exploiting the large-scale testing capabilities of publicly available experimental platforms.**

***Index Terms*—IAB, O-RAN, 5G, Colosseum**

I. Introduction

Radio Access Network (RAN) densification is a key technique to boost the coverage and performance metrics of current and future generations of mobile radio networks [1]. However, these ultra-dense deployments come with increased costs and complexity for provisioning wired backhaul to each base station [2]. To address this, the 3rd Generation Partnership Project (3GPP) has introduced Integrated Access and Backhaul (IAB) in its Release 16 for NR [3]. With IAB, the backhaul traffic is multiplexed on the air interface together with regular User Equipments (UEs) access traffic. This effectively creates a wireless mesh network of Base Stations (BSs) where only a few require an expensive wired connection to the Core Network (CN) (i.e., the IAB-Donors). Hence the cost-reduction potential through wireless relays (i.e., the IAB-Nodes) [4]. Additionally, IAB is especially relevant for millimeter wave (mmWave)-based radio access, where inexpensive network densification is a fundamental necessity [5].

While the standardization process has reached a sufficient maturity level, the open challenges brought about by integrating access and backhaul are still open. Consequently, IAB offers optimization opportunities at all layers of communication abstraction. At the lowest levels, specialized IAB-aware techniques are required to ensure a fair and effective resource allocation among UEs and Mobile Terminations (MTs) [6, 7]. At the same time, backhaul and access transmission multiplexing must be managed to minimize interference [8]. Furthermore, adaptive topology reconfiguration mechanisms must be provisioned to maintain resiliency against link failures, traffic unbalances and anomalous user distribution [9]. Overall, these sophisticated management procedures require control primitives that go beyond what has been specified by 3GPP.

The unprecedented paradigm shift brought about by the O-RAN architecture, developed by the O-RAN Alliance, promises to enable programmatic control of RAN components through open interfaces and centralized control loops [10]. As such, it is the ideal candidate to unlock the potential optimization and management gains awaiting in IAB. However, the current O-RAN architecture is tailored to traditional RAN deployments, and an extension to enable IAB control

This work was partially supported by NGIAtlantic.eu project within the EUHorizon 2020 programme under Grant No. 871582, by the U.S. National Science Foundation under Grant CNS-1925601, and by OUSD(R&E) through Army Research Laboratory Cooperative Agreement Number 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.

is required. The first contribution of this work resides in a discussion on how the O-RAN architecture, interfaces, and control loops can be extended to IAB scenarios, with the ultimate goal of allowing large-scale, data-driven control and management of 5th generation (5G) IAB networks.

Additionally, to foster prototyping and testing with IAB and O-RAN, we propose a comprehensive framework where researchers can easily deploy an end-to-end O-RAN-enabled IAB network with Over-The-Air (OTA) and hardware-in-the-loop emulation capabilities. In line with O-RAN core concepts, our framework is designed to be open, accessible and flexible by leveraging on open-source software and Commercial Off-the-Shelf (COTS) hardware. The framework builds on IABEST, the first large-scale accessible and open IAB testbed presented in [11]. This testbed has been enriched to produce a complete O-RAN IAB experimental solution, effectively replicating the proposed O-RAN-IAB integrated architecture. In particular, IAB-Donors and IAB-Nodes have been equipped with custom-developed agents for the so-called E2 and O1 standard interfaces. These additions enable the controllers introduced by the O-RAN architecture to manage IAB-Nodes, effectively representing the first publicly available O-RAN-enabled IAB prototyping and testing solution.

To further facilitate experimental research activities, we have packaged and integrated the entire framework into OpenRAN Gym, a publicly-available research platform for data-driven O-RAN experimentation at scale [12]. Through OpenRAN Gym, researchers can swiftly deploy and test the proposed framework over large-scale and publicly available hardware experimental platforms, such as the PAWR testbeds and Colosseum [13, 14]. Notably, we showcase how Colosseum can be leveraged for large-scale IAB testing through hardware-in-the-loop channel emulation to create sophisticated deployment scenarios. A tutorial on how to deploy an O-RAN-driven IAB network, together with the source code of all the framework, is available on the OpenRAN Gym website.[1] Finally, we use Colosseum to validate the proposed framework numerically. In particular, we test the attainable performance in a controlled radio scenario and in a more realistic deployment in which we reconstruct a part of Florence, Italy.

The remainder of this paper is organized as follows. Section II analyses the challenges of extending O-RAN to 5G IAB networks. Section III contains a description of the proposed frameworks, focusing on the O-RAN extensions that have been included in [11]. Section IV contains the results of the experiments we performed to validate our framework by exploiting the large-scale testing capabilities of Colosseum. Finally, Section V concludes the paper and discusses future extensions.

II. Integrating IAB in Open RAN

As discussed in Section I, IAB represents a scalable solution to the need for backhaul in ultra-dense 5G and 6G deployments. At the same time, however, the wireless backhaul introduces additional complexity to the network deployments: new parameters and configurations that need to be tuned—and possibly, adapted dynamically—to get the best performance out of the network and to seamlessly adjust to updated conditions in the scenario and in the equipment status. For example, it is possible to optimize the IAB network performance by properly selecting the connectivity of IAB-Nodes to their parents [9], or by appropriately allocating resources to backhaul and access flows sharing the same air interface [6].

