

AI-Powered Data Synthesis for Advanced Simulation in 5G/6G mmWave Integrated Access and Backhaul Networks

Amir Ashtari Gargari*, Farhad Rezazadeh*, Marco Giordani†, Sandra Lagén*, Lingjia Liu‡, Andra Lutu§, Michele Zorzi†

*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Barcelona, Spain,
email: name.surname@cttc.es

‡Wireless Research Group, Virginia Tech, Blacksburg, USA, email: ljliu@vt.edu

§Telefónica Research, Madrid, Spain, email: andra.lutu@telefonica.com

†Department of Information Engineering, University of Padova, Italy, email: name.surname@dei.unipd.it

Abstract—Integrated Access and Backhaul (IAB) is a cost-effective and adaptable solution for the deployment of ultra-dense next-generation (5G and 6G) cellular networks to increase the likelihood of Line-of-Sight (LOS) coverage. This technology allows wireless backhaul connections to be established using the same technology and specifications as available in the access links. However, the absence of a physical testbed or a dataset that can be used for simulation in the millimeter wave (mmWave) band prevents researchers' validation of the proposed algorithms in the IAB scenario. In this paper, we propose a novel data generator based on Generative Adversarial Network (GAN), trained on a real dataset from a mobile network that operates in Europe, and maintains a significant market share that returns accurate traffic data for an IAB network. Furthermore, we integrate this data generator with the SeBaSi simulator (an IAB simulator based on Sionna) which permits to obtain accurate, data-consistent, realistic, and end-to-end IAB simulation results. The performance results indicate that the data generator successfully passes the Kolmogorov–Smirnov (KS) criterion, so it could operate as a verified data generator. Furthermore, we use the SeBaSi simulator, integrated with the data generator, to evaluate the performance of an IAB network in the London City scenario.

Index Terms—6G, GAN, IAB, Self-backhauling, Wireless Backhaul, GAN, Data Generation

I. INTRODUCTION

The significant increase in data rate capacity at sub-terahertz (THz) and millimeter wave (mmWave) frequencies that enable data-hungry use cases such as Extended Reality (XR) and mobile metaverse applications in 5th generation (5G) and 6th generation (6G) cellular systems. To mitigate the effects of the severe propagation environment at higher frequencies, wireless networks will be deployed with an exceptionally high density to increase the probability of Line-of-Sight (LOS) coverage. The 3rd Generation Partnership Project (3GPP) has standardized an extension of 5G New Radio (NR), known as Integrated Access and Backhaul (IAB) [1], to make ultra-dense deployments more feasible and economically sustainable. This extension reduces the required number of fiber drops by utilizing the same waveform and

protocol stack to provide wireless backhaul for Next Generation Node Bases (gNBs), which are the IAB nodes [2].

IAB is a wireless backhauling technique initially suggested in 3GPP Release 16 [3]. By enabling self-backhauling, it provides a practical solution to the challenges encountered by dense mobile networks. This technique employs a chain structure in which numerous IAB nodes are ultimately linked to an IAB donor, combining the resources of access [4] and backhaul links in the gNB (an example of this network is shown in Fig. 1). The 3GPP NR Release 18 [1] currently includes a Work Item (WI) that is dedicated to the examination of architectures, radio protocols, and the physical layer for IAB. The objective is to facilitate the sharing of radio resources between access and backhaul links. This WI on IAB anticipates a more sophisticated and adaptable solution, featuring dynamic resource multiplexing, multihop communications, and a plug-and-play design for low-complexity deployments.

Numerous research papers have investigated the development of IAB networks for both backhaul and access links, focusing especially on the identification of optimal solutions for resource management [5], [6]. Additionally, IAB can utilize a significantly greater bandwidth at mmWaves than in legacy sub-6 GHz spectrum, and the inherent directionality at this frequencies mitigates the interference of concurrent access and backhaul transmissions. Nevertheless, the design of a high-performance IAB network remains an open research challenge, despite the consensus regarding IAB's capacity to reduce deployment and management costs. Multi-Hub (MH) backhauling, an important aspect of IAB technology, has the potential to enhance network throughput and coverage [7], [8]. In order to effectively address the blockage issue and interference management in an IAB network, it is crucial to leverage spatial reuse through MH backhauling [9], [10]. So, the IAB network design and in particular backhaul network is an important and complicated problem because it involves managing the network's topology, choosing routes, and adjusting the resources that are shared between backhaul

and access links [11]. The MH-IAB network deployment and configuration are flexible, which can present additional challenges. Consequently, it is important to implement an efficient network design for the access and backhaul connections of MH-IAB, which should consider the interference generated by in-band wireless backhaul and blockage issues.

