Joint Routing and Energy Optimization for Integrated Access and Backhaul with Open RAN

Gabriele Gemmi*§, Maxime Elkael†§, Michele Polese‡, Leonardo Maccari*, Hind Castel-Taleb†, Tommaso Melodia‡

*Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy.
†SAMOVAR, Telecom Sud-Paris, Institut Polytechnique de Paris, France
‡Institute for the Wireless Internet of Things, Northeastern University, Boston, MA, U.S.A.
gabriele.gemmi@unive.it

***Abstract*—Energy consumption represents a major part of the operating expenses of mobile network operators. With the densification foreseen with 5G and beyond, energy optimization has become a problem of crucial importance. While energy optimization is widely studied in the literature, there are limited insights and algorithms for energy-saving techniques for Integrated Access and Backhaul (IAB), a self-backhauling architecture that ease deployment of dense cellular networks reducing the number of fiber drops. This paper proposes a novel optimization model for dynamic joint routing and energy optimization in IAB networks. We leverage the closed-loop control framework introduced by the Open Radio Access Network (O-RAN) architecture to minimize the number of active IAB nodes while maintaining a minimum capacity per User Equipment (UE). The proposed approach formulates the problem as a binary nonlinear program, which is transformed into an equivalent binary linear program and solved using the Gurobi solver. The approach is evaluated on a scenario built upon open data of two months of traffic collected by network operators in the city of Milan, Italy. Results show that the proposed optimization model reduces the RAN energy consumption by 47%, while guaranteeing a minimum capacity for each UE.**

***Index Terms*—Energy Optimization, Integrated Access and Backhaul, O-RAN, 5G**

I. Introduction

Ultra-dense deployment and millimeter wave (mmWave) have been portrayed as the solution to meet the stringent requirements standardized with 5G in terms of data rates [1]. As actually proven by many studies, mmWave is capable of providing multi-gigabit connectivity to User Equipments (UEs) [2], [3], and Integrated Access and Backhaul (IAB) has been proven to be an effective way to reduce the deployment costs [4]. This technology, introduced in 3GPP Release 16, allows, in fact, connecting only a subset of the Next Generation Node Bases (gNBs), called IAB-donors, to the fiber backhaul while the rest of the gNBs, called IAB-node, rely on in-band wireless communication to reach one of the donors, forming a multihop wireless network.

This work was partially supported by Agence Nationale pour la Recherche through the AIDY-F2N project, grant number ANR-19-LCV2-0012, 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.

§Equal contribution

By dynamically activating and deactivating gNBs with respect to the current load of the network, it is possible to reduce the energy footprint of the system, switching off IAB-nodes that are not strictly necessary to match the requested level of service. This is of critical importance since energy consumption accounts for up to 60% of the Operational Expenditure (OpEx) [5]. Research on energy optimization techniques for traditional networks usually assumes that all gNBs are connected to the fiber backhaul. The presence of wireless-only IAB-nodes, however, increases the complexity of the scenario. In fact, the deactivation of an IAB-node might disrupt the service of another IAB-node that is relying on it. Moreover, IAB-nodes need to support periodic wake-up to update their radio state in order to dynamically change the topology when needed.

Thanks to the effort led by the Open Radio Access Network (O-RAN) Alliance, which has opened the Radio Access Network (RAN) architecture by introducing interfaces such as the E2 and O1 and the concept of RAN Intelligent Controller (RIC) [6], it is now possible to integrate custom closed-loop control logic in the RAN. This has been already studied for several applications, such as network slicing [7], and, more recently, an extension of the O-RAN architecture has been proposed for IAB [8].

Our novel optimization approach takes advantage of this closed-loop control framework to overcome the limitations discussed above and dynamically minimize the number of active IAB-nodes, while maintaining a minimum capacity per UE. The optimization—based on input data that can be obtained through O-RAN interfaces—generates a topology tree over which we route the traffic from each UE to the IAB-donor, deactivating IAB-nodes that are not needed and distributing UEs across the available gNBs.