As for traditional RAN deployments with fiber-based backhaul [15], there is a case to be made for providing IAB RAN equipment with primitives for flexible, dynamic, data-driven programmatic control. This requires providing endpoints to expose telemetry, measurements, and analytics from IAB-Nodes, as well as parameters and control knobs to enable the optimization. So far, the Open RAN paradigm has been successfully applied to non-IAB networks to achieve the same goals, thanks to interfaces that give access to 3GPP Key Performance Measurements (KPMs) and control parameters in the RAN nodes [16, 17]. The Open RAN vision, which is being developed into technical specifications by the O-RAN Alliance, includes controllers that run custom control loops, i.e., the RAN Intelligent Controllers (RICs). The O-RAN Alliance has defined control loops and related RICs that can operate at a time scale of 10 ms to 1 s (i.e., *near-real-time*) or more than 1 s (i.e., *non-real-time*) [18]. The near-real-time, or near-RT, RIC is connected to the RAN nodes through the E2 interface, while the non-real-time RIC, which is part of the network Service Management and Orchestration (SMO), interacts with the RAN through the O1 interface, as shown in the left part of Figure 1. Other interfaces from the non-RT RIC/SMO include A1 to the near-RT RIC, for policy guidance and Artificial Intelligence (AI)/Machine Learning (ML) model management, and the O2 interface to the O-Cloud, which is an abstraction of the virtualization infrastructure that can support the deployment of O-RAN functions. The use of standard interfaces makes it possible to run even third-party applications in the controllers, the so-called *xApps* and *rApps* for the Near Real-time RAN Intelligent Controller (near-RT RIC) and Non-Real-Time Ran Intelligent Controller (non-RT RIC), respectively.

The 3GPP already provides control and adaptation capabilities through the IAB Backhaul Adaptation Protocol (BAP) layer, the F1 interface, and the Radio Resource Control (RRC) layer across the IAB-Donor Central Unit (CU) and the IAB-Node Distributed Unit (DU). How and when control and adaptation of such configurations could be performed, however, is left to the vendor implementation. This is where an extension of the O-RAN architecture to IAB networks can play a role, exposing IAB-Donor and IAB-Node functions to the RICs. These can leverage a centralized point of view on the RAN and a wealth of analytics and information usually unavailable in the individual IAB-Donors and Nodes. For IAB, this could translate into effective multi-donor coordination with reduced interference and agile topology adaptation across different IAB-Donor domains, and dynamic resource allocation with—

[1] https://openrangym.com/tutorials/iab-tutorial

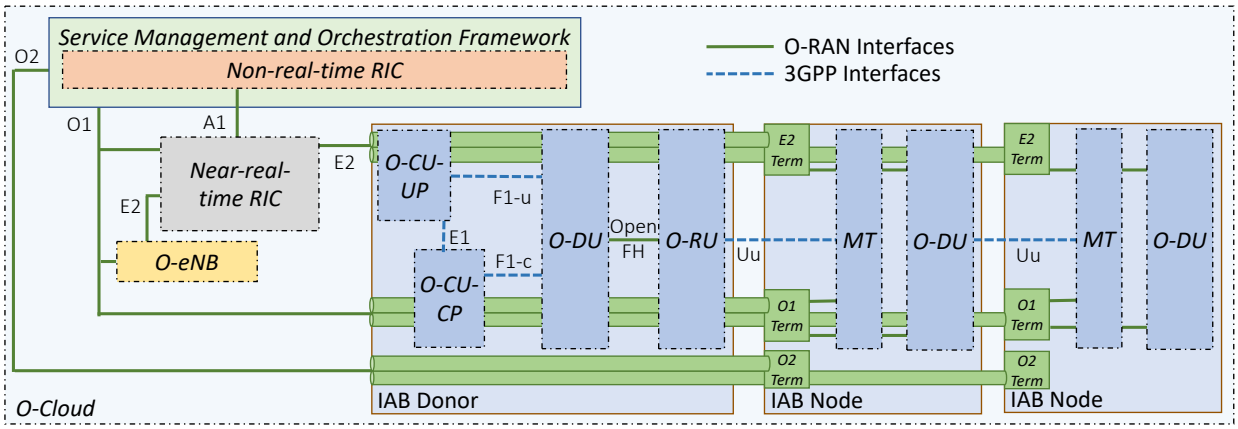


Fig. 1: IAB and O-RAN integrated architectures.

for example—data-driven proactive congestion identification and resolution across access and backhaul links.

A. Extensions to Open RAN

Extending the O-RAN architecture and interfaces to IAB deployments, however, presents some design and architectural challenges. Primarily, supporting O-RAN interfaces in IAB-Nodes means either (i) terminating the interfaces at the IAB-Donor; or (ii) transporting their data over the wireless backhaul. The first option is simpler, does not require architectural updates, but at the same time limits the control and reconfiguration to what is available in the IAB-Donor, without insight on the IAB-Nodes. The second option, instead, provides more granular access at the cost of additional complexity and tunneling of data over the wireless backhaul.

