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 5th generation (5G), 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 to prototype and test Open-RAN-based solutions over end-to-end, over-the-air 5G IAB networks. Finally, we validate the framework with 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

Ultra-dense deployments of next-generation 5G and 6G networks come with increased costs and complexity for provisioning of wired backhaul to each base station. To address this, the 3rd Generation Partnership Project (3GPP) has introduced Integrated Access and Backhaul (IAB) in its Release 16 for NR []. With IAB, the backhaul traffic is multiplexed on the air interface with access traffic of regular User Equipments (UEs), enabling wireless backhaul and nodes that can be deployed with reduced cost [] and increased flexibility (e.g., for on-demand networking) []. In an IAB deployment, the base station connected to the wired backhaul is called IAB donor, while the others are IAB nodes. Each IAB node include a Mobile Termination (MT), which acts as a UE and connects to the upstream Distributed Unit (DU), and a DU to provide downstream connectivity to other UEs and IAB nodes. A dedicated layer in the IAB nodes, the Backhaul Adaptation Protocol (BAP) layer, provides routing and forwarding functionalities to IAB nodes [].

While the standardization process has reached a sufficient maturity level, the open challenges brought about by the integration of 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 MTs. At the same time, backhaul and access transmission multiplexing must be managed in order to minimize interference. Furthermore, adaptive topology reconfiguration mechanisms must be provisioned to maintain resiliency against link failures, traffic unbalances and anomalous user distribution. 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 Open Radio Access Network (RAN) architecture, developed by the O-RAN Alliance, promises to enable programmatic control of RAN components through open interfaces and centralized control loops. As such, it represents the ideal candidate to unlock the aforementioned optimization and management gains for IAB. However, the current O-RAN architecture is tailored to traditional RAN deployments and an extension to

---

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.

enable IAB control 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) capabilities. In line with O-RAN's core concepts, our framework is designed to be open, accessible and flexible by exploiting 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 [1]. 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, donors and nodes has been equipped with custom-developed agents for the E2 and O1 interfaces. These additions enable nodes to be controlled by O-RAN's Near Real-time RAN Intelligent Controller (Near-RT RIC) and Non-Real-Time Ran Intelligent Controller (Non-RT RIC), 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 [**?**]. Through OpenRAN Gym, researchers can swiftly deploy and test the proposed framework over large-scale and publicly available hardware experimental platforms, such as PAWR and Colosseum. Particularly, we showcase how Colosseum can be exploited for large-scale IAB testing by exploiting its hardware-in-the-loop channel emulation capabilities that allows to recreate complex radio-frequency scenarios. Finally, we exploit Colosseum to produce a numerical validation of the proposed framework. In particular, we test the attainable performance both in a controlled radio scenario and in a more realistic deployment based on the city 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 [1]. Section IV contains a numerical analysis campaign where the proposed framework has been validated by exploiting the large-scale testing capabilities of Colosseum. Finally, Section V concludes the paper and discusses future extensions.

II. EXTENDING OPEN RAN TO SUPPORT IAB

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 [], or by properly allocating resources to backhaul and access flows sharing the same air interface [].

As for traditional RAN deployments with fiber-based backhaul [], 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 []. The Open RAN vision, which is being developed into technical specifications by the O-RAN Alliance, includes external controllers that run custom control loops based on (possibly third-party) applications, 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*) []. 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 3GPP already provides control and adaptation capabilities through the IAB BAP layer, the F1 interface, and Radio Resource Control (RRC) layer across the IAB donor Central Unit (CU) and IAB node 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 exploit a centralized point of view on the RAN and a wealth of analytics and information that is usually not available in the individual IAB donors and nodes. For IAB, this could translate into effective multi-donor coordination with reduction of interference and agile topology adaptation across different IAB donor domains, and dynamic resource allocation with—for example—data-driven proactive congestion identification and resolution across access and backhaul links.

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.