Prior to actual deployment, it is essential to validate all suggested methods for IAB, including IAB nodes and donor placement and backhaul schedulers [12]. In this sense, given the absence of publicly available testbeds or experimental setups, network simulation is a more reasonable choice for the performance evaluation of IAB networks. Among other solutions, the *ns3-mmwave-IAB*¹ [10] module is built on top of the ns-3 simulator, and can be used to implement, design, dimension and evaluate end-to-end IAB networks. However, the module had not been upgraded to support the newest 5G-NR standard specifications, and was incapable of handling the simulation of large-scale network deployments. Recently, we introduced a simulator called Self-Backhauling-Simulator (SeBaSi) [13], which is publicly available² and has the ability to replicate an IAB network with much stronger connections to the 3GPP standard than the current version of the *ns3-mmwave-IAB* module. Nevertheless, the design of SeBaSi has not been trained or validated with real data from IAB network deployments, which could lead to discrepancies between simulation results and real traffic patterns and deployments. Because the number of real datasets is extremely limited, it may be difficult to model a large number of scenarios and to train enough generality [14]. Therefore, it is crucial to learn from the real and available datasets, and replicate similar data in more different and heterogeneous scenarios.

In this context, GAN has become a popular technique to create datasets starting from real data, including data from wireless communication networks [15]. Generative Adversarial Networks (GANs) is a unique type of Deep Neural Network (DNN) that can produce data by acquiring the precise statistical characteristics of a given dataset through indirect methods. In this research, we propose a new real dataset generator based on GAN for the IAB network. This generator can produce realistic IAB traces in various scenarios and has been trained using a real dataset. The statistical evaluation, based on the Kolmogorov–Smirnov (KS) test, demonstrates that the generator and the actual dataset are comparable. We integrate the proposed generator into the SeBaSi simulator to enable full-stack simulation of MH-IAB networks using real data, which guarantees the accuracy and realism of the results. For instance, we demonstrate the performance of an MH-IAB deployment in London City and present the main simulation metrics and results.

The rest of the paper is organized as follows. Sec. II introduces the IAB system model and a summary of the SeBaSi simulator. Sec. III describes the proposed data generator framework. Sec. IV demonstrates statistically results

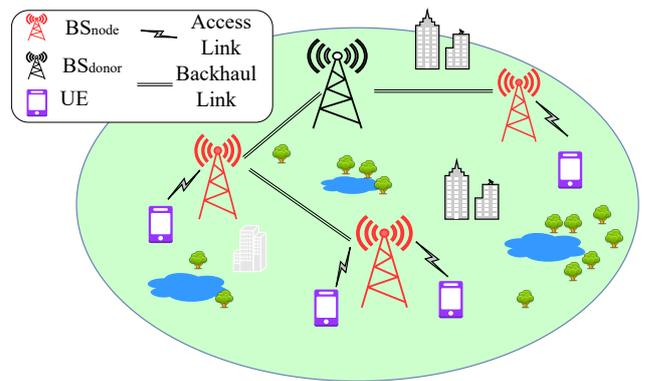


Fig. 1: Illustration of the IAB scenario (with backhaul and access link), with $N_I = 3$ IAB nodes.

and numerical results for the London City scenario. Finally, Sec. V concludes the paper.

II. IAB SYSTEM MODEL

This section provides an overview of the system model and of our assumptions in Sec. II-A, as well as a brief introduction to SeBaSi in Sec. II-B.

A. System Model

We consider a Time Division Multiple Access (TDMA) system, as illustrated in Fig. 1, where N_U User Equipments (UEs) exchange data with a single IAB donor that has fiber connectivity to the Core Network (CN). In order to provide stable coverage, the donor is supported by N_I IAB nodes, which may be linked either directly to the donor or to nearby base stations, potentially creating a multi-hop wireless backhaul. Without making any assumptions that limit the scope of the situation, we only consider uplink traffic.

We divide the time resources into T radio subframes, each with a duration of $T_{sub} = 1$ ms, whereas all nodes are equipped with transmission buffers. Consequently, the data that node i transmits to gNB k (either the IAB donor or an IAB node) during subframe t is stored in its buffer $B_k(t)_i$. This data represents either the packets that the donor has successfully received or the data that will be forwarded to the next hop along the path during subframe $t + 1$ in the case of IAB nodes. We assume that the backhaul links operate in the mmWave spectrum, and each IAB node is equipped with two Radio Frequency (RF) chains, so that two antenna systems can be flexibly used for the backhaul and access communications. The Sionna Ray Tracing (RT)³ tool is used to model the channel and calculate Power Spectral Density (PSD) in this research. The Signal to Interference plus Noise Ratio (SINR) of a packet from source node s to destination node d , $\delta_{s,d}$ can be expressed as

$$\delta_{s,d} = \frac{|h_{s,d}|^2 \sigma_x^2}{\sigma_n^2 + \sum_{i \in \mathcal{I}} \sigma_i^2}, \quad (1)$$

¹<https://github.com/signetlabdei/ns3-mmwave-iab>

²https://github.com/TUDA-wise/safehaul_infocomm2023

³<https://nvlabs.github.io/sionna/api/rt.html>

where $h_{s,d}$ represents the equivalent channel response between the communication endpoints, \mathcal{I} denotes the set of interferers, σ_x^2 , σ_i^2 and σ_n^2 are the powers of the transmitted signal, the i -th received interfering signal, and the thermal noise at the receiver, respectively. The corresponding access (backhaul) throughput $\beta_{j,k}^A(t)$ ($\beta_{s,d}^B(t)$) reads

$$\beta_{s,d}^A(t) = \frac{1}{T_{sub}} \sum_{l=1}^{B_s^t} \mathbb{1} \left\{ \hat{b}_l(\delta_{s,d}) = b_l \right\}, \quad (2)$$

whereas gNB $k = 0, \dots, N_I$, with index 0 denoting the IAB donor, receives data from node s , where B_s^t denotes the number of bits transmitted from user (IAB node) j to gNB d during subframe t and $\hat{b}_l(\delta_{j,k})$ is the l -th decoded bit at the receiver, as a function of $\delta_{j,k}$.