The 3GPP already foresees performing SMO-like operations through the wireless backhaul interface [19]. Therefore, in this paper and in the architecture described in Figure 1 we consider the second option, which would provide tighter and more effective integration between O-RAN and IAB deployments. In general, the tunneling can be performed by encapsulating the O-RAN interfaces payloads into dedicated bearers. Note that this requires some interaction between functions of the control plane of the network and the transport in the user plane, e.g., through a dedicated Packet Data Unit (PDU) session between a local User Plane Function (UPF) in the IAB-Donor and in the IAB-Node MT. Then, a local interface termination can be installed in the IAB-Node, as it would in a traditional, fiber-equipped RAN node. The O-RAN traffic, in this case, would be multiplexed with user data on the wireless backhaul resources, and it needs to be properly prioritized to achieve the control goals while not harming users' performance or creating congestion.

E2 extension for IAB. The extension of the E2 interface likely requires one or multiple new, dedicated E2 Service Models (E2SMs). The E2SM represents the semantic of the E2 interface, i.e., the RAN function with which an xApp in the near-RT RIC interacts. For IAB, an extension of E2SM KPM [20] can be used to expose performance metrics related to the MT, besides the DU. Another near-real-time control target over E2 can include, for example, resource partitioning between backhaul and access traffic, or dynamic Time Division Duplexing (TDD) slot configuration to adapt to varying traffic on the access and backhaul.

O1 extension for IAB. The O1 interface would connect the SMO to the IAB-Node, e.g., to perform maintenance and updates of the components (MT and DU) of the IAB-Node. Compared to E2 near-real-time control, the O1 interface would run control loops at 1 s or more. Thus its traffic can be transported with lower priority than the E2 traffic. This makes a case for dedicated bearers and tunnels on the backhaul interface for *each* of the O-RAN interfaces.

O2 extension for IAB. This interface can be used to integrate the IAB-Nodes as resources in the O-Cloud. Compared to traditional virtualization infrastructure for the O-Cloud, the IAB-Nodes are available—and reachable over O2—only when a session is established from one IAB-Donor to the IAB-Node itself.

III. An Experimental Framework for IAB and O-RAN

Our proposed experimental framework packages the entire software chain required to run the O-RAN-enabled IAB network described in Section II in a multi-layer architecture. At the hardware level, our framework does not present any specific requirement. Indeed, every software component can run on COTS hardware like generic x86 machines and USRP Software-defined Radio (SDR). On the other hand, some software components are customized or designed from scratch to reproduce and support a 5G IAB network. In particular, we have adapted OpenAirInterface (OAI), an open source 5G RAN framework [21], to implement IAB-Donors, IAB-Nodes, and IAB-capable core functions. Additionally, we have integrated agents for the E2 and O1 interfaces in the IAB-Donor and IAB-Node, effectively implementing the architectural integration proposed in Section II. These interfaces are used by the non-real-time and real-time RICs packaged in our framework to control all the components of the deployed IAB

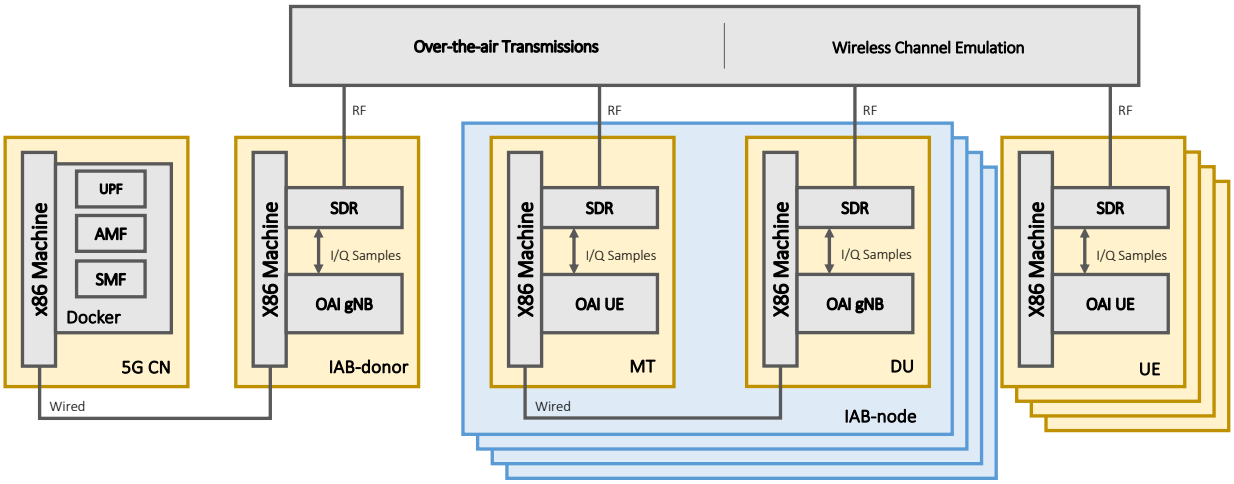


Fig. 2: Overview of the RAN architecture deployed over white-box hardware.

network. We now describe the aforementioned components, separating them into the RAN and O-RAN domains.