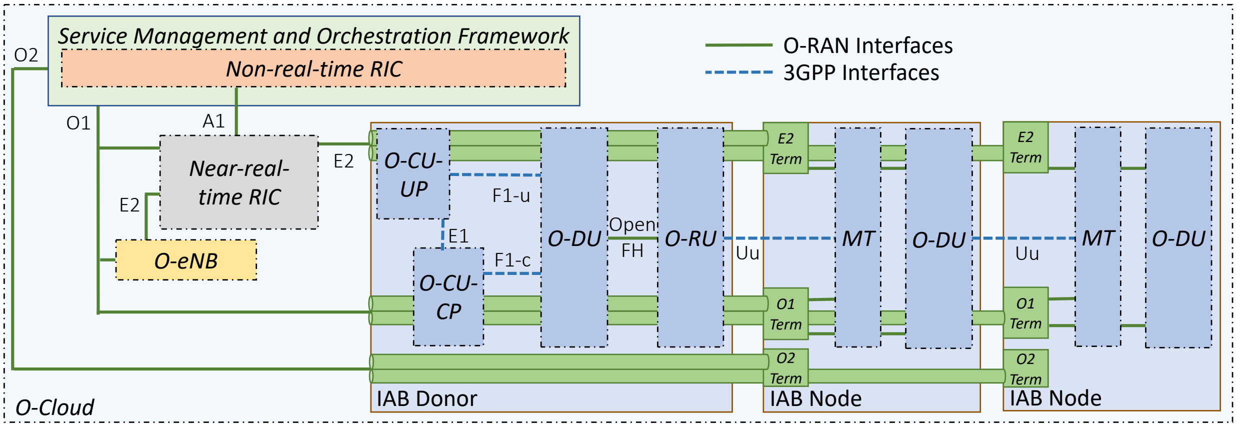


Fig. 1: IAB and O-RAN integrated architectures.

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 [**?**]. Therefore, in this paper and in the architecture described in Figure 1 we consider the second option, which would provide a 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 donor and the IAB node MT. Then, a local termination of the interface 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 [] can be used to expose performance metrics related to the MT, besides the DU. Other 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 maintainance 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 compared to the E2 traffic. This makes the 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 hardware-software chain required to run an O-RAN-enabled IAB network in a multi-layer architecture, as shown in Figure 2. At the hardware level, our framework does not present any specific requirement. Indeed, every software component can run on COTS hardware, such as generic x86 machines and USRP Software-defined Radio (SDR). Some software components are, on the other hand, either 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, to produce IAB donors, nodes and IAB-capable core functions. Additionally, we have integrated agents for the E2 and O1 interfaces in the IAB-donor and IAB-node. These interfaces are exploited by the non-real-time and real-time RICs packaged in our framework to control all the components of the deployed IAB network. In this section, we give a detailed description of the aforementioned components. Additionally, we describe how our framework can exploit the automation and scaling capabilities of large publicly-available testbeds, such as Colosseum.

A. Reproducing IAB on existing RAN Frameworks

According to the 3GPP [2], an IAB-donor hosts a CU and multiple DUs. An IAB-node is split into two functional blocks, i.e., the DU, which offers connectivity to downstream IAB-nodes and UEs, and the MT, through which the node connects