To the best of our knowledge, joint routing and energy optimization on multi-hop IAB topologies has never been studied before. Most studies focus on optimizing the IAB topology with different constraints, such as in [9] where the fiber-deployment cost is minimized or in [10] where the UE data rates are maximized. Other studies, more focused on power and energy-related optimization of IAB networks, exist, but they either optimize the energy consumption after the topology and routing have been chosen, such as in [11], or they are restricted to a single-hop architecture, such as in [12], where a low energy multiple access scheme is designed.

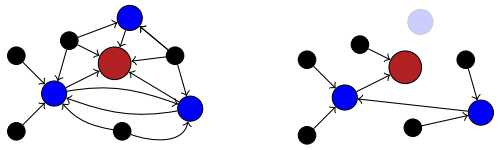
Optimization for energy consumption is a common concern in both this study and Wireless Sensor Networks (WSNs). However, the goals differ: WSNs aim to extend network lifetime due to their battery-powered nodes, as in [13], while our grid-powered nodes target different optimization criteria.

Hence, in this paper, we fill this gap by formulating the problem as a binary nonlinear program. Since its continuous relaxation is non-convex, it is not solvable using off-the-shelf solvers, therefore we show it can be transformed into an equivalent binary linear program, which we then solve using the Gurobi solver. The approach is evaluated on a scenario built upon open data of two months of traffic collected by network operators in the city of Milan, Italy, [14] together with detailed morphological data of the same area. Our optimization model manages to perfectly tune the number of active gNBs with the number of UE, reducing by 47% the total number of hours the gNBs (i.e., IAB-nodes) had been active, while maintaining a minimum downlink capacity for each UE equals to 80Mb/s. This shows how introducing dynamic optimizations enabled by the O-RAN architecture can effectively target improvements in energy efficiency.

II. System Model and Optimization

Let us first introduce the problem formally. We start from a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, called measurements graph, whose nodes can be either IAB-nodes or UEs. We denote the set of UEs as $\mathcal{U} \subset \mathcal{V}$ and the IAB-donor as $t \in V$. Each edge (u, v) of this graph represents a potentially usable wireless link between each node and it is weighted by its available capacity ($c(u, v)$), which depends on the channel quality. Since the goal is to find a tree representing the routing from each UE to towards the donor t, the edges of the graph will be directed accordingly. Access links (originating from the UEs) will always have UEs as source and IAB-nodes as destination. Backhaul links involving t will always point towards it, as it is always the destination to reach the core. The links between IAB-nodes instead can be used in one or the other direction to build the IAB-tree, so the measurements graph contains a couple of links per each neighbor IAB-node pair. Figure 1a and Figure 1b report an example of a measurements graph and a possible IAB-tree.

Local detailed information on the feasibility of wireless links between UEs and IAB-nodes is available on each gNBs. The O-RAN architecture allows extensions to standard interfaces so that we can assume that the local information can be collected by an rApp, running on the non-real-time RIC, which reconstructs the measurements graph we mentioned above. Then, the optimization algorithm periodically runs and pushes the optimized topology to the RAN through the O1 interface. Note that we take into consideration periodic updates of the topology with a period in the order of minutes, so we assume that disabled nodes wake up to receive an updated topology with a similar schedule. This schedule is also perfectly compatible with the non-real-time RIC closed-loop time constraints. Without loss of generality in the following model we will assume the optimization of a single tree, but the proposed optimization model can be trivially adapted to optimize multiple trees.



(a) (b)

Fig. 1: Example of a measurements graph $\mathcal{G}$ (a) and a possible IAB Tree $\mathcal{T}$ (b). IAB-donors are depicted in red, IAB-nodes in blue, UEs in black, and deactivated IAB-nodes in light blue.