A. RAN and Core Network Components

Figure 2 represents an overview of the radio access functional components that enable end-to-end communication in our framework. In particular, we provide the following: a minimal yet functional deployment of 5G CN functions, software-defined IAB-Nodes and IAB-Donors and software-defined UEs.

IAB-Nodes and IAB-Donors. According to 3GPP specifications [4], an IAB-Donor hosts a CU and multiple DUs. Similarly, IAB-Node is split into a DU and an MT. Functionally, these have the task of enabling downstream and upstream connectivity, respectively. At the time of writing, OAI's main branch implementation of the CU/DU functional split does not support multiple DUs connected to a single CU [22]. This limitation is incompatible with the IAB architecture. Consequently, we employ a full OAI Next Generation Node Base (gNB) in place of both CU and DU. In other words, the IAB-Nodes and IAB-Donors in our framework do not follow 3GPP split 2. Instead, these components are deployed as monolithic gNBs. As for the MT, an open-source implementation is currently unavailable. However, this component is functionally equivalent to a UE, as it connects to upstream nodes using the same resources and protocols. Consequently, we have selected OAI's software-defined UE to act as MTs in the proposed framework. This results in a system where a single IAB-Node is made up of two concurrently running instances: an OAI gNB—acting as a DU—and an OAI UE—acting as a MT. In the resulting architecture, IAB-Nodes are naturally deployed over two separate machines, hosting the gNB and the UE, and connected out-of-band as it is shown in Figure 2. Alternatively, the two software components can run on a single x86 machine, provided that sufficient computing power is available. While this architecture does not require any particular modification to OAI's implementations, we have added a signaling functionality through which the IAB-Nodes or IAB-Donors can discern connected MTs from connected UEs. This has been achieved through proper manipulation of the UE Capability messages. Such information can be exploited to enable IAB-aware optimization solutions in the gNB.

Core Network Functions. A minimal set of 5G CN functions have been included in our framework: Network Repository Function (NRF), Access and Mobility Management Function (AMF), Slicing Magangement Framework (SMF) and User Plane Function (UPF), all based on the OAI 5G core implementation. All these functions run as containers on a single x86 machine, as shown in Figure 2. Due to the selected IAB system design, the UPF required modifications to enable IAB operations. As previously mentioned, UEs acts as MTs in IAB-Nodes, connecting to upstream nodes. The established GPRS Tunneling Protocol (GTP) tunnels are then used to provide direct connectivity between the DU of the node and the CN functions. In other words, MT-acting UEs relay the backhaul traffic of the IAB-Nodes. However, OAI's UPF implementation lacks support for the required forwarding capability,[2] as any packet whose destination is not a UE is dropped. Therefore, we have implemented a minimal version of framed routing [23] in OAI UPF, enabling UEs to act as intermediate nodes.

User Equipment From the perspective of the UE, an IAB network deployed using the components described above is entirely standard-compliant. As such, both software-defined UEs (as shown in Figure 2) and COTS UEs can be used in the proposed framework.

B. O-RAN Components

As mentioned in Section II, O-RAN defines a set of standardized and open interfaces with which the RAN exposes data collection and control primitives to the RICs. In the proposed framework, we have enabled IAB-Nodes and IAB-Donors to be O-RAN-compatible by integrating software agents for the

[2]To the best of the authors' knowledge, there is no available open source implementation that supports this operating mode.

E2 and O1 interfaces into the codebase of OAI. Furthermore, our framework comprises a near-RT RIC and a non-RT RIC.

E2 interface integration. The E2 interface is functionally split into two protocols: E2AP—tasked with establishing a connection with the near-RT RIC—and E2SM—which implements specific monitoring and control functionalities, namely Service Models (SMs), as discussed in Section II. In the software implementation we provide, E2AP has been adapted from O-RAN Alliance Software Community reference implementation and, as such, it is entirely compliant with O-RAN. On the other hand, the SMs provided by the O-RAN alliance are defined using ASN.1: a powerful production-ready abstract description language which is, however, cumbersome and challenging to use in the fast-paced research and development environments targeted by our framework. In light of this, we employ custom SM that are defined through Protocol Buffers (protobuf)—an abstract definition language that is easier to handle and allows for fast prototyping and testing, facilitating the development of IAB-aware control solutions. Since the E2 interface is such that the E2SM messages are encoded and decoded only in the RAN and xApp, the custom SM definitions are transparent to the RIC, allowing our proposed solution to retain generic O-RAN compliance. At the time of this writing, we have implemented a set of protobuf messages that can be used to reproduce both the KPM and RAN Control (RC) SMs [10]. These can be used to develop data collection and control xApps, respectively.