B. SeBaSi

SeBaSi [16] is a system-level simulator, built on top of the open-source SionnaTM [17] simulator, which is used for modeling the physical layer of 5G and beyond 5G networks. SeBaSi is specifically designed to model 3GPP Release 17 IAB cellular networks. It is written in Python and operates on top of any link-level simulator, such as Sionna, and simulates essential components. In order to incorporate self-backhauling IAB capabilities into Sionna, we have successfully integrated several system-level features into SeBaSi. The extensions, described in detail in [13], consist of a scheduler at the Medium Access Control (MAC) level, layer-2 buffers, and algorithms for selecting the backhaul path. In addition, we implemented 5G-NR procedures, such as codebook-based beamforming and SINR computation, to improve the alignment of Sionna's physical layer with the latest 5G-NR standards. In addition, we enhanced SeBaSi to include support for sub-THz links in the backhaul [18]. This is enabled by the extension of SeBaSi channel modeling to support sub-THz by simulated traces in Terasim [19]. So the links can be configured to function at either mmWave, sub-THz, or a combination of both frequencies. The purpose of this enhancement is to evaluate the performance of sub-THz frequencies for IAB, which is in line with the latest research and standardization activities on 6G. At the physical layer, Sionna and SeBaSi implement the 3GPP TR 38.901 model for the mmWave channel, even though the most recent version of SeBaSi also includes a built-in ray tracer tool to model the channel. To address routing within the wireless backhaul network, we implemented the Backhaul Adaptation Protocol (BAP) layer in the upper layers of SeBaSi [20]. This layer uses a MAC-level scheduler that operates in a TDMA manner. Additionally, it utilizes hop-by-hop Radio Link Control (RLC) channels to simulate layer-2 buffering and data transmission.

SeBaSi enables users to customize various simulation parameters, including the duration and mode of the simulation, the size of the packets, and the rate of data transmission from either individual user equipment or the entire system. The simulation modes being considered are the *run* mode and the *debug* mode. The debug mode offers extra control signals

and related information. In addition, users have the ability to personalize the scenario by selecting the quantity and location of UEs and base stations, as well as the IAB topology, which refers to the wireless backhaul links between gNBs. The backhaul scheduler algorithm, which determines which backhaul links to schedule in each time slot, allows users to either create custom policies or select from predefined options such as SCAROS [21], MLR [6], Safehaul [13], and SINR-based [18].

The simulator generates a comprehensive collection of system-level Key Performance Indicators (KPIs), including end-to-end latency, throughput, and packet drop rate. Each of these metrics can be collected and shown for each IAB node or for the entire network. Furthermore, SeBaSi provides internal and/or lower layer metrics, such as the timestamp of packet generation and arrival, destination UE, and the backhaul path. In addition, it provides information on the load of each IAB node for each time step, including both the access and the backhaul interfaces.

III. DATA GENERATION FRAMEWORK

Data Driven-SeBaSi (DD-SeBaSi) is a framework that models IAB networks using a real dataset integrated on top of SeBaSi. DD-SeBaSi incorporates all the existing features of SeBaSi and introduces an additional functionality that allows users to select between generating parameters (e.g., the number of connected UEs per gNB, system Rate, and Reference Signal Received Power (RSRP)) randomly or using data traffic, either from a real dataset or from our data generator. Figure 2 demonstrates the proposed data generator in collaboration with SeBaSi. In order to accomplish this, we initially collect the dataset (described in Sec. III-A). Next, we employ the GAN architecture to train the model using this dataset and generate new synthetic data which is still accurately representative of the original data. Ultimately, we integrated the GAN model into the SeBaSi as a new feature to have DD-SeBaSi. It is noteworthy that users have the ability to define any scenario, and the data generator will attempt to create data that is relevant to that scenario, using the real dataset as a basis. In the following sections, we will begin by providing an overview of the data collection process. Next, we will present the proposed data generator. Finally, we will demonstrate how we seamlessly incorporate it into SeBaSi.

A. Dataset collection

In order to conduct our investigation, we utilize real-world datasets that we collected from a mobile network that operates in Europe, and maintains a significant market share. The dataset comprises individual measurement samples of a variety of metrics from end-user devices that were collected during 2 months in 2023. The dataset contains thousands of samples, each of which is associated with the corresponding radio sector identity and geographical coordinates. We concentrate our analysis on London, which is the primary innovation hub for the operator as a result of the high population density and growing service demand.