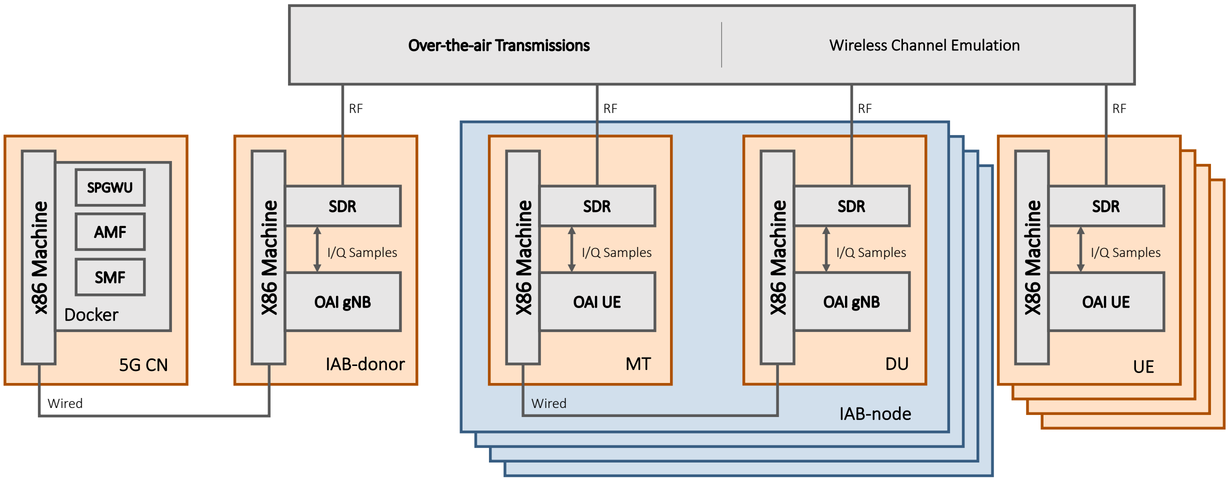


Fig. 2: Overview of the framework architecture

upstream. As OAI's implementation of the CU/DU functional split is not ready for large scale testing [3], our IAB testbed employs a full OAI's gNB in place of both CU and DU, as shown in Figure 2. An OAI UE acts as MT in each IAB-node, connecting upstream and establishing a GPRS Tunneling Protocol (GTP) tunnel for the IAB-node's DU to reach the 5G Core (5GC). This architecture requires UEs to work as intermediate nodes, which is not supported by GTP. Therefore, we have implemented a minimal version of framed routing [4] in the OAI's 5GC SPGWU packet gateway function.

B. *Integration of O-RAN E2 and O1 interfaces*

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 donors to be O-RAN-compatible by integrating software agents for the E2 and O1 interfaces into the codebase of OAI.

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). In the software implementation we provide, E2AP has been adapted from O-RAN Alliance's 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 environment where our framework is employed. Indeed, ASN.1 definitions require considerable effort during the description phase, and they are not easily manipulated when compiled for a specific language and integrated into the codebase. For example, the ASN.1 definitions for KPM given by O-RAN span across 459 lines of abstract syntax [**?**]. When compiled for C++, these results in 482 `.h` and `.c` files. In light of these considerations, we employ custom SM that are still defined through an abstract definition language but are easier to handle and allow for fast prototyping and testing. In particular, we base our SMs on Protocol Buffers: an abstract description language and a serialization tool that allows to define generic data structures (messages), which are then compiled for a variety of programming languages. 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 E2 Agent to retain generic O-RAN compliance, while easing the testing phase of control solutions.

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 Next Generation Node Bases (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 [5] [6]. 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 glsoai codebase by implementing a scheduler that, running on a dedicated thread, periodically sends VES notifications in JSON format over HTTP. This format and communication protocol have 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 the channel state information between IAB nodes and failure events such as RRC or ULSCH 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, such as 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 the experiments and testing that will be detailed in the following section.

C. Deploying and managing large-scale experiments

In general, IAB networks are expected to include a numerous amount of IAB nodes and UEs and the proposed framework is capable of scaling to such numbers. However, managing experiments with tens or more RAN components can be tedious and time consuming. 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 Core Network (CN) functions and ends with the terminal IAB nodes. To facilitate experimenting at such a large scale, we have developed *iab-manager* [**?**]: 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 entry-point 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 has been designed to connect to all the machines involved in the experimentation and configure them according to roles that has been assigned to each. In particular, once the machine list and roles are specified by the user, the manager sequentially activates each network component until that the final deployment is ready for experimentation, greatly simplifying the setup phase.