O1 interface integration. In order to properly manage all the different aspects of networked elements, the O1 interface defines various Management Services (MnS), which can be used either from the managed entities (the gNBs) to report information back to the RIC or from the managing entity (the SMO and the rApps running on it) to deploy configurations changes, transfer files or update the software on the managed entities [10, 24]. Among all the different MnS, we have focused our contribution on implementing the Heartbeat MnS, which periodically transmits heartbeats; the Fault Supervision MnS, which reports errors and events; and the Performance Assurance MnS, which streams performance data. Those MnS have been integrated into the OAI codebase by implementing a scheduler that, running on a dedicated thread, periodically sends Virtual Network Function (VNF) Event Stream (VES) notifications in JSON format over HTTP. This format and communication protocol has been chosen among the different options defined in the standard, as it is widely known and easily extendable by other researchers. As of now, our implementation reports performance metrics, such as the throughput and information on the channel quality between IAB-Nodes, and failure events, such as RRC or Uplink Shared Channel (UL-SCH) failures, which can be used in rApps to monitor and optimize the backhaul network. Provisioning MnS, which can be used by the rApps to deploy configuration changes (e.g., topology optimizations), have not been implemented by following the O1 specifications, as it would have needed major reworks in the OAI codebase. Instead, we have taken advantage of *IAB-Manager*, a software component we developed to orchestrate IAB experiments, as discussed next.

Parameter	Value
Area Size for realistic deployment	0.627 km^2
gNB Density	45 gNB/km^2
IAB-donors/ IAB-nodes ratio	1/10
Emulated center frequency	28 GHz
Bandwidth	40 MHz
Scheduler	7 2 1
Subcarrier Spacing	30khz
Colosseum Base loss	50 dB
3GPP Channel Model	Urban Micro
MIMO layers	1

TABLE I: Table of System Settings

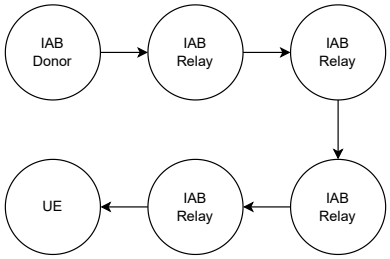
IAB-Manager. IAB networks are generally expected to include several IAB-Nodes, and the proposed framework can scale to such numbers. However, managing experiments with tens or more RAN components can take time and effort. Indeed, each component is potentially hosted by a dedicated machine, and setting up an IAB deployment requires each one to be activated and configured according to a sequence that starts from the CN functions and ends with the terminal IAB-Nodes. To facilitate experimenting at such a large scale, we have developed *IAB-Manager* [11]: a software component that can automate the IAB network deployment and testing through a command line interface and an Application Programming Interface (API). In particular, *IAB-Manager* is a single entrypoint for controlling the entire experiment: network components and radio environment setup (in case of wireless channel emulation), topology and routing management and reconfiguration, automated testing, and result collection. From a functional perspective, the manager connects to the machines involved in the experimentation and configures them according to the assigned roles. In particular, once the user specifies the roles, the manager sequentially activates each network component until the final deployment is ready for experimentation, greatly simplifying the setup phase. Additionally, as previously mentioned, *IAB-Manager* executes the network configuration changes mandated by the rApps.

RAN Intelligent Controllers. The proposed framework packages a near-RT RIC and a non-RT RIC. Both are compliant with the standard and based on O-RAN Software Community reference implementations.

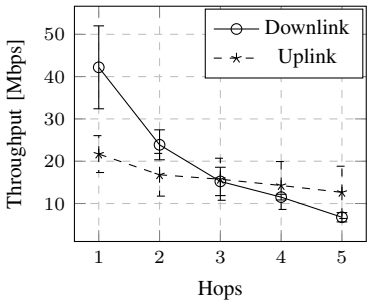
IV. Validation and Results

This section focuses on validating our proposed framework from an experimental perspective. In particular, we are interested in giving an initial characterization of some fundamental Key Performance Indicators (KPIs) of the deployments allowed by our IAB framework while validating its correct functioning.

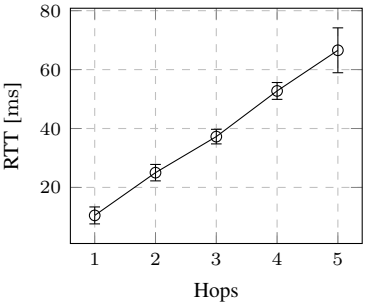
While the openness and flexibility of the software components are such that the framework can run on generic hardware, we chose to base our validation campaign on Colosseum [14]. The Colosseum testbed is a publicly available large-scale testing platform with hardware-in-the-loop capabilities. It



(a) Linear IAB topology.



(b) Throughput measurements for the linear topology.



(c) Round Trip Time (RTT) measures for the linear topology.

Fig. 3: Topology and results for the linear chain.

comprises 128 Standard Radio Nodes (SRNs), each composed of a powerful x86 computing node and an USRP X310 SDR. All the components in the proposed framework can be easily deployed on SRNs. Every SRN radio is interconnected by an FPGA mesh that emulates arbitrary radio channels defined through tapered delay models. With the capability of emulating complex scenarios of tens of entities, Colosseum makes it possible to deploy large IAB networks over complex propagation scenarios. As such, it represents an ideal validation platform for our framework. Furthermore, Colosseum is open to the research community, and the validation tools are made available, allowing interested parties to start experimenting with a minimal initial effort.