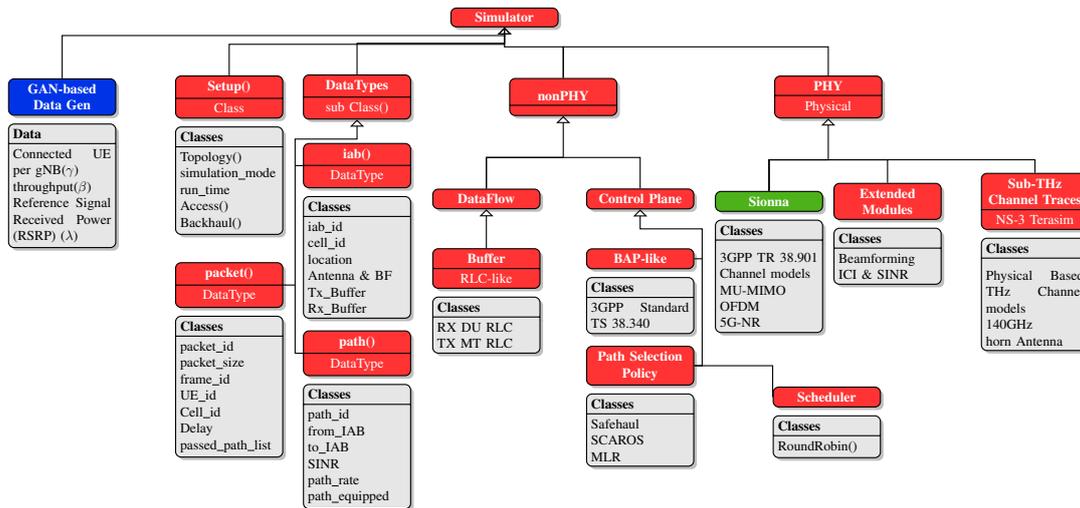


Fig. 2: Overall design of DD-SEBASI Framework extension. The blue block is the proposed data generator, the red blocks are SeBaSi, and the green block represents the Sionna simulator [17].

B. Data Generator

Within the collected dataset, we have selected three parameters that are likely to have a significant influence on the performance: the number of connected UEs per gNB (γ), the system rate per UE (β), and RSRP (λ). Given the unavailability of real-world mmWave deployments at the time of writing, all data is limited to cellular deployments operating at sub-6 GHz frequencies. As a first step, we establish a simulation scenario utilizing the Sionna RT tool to generate rays at mmWave spectrum to obtain RSRP values with corresponding deployment positions. RT is an excellent tool for modeling the environment and generating paths; however, it is restricted to specific scenarios. Therefore, in order to obtain the appropriate channel and, in particular, the RSRP value, it is necessary to model the environment with all details for each scenario. So to do simulation with DD-SeBaSi, first we model the London City scenario (same are real deployment) in RT to obtain proper RSRP values. We utilize exclusively real datasets for the γ and β parameters.

GANs are a distinct type of DNNs that have the capacity to generate data by learning the precise statistical characteristics of a given dataset through indirect methods. The interesting feature of GANs is that they can be trained using a limited amount of available real data to generate synthetic data in different scenarios and conditions. This synthetic data can then be used in data generation for wireless communication networks. GANs involve two key components: the generator G , which transforms a random sample from a uniform distribution into a sample that follows the data distribution, and the discriminator D , which assesses whether a given sample is representative of the data distribution or not. To learn a generator distribution p_g over the dataset, it constructs a mapping function $G(z)$ from a noise distribution $p_z(z)$ where z is the input noise to the output of generator (Xf). The discriminator $D(x)$ returns a single scalar value that represents the probability that X is derived from the real

dataset (Xr) rather than $p_g(Xf)$, therefore showing the authenticity of the data. The negative relationship between the two components of the GAN is reflected in the min-max equation, which is a fundamental component of the training objective. Concurrent training of both G and D is conducted, with parameters modified for G to minimize the cost functions $\log_{10}(1 - D(G(z)))$ and $\log_{10} D(x)$ accordingly, using the min-max value function $V(D, G)$ ⁴ which is defined as:

$$\min_G \max_D V(D, G) = \mathbb{E}_{X \sim \text{data}(X)} [\log D(X | y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z | y)))] \quad (3)$$

The min-max equation encapsulates the adversarial nature of GAN training, in which the generator and discriminator are perpetually enhancing themselves in response to each other's advancements, resulting in improved data generation over time. The underlying premise of min-max is that $D(X)$ is attempting to optimize its accuracy by increasing the probability of distinguishing between real and fake data. $G(X)$ is attempting to reduce the discriminator's capacity to do so by ensuring that the fake data appears as real as possible. $D(X)$ should be unable to differentiate between real and fake data when the GAN reaches equilibrium, which implies that $D(G(z)) = 0.5$ for all z . At this step, $D(X)$ is at its most confused, and the generator generates data that is indistinguishable from real data.

Figure 3 illustrates the architecture of the DD-SeBaSi GAN data generator model, whereas Z , Xr , and Xf represent the input noise, $\{\gamma, \beta, \lambda\}_{real}$, $\{\gamma, \beta, \lambda\}_{fake}$ ⁵, respectively. Figures 3a and 3b represent the architecture of G and D whereas G and D take as input Z , and $\{Xr, real\}$ or $\{Xf, fake\}$, respectively.