IV. VALIDATION AND RESULTS

In this section, we validate our proposed framework by testing it on Colosseum. Colosseum is a large-scale channel emulator with hardware-in-the-loop capabilities. It comprises 128 Standard Radio Nodes (SRNs), each one composed by a powerful compute node equipped with an USRP X310 SDR. Every SRN is interconnected by an FPGA mesh that emulates arbitrary radio channels defined through tapered delay models. In our packaged solution, we make available a linear topology, where all channels are ideal, and three realistic deployment scenarios where the radio channels recreate the urban settings of three different cities. The numerical validation presented in this section has been carried out over the linear topology and one of the realistic scenarios.

A. Experiments with a linear chain

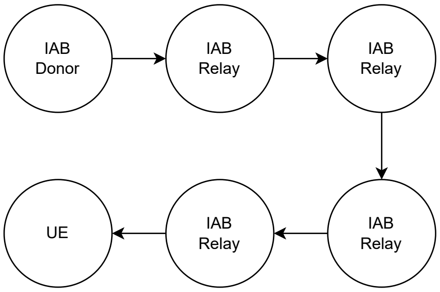


Fig. 3: Linear IAB topology

In order to validate the proposed IAB implementation we first ran a set of experiments on a linear topology composed of one IAB donor, four IAB relay, and one UE. This allowed us to determine how the topological distance from the donor affected KPIs such as the latency and the capacity of the network. In this scenario, the channels are ideal: a 0dB pathloss is selected for nodes that are connected in the linear topology, and an infinite pathloss is set for all the other channels, effectively avoiding interference[1]. For this scenario, a carrier frequency of 3.6GHz has been selected, with a total bandwidth of 40MHz.

Figure 5 shows both the downlink and uplink TCP throughput against the number of hops. The values of 35Mbps and 8Mbps at the first hop represents the maximum throughput attainable in the testing settings. This upper bound is far from the theoretical maximum allowed by the available bandwidth. In this case, the bottleneck is caused by the limited performance of OAI's UE, whose implementation is inefficient. As the number of hops increase, the downlink throughput experiences a linear decrease of around 8Mbps per-hop. The performance drop is expected, and it is due to the congested UE RX pipeline, which results in both data and ACK packets being lost at each hop. 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 causes packet fragmentation with a further negative impact on the overall performance. 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 transmission capabilities of OAI's UE are such that the gNBs are far from being overwhelmed. As such, the TCP rate is correctly adjusted according to the terminal UE TX buffer and all the ACK packets can successfully reach the destination.

As for the Round Trip Time (RTT) shown in Figure 4, the first hop latency is around 11ms. This value represents the processing delay, plus a small fixed propagation delay that is, however, the same for each hop. As the number of hops increase, the RTT experiences a linear increase that is comparable with the first hop latency, as expected. This shows how the system does not introduce any spurious latency when the network is unloaded.

B. Validation over realistic RF scenarios

To further validate our IAB implementation, and to provide a complete experimental environment to other researchers interested in investigating IAB networks, we have devised three realistic scenarios focused on IAB network backhauls. Since the generation of realistic scenarios requires the availability of highly precise 3D models of the environment, we started

---

[1] It must be noted that, as shown in a recent analysis, Colosseum's architecture includes an unavoidable base loss of 50dB [7], independently of the selected channel characteristics.

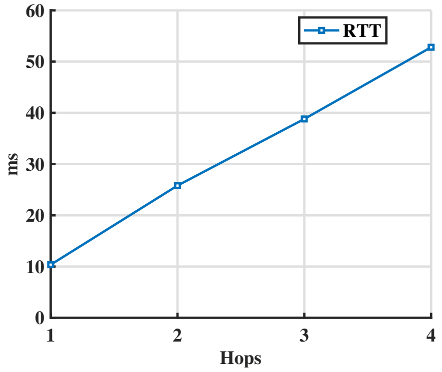


Fig. 4: RTT measures for the linear topology.