A. Experiments with a linear chain

We start by evaluating the performance of an IAB network deployed in a tightly controlled scenario. To this end, we consider a 5-hop linear topology, as shown in Figure 3a. As detailed in Section II, each IAB-Node comprises an MT and a DU, bringing this experiment's overall radio node count to 10. In order to characterize the upper-bound performance of the proposed framework, we employ an ideal propagation scenario. Through properly manipulating Colosseum's channel emulator, a 0 dB pathloss model is selected for nodes connected in the linear topology, and an infinite pathloss is set for all the other channels, effectively suppressing any possible interference. In other words, this radio scenario is equivalent to connecting the SDRs with coaxial cables.[3] Transmissions occur on band n78 with 106 Physical Resource Blocks (PRBs) available, for a total of 40MHz bandwidth.

Figure 3b shows the downlink and uplink TCP throughput against the number of hops, as measured between the core network and the specific MT/UE. The first-hop values of 47 Mbps in DL and 21 Mbps in UL represent the maximum throughput attainable in the testing settings. This upper bound is far from the theoretical maximum allowed by the available bandwidth. It is limited by several factors that depend on the experimental platform, OAI software implementation, and system design. Most notably, the strongest detractor to the final throughput performance is given by the OAI implementation of the software-defined UE, which is employed to build the MT. In particular, the OAI UE is inefficient in reception and transmission, thus becoming a bottleneck for the entire communication chain. Efforts are ongoing to improve the performance and stability of this software framework. Furthermore, the frameworks' system design is such that each IP packet is encapsulated into as many GTP packets as the number of hops. This increased overhead can cause packet fragmentation with a further negative impact on the overall performance. Furthermore, even if the emulated channel is set to a 0 dB pathloss, Colosseum's architecture includes an unavoidable base loss of 50 dB [25] due to characteristics of the hardware architecture. This, together with the aforementioned inefficiencies, make such that packet drops and subsequent retransmissions happen also in this ideal scenario.

As the number of hops increase, the downlink throughput experiences a sharp decrease before stabilizing on a per-hop loss of around 6 Mbps. The notable throughput loss experienced at the second hop can be explained by observing the standard deviation of the throughput, represented by the whiskers in Figure 3b. This value is at its maximum for the first hop, suggesting that the first radio link is unstable due to the RX pipeline of the MT being overwhelmed. This substantial variability is caused by packet loss and retransmissions and internal buffer overflow, which negatively affect the performance of the second hop, as it is noticeable in the numerical results. At the same time, the second hop's throughput standard deviation is lower, as the decreased traffic volume causes less drops in the involved MTs. This stabilizing effect propagates down the topology, as both the decreasing standard deviation and the linear per-hop loss testify. On the other hand, the uplink throughput is relatively stable and close to the upper bound, even at the fourth hop. This is because the limited OAI UE performance and BS scheduling process limits the uplink traffic volume, and the gNBs are far from being overwhelmed. On the other hand, since the uplink throughput does not significantly decrease from the maximum, the UE's congestion level remains relatively stable and high, as proven by the constant standard deviation values.

RTT is measured when the network is unloaded, that is

[3]

Fig. 4: Realistic deployment scenario in Florence, Italy. Donors are represented in red, while IAB-Nodes are represented in yellow.

when there is no traffic flowing through the IAB network. As shown in Figure 3c, the first hop latency is around 11 ms. This value represents the base processing delay plus a small fixed propagation delay that is, however, the same for each hop. As the number of hops increases, the RTT experiences a linear increase comparable with the first hop latency, as expected. This shows how the system does not introduce any spurious latency when the network is unloaded. Finally, the relatively higher RTT standard deviation of the last hop (as represented by the whiskers in Figure 3c) suggests that multiple packet retransmissions are required.

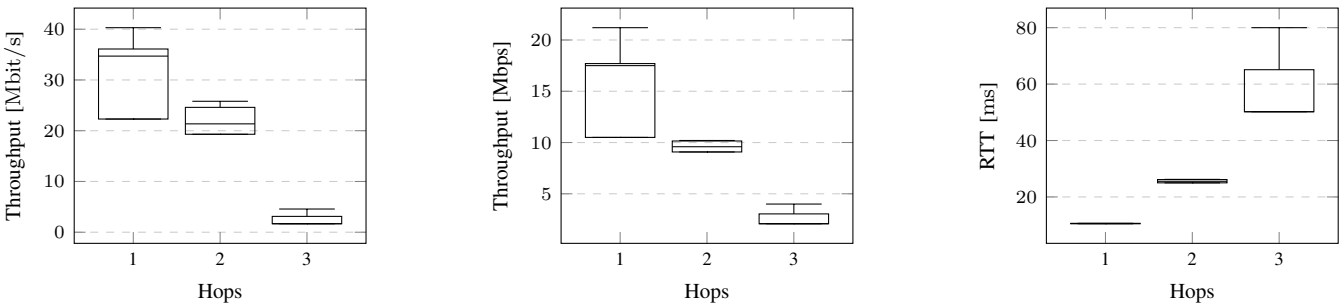
B. Validation over realistic RF scenarios