⁴For a more detailed description of the GAN training process, we refer the interested readers to [22].

⁵During training we refer to output of Generator(Xf) as fake, after training, during the simulation campaign in DD-SeBaSi, we refer with $G, \{\gamma, \beta, \lambda\}_{fake} = \{\gamma, \beta, \lambda\}_G$

Finding the best parameters for training a GAN is a challenging task because it involves optimizing multiple hyperparameters, including the learning rate, batch size, epoch count, generator and discriminator layer count, activation functions, regularization techniques, and more. We performed training on the model using several combinations of hyperparameters. We evaluated the performance using a specific measurement and chose the hyperparameters that produced the most advantageous outcomes. The most favorable hyperparameters are as follows. We utilize the rectified linear unit (ReLU) as the activation function. The training method comprises 80,000 epochs, with a batch size of 32. We employ Adam as the optimizer, utilizing a learning rate of 0.001. The loss function for the generator G and discriminator D is obtained by utilizing the mean absolute error for G and binary cross-entropy for D .

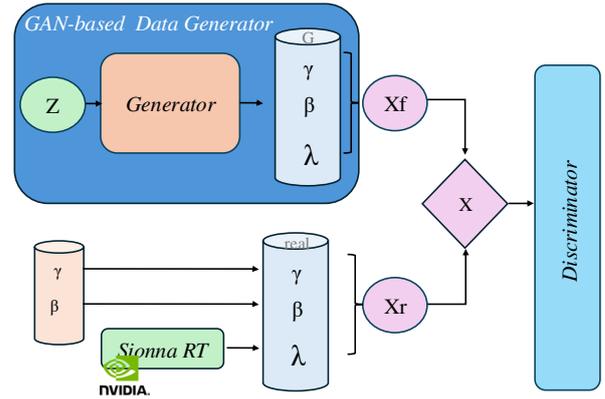
IV. PERFORMANCE EVALUATION

In Sec. IV-A we provide a statistical analysis that validates the accuracy of the proposed data generator. In Sec. IV-B we validate via simulation the implementation of the DD-SeBaSi simulator using synthetic data from the data generator, and evaluate the performance of an IAB network considering different backhaul schedulers.

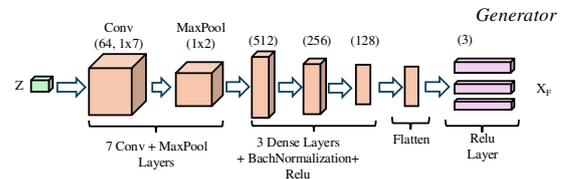
A. Statistical Analysis

The initial step in using the data generator in the SeBaSi simulator is to verify that the proposed data generator is consistent with the actual dataset [23]. Graphically representing the Cumulative Distribution Functions (CDFs) of two distributions is an effective approach to visualize the degree of similarity or any strong disparity between them. In order to accomplish this, we plot the CDF of real and generated values of γ , β , λ in Fig. 4. While we see that the two curves almost overlap, which indicates that our data generator is accurate, we use the KS test to formally verify that the generated data is statistically consistent with the actual dataset. KS is an important metric used in statistical analysis to compare the distributions of two datasets. Its purpose is to determine whether a dataset adheres to a specific distribution [24]. The test yields a P-value that signifies the likelihood of achieving the observed disparity in distributions due to random chance. A higher P-value suggests that the two datasets are probably sampled from the same distribution. The D-value, also referred to as the KS statistic, is a quantitative measure that captures the largest, also referred to as vertical distance, between the CDFs of two datasets under comparison. Consequently, the datasets exhibit greater similarity when the D-value is small.

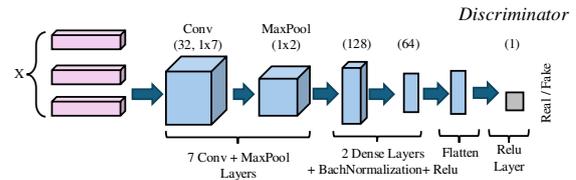
We conducted the KS for the CDF of $\{\gamma, \beta, \lambda\}_{real}$ and $\{\gamma, \beta, \lambda\}_G$. The results of the test are presented in Table I. It is shown that all parameters ($\{\gamma, \beta, \lambda\}$) are passing the test very well: An extremely low D-value and high P-value for all parameters suggest that the distribution of the two datasets is comparable. Therefore, our KS test does not reject the null hypothesis, which indicates that there is insufficient evidence



(a) GAN-based Dataset generator framework architecture



(b) Architecture of Generator of GAN



(c) Architecture of Discriminator of GAN

Fig. 3: Structure of the proposed data generator, including DNN-based architecture

to demonstrate that the sample distribution deviates from the reference distribution.

TABLE I: A goodness of fit (two sample KS Test) for.