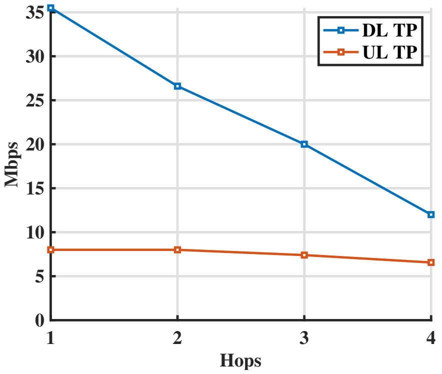


Fig. 5: Throughput measures for the linear topology.

by selecting three large cities in Europe. For each city, we extracted a densely urbanized area in the city center onto which we applied a gNB placement heuristic [8] to find the optimal locations for the IAB nodes. We have then uniformly and randomly placed UEs on the public space. Finally, taking advantage of a raytracing propagation tool we have characterized the RF channel between each pair of devices (both UEs and IAB nodes). Tab. I details the characteristics of each area and Fig. 6 shows a map with the buildings, the gNB, and the links in Line of Sight (LoS).

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 the 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 it 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 it is available to the the research community.

Area	Size (km^2)	# gNB	Coordinates	Frequency
Barcelona	0.77	23	2.168 E, 41.389 N	28GHz
Florence	0.63	20	11.232 E, 43.786 N	28GHz
Luxemburg	1.10	35	6.115 E, 49.605 N	28GHz

TABLE I: Charactercs of the three different scenarios.

REFERENCES

[1] 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. New York, NY, USA: Association for Computing Machinery, 2022, p. 772–774. [Online]. Available: https://doi.org/10.1145/3495243.3558750

[2] 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.

[3] (2021) F1 interface - data plane status. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/-/wikis/f1-interface

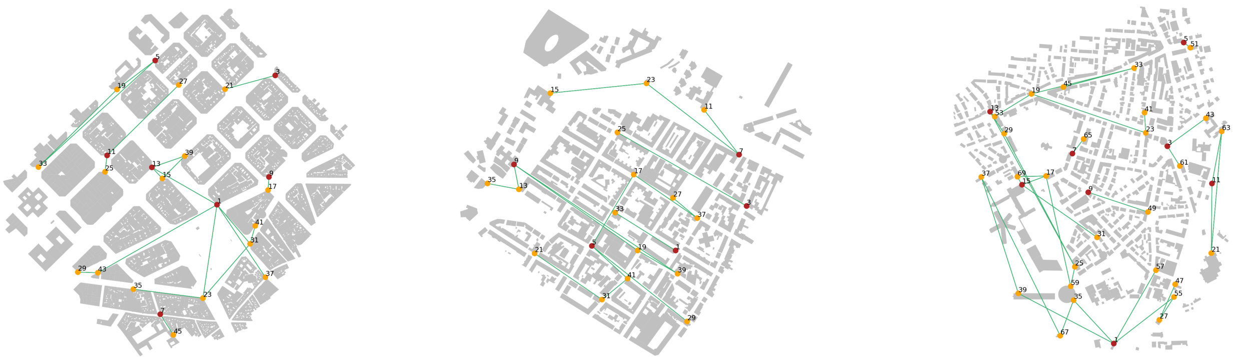
[4] 3GPP, "Interface between the control plane and the user plane nodes," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 29.144, 6 2022, version 16.4.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3111

[5] 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.

[6] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," 2022. [Online]. Available: https://arxiv.org/abs/2202.01032

[7] 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. New York, NY, USA: Association for Computing Machinery, 2022, p. 45–52. [Online]. Available: https://doi.org/10.1145/3556564.3558243

[8] 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.



(a) Barcelona (b) Florence (c) Luxemburg

Fig. 6: Map and topology of the three different scenarios.