After having validated the system performance in a controlled environment, we move to more realistic urban scenarios, representing the typical deployment environment of an IAB network. We start by selecting a densely populated area of a city, from which we extract a precise 3D model of the environment. On top of this, we apply a coverage planning heuristic to find the optimal locations for the IAB-Nodes [26]. We then take advantage of a viewshed algorithm implemented on GPU and the above-mentioned 3D models to evaluate the Line of Sight (LoS) between each pair of locations and to produce the so-called visibility graph [27]. Then, we characterize the propagation according to the 3GPP Urban Micro parametric model [28], and we produce a tapered-delay representation of the communication channels, which Colosseum can then emulate. This process has been carried out for several European cities and four different scenarios are made available.[4]

Motivated by the fact that IAB is unanimously considered as a key enabler of mmWave RAN [29], we are interested in providing an experimental solution that enables testing in such conditions. While Colosseum is not directly capable of operating at frequencies higher than 6 GHz, we can approximate these radio scenarios by reproducing the most relevant propagation characteristics of mmWaves, namely the extremely directive transmissions through beamforming and the increased pathloss [30]. In particular, the pathloss between nodes that are not directly connected in the provided topologies has been set to infinite. The resulting suppression of inter-node interference might appear too ideal at first. However, this is compatible with the highly directive transmissions typical of mmWave, where interference in static conditions (i.e., as in a backhaul IAB topology) can be practically neglected [31]. A more refined mmWave channel emulation will be subject of future extensions. In addition, since Colosseum's channel emulation happens in base-band, we can apply arbitrary pathloss independently of the radio frequency employed during the experiments. Thanks to this flexibility, we could compute pathloss for a carrier frequency of 28 GHz and apply them to LoS links. Nonetheless, the scenarios made available to the Colosseum community are available for both 3.6 GHz and 28 GHz, both with and without inter-node interference suppression.

For the experimental evaluation presented in this work, we

[4]https://colosseumneu.freshdesk.com/support/solutions/articles/61000303373-integrated-access-and-backhaul-scenarios



(a) Downlink. (b) Uplink. (c) Round Trip Time.

Fig. 5: Measurements for the realistic scenario.

have selected a scenario based on the city of Florence, Italy. Figure 4 shows both the urban layout and the IAB deployment, which is extended over $0.7\ \mathrm{km}^2$ and comprises 21 nodes (3 of which are IAB-Donors). To determine which nodes are going to become IAB-Donors, we have applied the group closeness centrality metric [32] to the visibility graph. This centrality metric selects k nodes such that their distance to all the other nodes is minimized. Then, we have determined the IAB topology as a Shortest-Path Forest computed over the visibility graph of the area with the well-known Dijktra's Algorithm. Similar to what has been done for the previous analysis, we characterize the throughput and latency at each hop in the network. In this case, however, the different link lengths cause performance variations in the per-hop throughput and latency. As such, we employ box plots to synthetically describe the network performance statistics in Figure 5. In particular, consider Figure 5a. Here the bottom and top edges of each box represent the first and third quartile of the downlink throughput measurements taken at all the different hops in the scenario. Similarly, the central marks indicate the median, and the whiskers represent the extreme data points. The plotted values indicate how the realistic pathloss introduced in the study scenario causes lower performance than the ideal case previously analyzed, independently of the considered hop. The same can be noted for the uplink throughput, as shown in Figure 5b. In both cases, the decreasing per-hop throughput trend is conserved. However, the throughput variability is the same for the two transmission directions. This is because, as opposed to the ideal scenario, the link length now represents the main performance-determining factor. This is testified by the significant distance between the first and third quartile of the first hop in both downlink and uplink throughput, which is consistent with the high variance of the first hop length in the topology of study. As for the second and third hop, the relatively closer quartiles are motivated by lower link length variations for these hops in the considered topology. Finally, the upper whiskers represent the performance of the shortest links, giving a further characterization of the system performance in this realistic scenario.

Figure 5c shows the RTT statistic through the same plotting technique. Differently from the throughput, the latency is not affected by the link length variations in the considered scenario for the first two hops. Additionally, the RTT increase at hops 1 and 2 is consistent with the one experienced in the controlled scenario. On the other hand, the high RTT variance of the third and last hop suggests a high probability of requiring retransmissions along the IAB path.

V. Conclusions

In this work, we have discussed the motivations and challenges of integrating IAB with O-RAN. On this matter, we have proposed possible architecture extensions that enable dynamic control and data collection over 5G IAB networks through O-RAN intelligent controllers. We have implemented the proposed integrated architecture and packaged it into the first publicly available experimental framework enabling at-scale testing and prototyping of O-RAN-based solutions applied to IAB networks. The system comprises all the software components required to establish end-to-end connectivity, plus custom-developed E2 and O1 agents that allow software-defined IAB-Nodes to be O-RAN-compliant. The framework is designed to be open and accessible and can be deployed over COTS hardware. We numerically validated the framework exploiting the large-scale testing capabilities of Colosseum, showing the system's performance over both an ideal linear topology and more sophisticated realistic deployments. Finally, the framework has been packaged and released into OpenRAN Gym and is available to the research community.