Parameter	P-value	D-value
γ	0.714	0.073
β	0.795	0.0124
λ	0.892	0.0091

B. IAB simulation Results

In this section we use SeBaSi to run IAB simulations in different scenarios, and used our GAN data generator, that was previously validated in Sec. IV-A, to model the channel of both the access and the backhaul links. Simulation results are given a function of the number of IAB nodes and

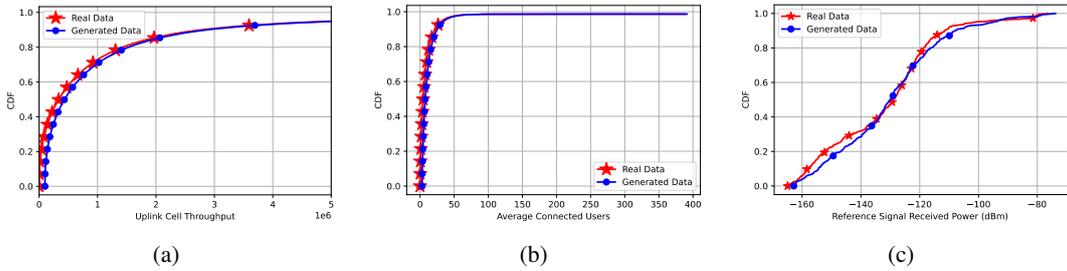
Fig. 4: CDF of α, β, λ comparison of generated and real datasets

TABLE II: Simulation parameters.

Parameter	Value
Carrier frequency and bandwidth	30 GHz and 400 MHz
IAB RF chains	2 (1 access + 1 backhaul)
Number of BS-nodes N	100
IAB Backhaul and access antenna array	8Hx8V and 4Hx4V
UE antenna array	4Hx4V
IAB and UE height	15 m and 1.5 m
IAB antenna gain	33 dB
Noise Figure	10 dB

for different scheduler implementations. We plot the mean throughput, latency, and packet drop rate.

Simulation Scenario We simulate the actual cellular network deployment configuration of London City. Specifically, we consider $N = 100$ 5G-NR base stations within a 15 Km^2 area, as shown in Fig. 5. The specific simulation parameters are outlined in Table II.

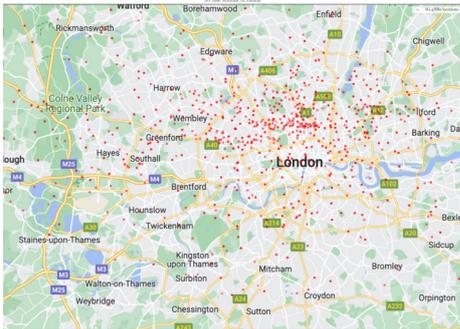


Fig. 5: Locations of BS-nodes (red dots) in London City

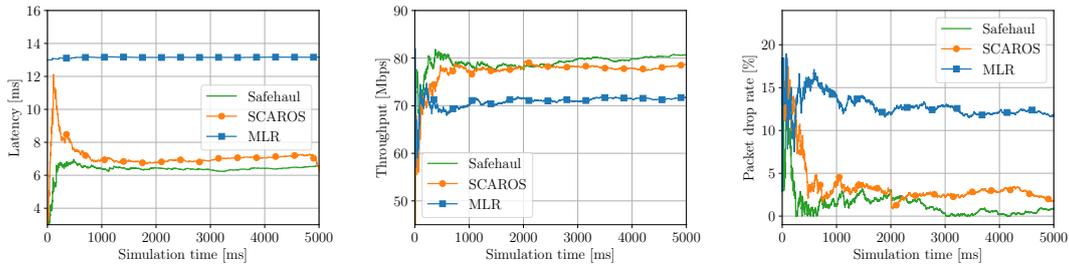
Numerical Results We use three different IAB schedulers available in SeBaSi: (i) Safehaul [13], a risk-averse learning method for ensuring reliability in mmWave systems which uses a Reinforcement learning algorithm to increase reliability in the network; (ii) Scalable and Robust Self-backhauling Solution (SCAROS), an online learning-based technique that reduces the average backhaul scheduling latency in the network [21]; and (iii) Maximum-local-rate (MLR), a greedy statistical method that aims at maximizing throughput by choosing links with the highest data rate. This approach operates offline, greatly facilitating its application in real-world situations, but at the cost of decreased performance.

1) *Scenario 1: Average Network Performance:* Evaluation of the algorithms' performance over time is essential for determining the rate at which the learning-based approaches, namely Safehaul and SCAROS, converge. Therefore, in Fig. 6

we display the mean latency, throughput, and packet loss rate of the IAB network over time. In Fig. 6a, we can observe that Safehaul rapidly converges to an average latency of approximately 6.5 ms which is 11% and 48.4% lower than the latency of SCAROS and MLR, respectively. The high performance of Safehaul stems from the joint minimization of the average latency and the expected value of its tail loss, which results in avoiding risky situations where latency goes beyond T_{\max} (maximum time before packet drop in the network, after which a packet is considered as dropped, 50 ms in our simulation campaign). This is not the case for SCAROS where we observe a high peak in the latency before convergence. *It is exactly the avoidance of such transients in Safehaul that leads to higher reliability in the system.* The reliability offered by Safehaul allows Mobile Network Operators (MNOs) to deploy self-backhauling in an online fashion and without disrupting the network operation. Figure 6b illustrates that the average transmission of the network is not adversely affected by the risk-aversion capabilities of Safehaul. Safehaul's performance is approximately 79.3 Mbps, which is 11.7% higher than that of MLR and comparable to SCAROS. The behavior observed in Fig. 6a is consistent with the performance depicted in Fig. 6c in terms of packet drop rate. Safehaul obtains the lowest packet drop rate among the reference schemes, which is 30.1% (84.0%) lower than SCAROS (MLR).