A. Optimization problem

Our optimization model aims to identify a tree, denoted as $\mathcal{T}$, which is a subgraph of $\mathcal{G}$, rooted in the IAB-donor, and whose leaves are the UEs. The primary objective of $\mathcal{T}$ is to minimize the energy consumption of the network. This consumption is generally viewed as a combination of static energy, which is expended even when an IAB-node remains idle, and dynamic energy, which depends on the volume of radio resources the IAB-node has to serve. As per the findings in [15], static energy constitutes over 70% of the peak energy consumption by a gNB. Given this significant skew towards static energy, our model simplifies to focus predominantly on static energy. The resultant objective is to construct the tree $\mathcal{T}$ such that the activation of IAB-nodes is minimized.

Additionally, since the whole network operates using the same spectrum, we assume that each node has a Time Division Multiple Access (TDMA) scheduler that operates using a round-robin policy to schedule the inbound traffic and a dedicated radio device to relay the outbound traffic. This additional constraint—which follows guidance from 3GPP technical documents [16]—differentiates our model from a classical multicommodity flow problem, where adjacent edges do not have to share the same time resources as in a wireless network.

We begin the formulation of the problem as a binary multi-commodity flow problem. In such a problem, we have to route a set $\mathcal{K}$ of commodities on the graph, each using a single path. A commodity $k \in \mathcal{K}$ is defined as a triplet s_k, t_k, d_k where s_k is the source node (in our case, a UE), t_k is the destination node (in our case, the IAB-donor t) and $d_k \in \mathbb{R}$ is the bandwidth to reserve on the path from s_k to t_k. These commodities are decided by the Mobile Network Operator (MNO) beforehand, depending on the minimum capacity it wants to guarantee to its customers, and might be differentiated by different classes. The MNO can feed this information to the rApp running the optimization problem. We denote by $\mathcal{N}_{out}(v)$ the outer neighbors of node v and by $\mathcal{N}_{in}(v)$ its inner neighbors. The cardinality of these sets (*e.g.,* the outer and inner degrees) are denoted by $out(v)$ and $in(v)$, and their sum (the degree of the node) $deg(v) = out(v) + in(v)$. Let us introduce the binary variables $a(v)\ \forall v \in \mathcal{V}$ which indicate whether node v is turned on or sleeping, binary variables $f_k(u, v)$ which indicate whether commodity k uses edge $(u, v) \in \mathcal{E}$, and binary variables $f(u, v)$ which indicate whether (u, v) is used by any commodity. We define the problem as the following binary non-linear programming problem:

$$\min \quad \sum_{v \in \mathcal{V}} a(v) \tag{1}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} f_k(u,v)\cdot d_k \le c(u,v) \times \frac{1}{\sum\limits_{w\in\mathcal{N}_{in}(v)} f(w,v)} \quad \forall (u,v)\in\mathcal{E}, \forall k\in\mathcal{K} \tag{2}$$

$$\sum_{v\in\mathcal{V}} f_k(u,v) - \sum_{v\in\mathcal{V}} f_k(v,u) = 0 \qquad \forall u\in\mathcal{V}, \forall k\in\mathcal{K} \tag{3}$$

$$\sum_{v\in\mathcal{V}} f_k(s_k,v) - \sum_{v\in\mathcal{V}} f_k(v,s_k) = 1 \qquad \forall k\in\mathcal{K} \tag{4}$$

$$\sum_{v\in\mathcal{V}} f_k(v,t_k) - \sum_{v\in\mathcal{V}} f_k(t_k,v) = -1 \qquad \forall k\in\mathcal{K} \tag{5}$$

$$a(v) \ge \frac{1}{deg(v)} \left[\sum_{\forall u\in\mathcal{N}_{in}(v)} f(u,v) + \sum_{\forall u\in\mathcal{N}_{out}(v)} f(v,u) \right] \quad \forall (v)\in\mathcal{E} \tag{6}$$

$$f(u,v) \ge f_k(u,v) \qquad \forall (u,v)\in\mathcal{E}, k\in\mathcal{K} \tag{7}$$