References

[1] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[2] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, 2014.

[3] T. Inoue, "5G NR Release 16 and Millimeter Wave Integrated Access and Backhaul," in *2020 IEEE Radio and Wireless Symposium (RWS)*, 2020.

[4] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 62–68, 2020.

[5] W. Feng, Y. Wang, D. Lin, N. Ge, J. Lu, and S. Li, "When mmWave Communications Meet Network Densification: A Scalable Interference Coordination Perspective," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1459–1471, 2017.

[6] Y. Zhang, V. Ramamurthi, Z. Huang, and D. Ghosal, "Co-optimizing performance and fairness using weighted pf scheduling and iab-aware flow control," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[7] B. Zhang, F. Devoti, I. Filippini, and D. De Donno, "Resource allocation in mmWave 5G IAB networks: a reinforcement learning approach based on column generation," *Computer Networks*, vol. 196, p. 108248, 2021.

[8] "Coordinated parallel resource allocation for integrated access and backhaul networks," *Computer Networks*, vol. 222, p. 109533, 2023.

[9] S. Ranjan, P. Chaporkar, P. Jha, and A. Karandikar, "Backhaul-Aware Cell Selection Policies in 5G IAB Networks," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021.

[10] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, pp. 1–1, 2023.

[11] E. Moro, M. Polese, I. Filippini, S. Basagni, A. Capone, and T. Melodia, "IABEST: An Integrated Access and Backhaul 5G Testbed for Large-Scale Experimentation," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, ser. MobiCom '22. Sydney, NSW, Australia: Association for Computing Machinery, 2022.

[12] 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," *Computer Networks*, vol. 220, pp. 1–11, January 2023.

[13] A. Gosain, "Platforms for Advanced Wireless Research: Helping Define a New Edge Computing Paradigm," in *Proceedings of the 2018 on Technologies for the Wireless Edge Workshop*, ser. WirelessEdge '18. New Delhi, India: Association for Computing Machinery, 2018.

[14] L. Bonati, P. Johari, M. Polese, S. D'Oro, S. Mohanti, M. Tehrani-Moayyed, D. Villa, S. Shrivastava, C. Tassie, K. Yoder *et al.*, "Colosseum: Large-scale wireless experimentation through hardware-in-the-loop network emulation," in *2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2021.

[15] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent O-RAN for Beyond 5G and 6G Wireless Networks," in *2022 IEEE Globecom Workshops (GC Wkshps)*, 2022.

[16] H. Lee, J. Cha, D. Kwon, M. Jeong, and I. Park, "Hosting AI/ML Workflows on O-RAN RIC Platform," in *2020 IEEE Globecom Workshops (GC Wkshps)*, 2020.
[17] B. Brik, K. Boutiba, and A. Ksentini, "Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 228–250, 2022.
[18] 3GPP, "NR; NR and NG-RAN Overall description; Stage-2," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.300, 09 2022, version 17.2.0.
[19] ——, "NG-RAN; Architecture description," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.401, 09 2022, version 17.2.0.
[20] O.-R. WG3, "O-RAN Near-Real-time RAN Intelligent Controller E2 Service Model (E2SM) KPM 2.0," Open-RAN Alliance, Technical Specification (TS), 7 2021, version 2.0.0.
[21] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, and T.-T. Nguyen, "OpenAirInterface: Democratizing innovation in the 5G Era," *Computer Networks*, vol. 176, p. 107284, 2020.
[22] (2021) F1 interface - Data plane status. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/-/wikis/f1-interface
[23] 3GPP, "Interface between the Control Plane and the User Plane Nodes," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 29.144, 6 2022, version 16.4.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3111
[24] O-RAN Working Group 10, "O-RAN Operations and Maintenance Interface Specification 8.0," O-RAN.WG10.O1-Interface.0-v08.00 Technical Specification, Oct 2022.
[25] 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 *Proceedings of the 16th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH '22. Sydney, NSW, Australia: Association for Computing Machinery, 2022.
[26] G. Gemmi, R. Lo Cigno, and L. Maccari, "On Cost-Effective, Reliable Coverage for LoS Communications in Urban Areas," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2767–2779, 2022.
[27] G. Gemmi, R. L. Cigno, and L. Maccari, "On the Properties of Next Generation Wireless Backhaul," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 1, pp. 166–177, 2023.
[28] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.901, 01 2020, version 16.1.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173
[29] M. Cudak, A. Ghosh, A. Ghosh, and J. Andrews, "Integrated Access and Backhaul: A Key Enabler for 5G Millimeter-Wave Deployments," *IEEE Communications Magazine*, vol. 59, no. 4, pp. 88–94, 2021.
[30] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE communications surveys & tutorials*, vol. 18, no. 2, pp. 949–973, 2015.
[31] P. Fiore, E. Moro, I. Filippini, A. Capone, and D. D. Donno, "Boosting 5G mm-Wave IAB Reliability with Reconfigurable Intelligent Surfaces," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022.
[32] A. Bavelas, "Communication patterns in task-oriented groups," *The journal of the acoustical society of America*, vol. 22, no. 6, pp. 725–730, 1950.