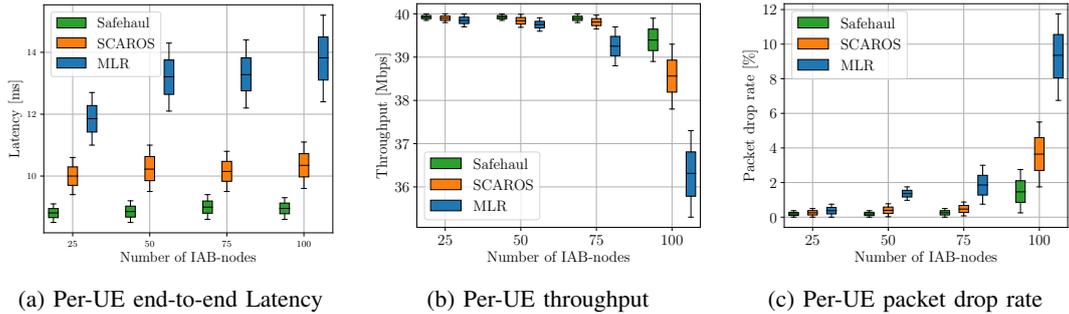
2) *Scenario 2: Impact of the IAB Configuration:* In Fig. 7 we evaluate the performance of the IAB network as a function of the number of IAB nodes, that we change from 25 to 100. Simultaneously, we augment the network's burden by increasing the number of UEs (2 UEs per IAB node). We demonstrate that Safehaul consistently obtains better performance in comparison to the reference schemes. This demonstrates that Safehaul accomplishes the intended optimization objective, which is the joint minimization of the average end-to-end latency and its anticipated tail loss. As the number of IAB-nodes increases, Safehaul is capable of maintaining a nearly constant latency, as illustrated in Fig. 7a. In particular, the variation of latency with Safehaul is 56.1% and 71.4% lower than that with SCAROS and MLR, respectively. Additionally, Safehaul obtains an 11.1% and 43.2% lower latency than SCAROS and MLR, respectively. The latter's high variance is a result of the lack of adaptation capabilities.

The average throughput of the learning-based approaches, i.e., Safehaul and SCAROS, remains constant as the number



(a) Average per-UE end-to-end latency (b) Average per-UE throughput (c) Average per-UE packet drop rate

Fig. 6: Average network performance for 50 UEs(Scenario 1).



(a) Per-UE end-to-end Latency

(b) Per-UE throughput

(c) Per-UE packet drop rate

Fig. 7: Network performance for $\{25, 50, 75, 100\}$ BS-node, 2 UEs per BS-nodes on average, and 40 Mbps per-UE source rate (Scenario 2).

of IAB nodes increases, as illustrated in Fig. 7b. Safehaul, on the other hand, achieves the lowest variation, differential of maximum and minimum value, in throughput, with a value of 0.90, as opposed to the benchmark schemes' 1.9 and 2.8. The reliability capabilities of Safehaul are corroborated by the results in Fig. 7c, where we plot the packet failure rate vs. the number of IAB nodes. It is worth noting that Safehaul consistently outperforms the reference schemes and exhibits the least variation in results (at least 47.3% lower than the benchmarks). When the greatest network size and traffic are taken into account, namely 200 BS-nodes and 400 UEs, Safehaul achieves a 49.3% and 81.2% lower packet drop rate than SCAROS and MLR, respectively.

Summary We present an example scenario of London City end-to-end performance metrics results, utilizing the GAN-based data generator to demonstrate how various schedules can be validated in the SeBaSi simulator. It is evident that from the obtained results, among the backhaul schedulers, Safehaul has the potential to attain superior performance, which is directly consistent with the results obtained in [13] for Manhattan City using a random data generator.

V. CONCLUSIONS AND FUTURE WORK

The scope of this research was to validate the integration between SeBaSi and the data generator, and present (for the first time) realistic IAB results, i.e., obtained considering synthetic (though validated with real traces) data for the IAB channel. The numerical results align with prior trends and demonstrate the applicability of approaches for implementation in the actual system. In future work, further real

datasets will be gathered from various operators to enhance the generality of the framework and enable the utilization of the data generator in other simulators, such as ns-3. In addition, we would extend simulations to do localization to find the best spot to install the IAB node in various scenarios.

VI. ACKNOWLEDGEMENT

This work was partially supported by the EU MSCA ITN project MINTS “Millimeter-wave Networking and Sensing for Beyond 5G” (grant no. 861222) and funded by Grant PID2021-126431OB-I00 funded by MCIN/AEI/10.13039/501100011033 grant (ANEMONE), Spanish MINECO grant TSI-063000-2021-54 (6G-DAWN ELASTIC), TSI-063000-2021-55 (6G-DAWN RESILIENT), TSI-063000-2021-56 (6G-BLUR SMART), TSI-063000-2021-57 (6G-BLUR JOINT), and “ERDF A way of making Europe” by Generalitat de Catalunya grant 2021 SGR 00770.” Also the work was partially supported by the SNS (Smart Networks and Services) JU and its members, funded by the European Union under Grant Agreement number 101139161.

REFERENCES

- [1] 3GPP, “NR; Integrated Access and Backhaul (IAB) radio transmission and reception,” Technical Specification (TS) 38.340, July 2024, v.18.5.0.
- [2] A. A. Gargari, A. Ortiz, M. Pagin, W. de Sombre, M. Zorzi, and A. Asadi, “Risk-averse learning for reliable mmwave self-backhauling,” *IEEE/ACM Transactions on Networking*, pp. 1–15, 2024.
- [3] 3GPP, “NR; Study on integrated access and backhaul,” Technical Specification (TS) 38.340, July 2019, v.16.0.0.

- [4] F. Rezazadeh, S. Barrachina-Muñoz, H. Chergui, J. Mangués, M. Benis, D. Niyato, H. Song, and L. Liu, "Toward explainable reasoning in 6g: A proof of concept study on radio resource allocation," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6239–6260, 2024.
- [5] O. P. Adare, H. Babbili, C. Madapatha, B. Makki, and T. Svensson, "Uplink power control in integrated access and backhaul networks," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2021, pp. 163–168.
- [6] M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource management for 5G NR integrated access and backhaul: A semi-centralized approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 753–767, Jul. 2022.
- [7] C. Fang, C. Madapatha, B. Makki, and T. Svensson, "Joint scheduling and throughput maximization in self-backhauled millimeter wave cellular networks," in *17th International Symposium on Wireless Communication Systems (ISWCS)*, 2021.
- [8] C. Madapatha, B. Makki, C. Fang, O. Teyeb, E. Dahlman, M.-S. Alouini, and T. Svensson, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1374–1389, Sep. 2020.
- [9] 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, Mar 2020.
- [10] M. Polese, M. Giordani, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "End-to-end simulation of integrated access and backhaul at mmwaves," in *IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2018.
- [11] C. Madapatha, B. Makki, H. Guo, and T. Svensson, "Constrained deployment optimization in integrated access and backhaul networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2023.
- [12] F. Rezazadeh, A. A. Gargari, S. Lagén, J. Mangués, D. Niyato, and L. Liu, "Genonet: Generative open xg network simulation with multi-agent llm and ns-3," *arXiv preprint arXiv:2408.13781*, 2024.
- [13] A. Ashtari Gargari, A. Ortiz, M. Pagin, A. Klein, M. Hollick, M. Zorzi, and A. Asadi, "Safehaul: Risk-Averse Learning for Reliable mmWave Self-Backhauling in 6G Networks," in *IEEE Conference on Computer Communications (INFOCOM)*, New York, USA, 2023.
- [14] J. Kim, Y. Ahn, and B. Shim, "Massive data generation for deep learning-aided wireless systems using meta learning and generative adversarial network," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 1302–1306, Sep 2022.
- [15] E. Ayanoglu, K. Davaslioglu, and Y. E. Sagduyu, "Machine learning in nextg networks via generative adversarial networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 480–501, Feb 2022.
- [16] A. A. Gargari, M. Pagin, A. Ortiz, N. M. Gholian, M. Polese, and M. Zorzi, "Demo:[sebas] system-level integrated access and backhaul simulator for self-backhauling," in *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2023, pp. 355–357.
- [17] J. Hoydis, S. Cammerer, F. A. Aoudia, A. Vem, N. Binder, G. Marcus, and A. Keller, "Sionna: An open-source library for next-generation physical layer research," *arXiv preprint arXiv:2203.11854*, 2022.
- [18] A. A. Gargari, M. Pagin, M. Polese, and M. Zorzi, "6G Integrated Access and Backhaul networks with sub-terahertz links," in *18th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2023, pp. 13–19.
- [19] Z. Hossain, Q. Xia, and J. M. Jornet, "Terasim: An ns-3 extension to simulate Terahertz-band communication networks," *Nano Communication Networks*, vol. 17, pp. 36–44, September 2018.
- [20] 3GPP, "NR; Backhaul Adaptation Protocol (BAP) specification," Technical Specification (TS) 38.340, May 2022, v.17.0.0.
- [21] A. Ortiz, A. Asadi, G. H. Sim, D. Steinmetzer, and M. Hollick, "SCAROS: A scalable and robust self-backhauling solution for highly dynamic millimeter-wave networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2685–2698, Oct. 2019.
- [22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct 2020.
- [23] X. Liao, J. Xie, and J. Zhou, "A data-driven optimization method for simulating arbitrarily distributed and spatial-temporal correlated radar sea clutter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, Oct 2023.
- [24] A. Zeimbekakis, E. D. Schifano, and J. Yan, "On misuses of the kolmogorov-smirnov test for one-sample goodness-of-fit," *The American Statistician*, pp. 1–18, May 2024.