$$\sum_{v\in\mathcal{N}_{out}(u)} f(u,v) \le 1 \qquad \forall u\in\mathcal{V} \tag{8}$$

$$a(v), f(u,v), f_i(u,v) \in \{0,1\} \tag{9}$$

Our objective in Eq. (1) is to minimize the number of nodes that are turned on, *e.g.,* the energy consumption of the network. In Eq. (6), the value of variable $a(v)$ is enforced to be 1 if any flow uses node v. Eq. (2) to (5) are multi-commodity flow constraints, where Eq. (3) to (5) enforce the equilibrium of the flow and Eq. (2) ensures the capacity constraints are respected. This constraint is different from the classic multicommodity flow problem, in which it would be

$$f_k(u,v)\cdot d_k \le c(u,v) \qquad \forall (u,v)\in\mathcal{E}, \forall k\in\mathcal{K}.$$

In fact, as mentioned above, in a wireless network the edges adjacent to the same node need to share the spectrum, typically by using TDMA with a specific scheduler. In our case, we have assumed that a Round Robin scheduler allocates equal resources to all the adjacent edges. Finally, the constraint in Eq. (7) ensures that an edge is activated if any commodity uses it and Eq. (8) makes sure all activated nodes have outer degree 1, which implies the network is a tree.

This model is non-linear because of the inverse function in Equation Eq. (2). We now propose an equivalent linearized version of the previous model. We prove the equivalence in Theorem 1.

In the linearized model below, we introduce binary variables $x_i(v) \;\; \forall v \in \mathcal{V}$. These variables are equal to 1 iff at least i of the inner edges incident to v are activated. This enables us to linearize the inverse function in Eq. (2) and to replace it with a weighted sum of those binary variables.

$$\min \quad \sum_{v\in\mathcal{V}} \sum_{i=1}^{in(v)} x_i(v) \tag{10}$$

$$\text{s.t.} \quad f(u,v)\cdot d_k \le c(u,v)\cdot \left(x_1(v) - \sum_{i=2}^{in(v)} \frac{x_i(v)}{(i-1)i} \right) \quad \forall (u,v)\in\mathcal{E} \tag{11}$$

$$x_i(v) \ge \left(\sum_{u\in\mathcal{N}_{in}(v)} f(u,v) - (i-1) \right) /in(v) \quad \forall v\in\mathcal{V}, \forall 1\le i\le d(v) \tag{12}$$

$$x_i(v) \in \{0,1\} \qquad \forall v\in\mathcal{V}, \forall 1\le i\le d(v) \tag{13}$$

$$(3), (4), (5), (7), (8), (9)$$

Theorem 1. *The BNLP (1) - (9) has the same optimal solution as the BLP (3) - (13).*

Proof. Let us first observe that $\left(\sum\limits_{u\in\mathcal{N}_{in}(v)} f(u,v) - (i-1) \right)$ is always positive if at least i inner edges of v are activated, and is nonpositive otherwise. also note that this sum is always lower or equal to $in(v)$. Hence, the right-hand side of Eq. (12) is between -1 and 1, and its value is positive if i inner edges are used. This, combined with the fact we are minimizing the sum of variables $x_i(v)$ means that in an optimal solution to problem $(3)-(13)$, $x_i(v)$ will be equal to 1 if at least i inner edges of v are activated and 0 otherwise.

Let us now observe that in Eq. (11), if n inner edges of v are activated, then $x_1(v), x_2(v), ...x_n(v)$ will be equal to 1. It follows that the sum $x_1(v) - \sum_{i=2}^{in(v)} \frac{x_i(v)}{(i-1)i}$ will be equal to $\frac{1}{n}$, *e.g.* the constraint is equivalent to constraint (2). Finally, observe that since we are building a tree, minimizing its number of edges is equivalent to minimizing its number of nodes, as a tree of n nodes always has exactly $n-1$ edges, meaning the objective function, Eq. (10), is equivalent to the objective in Eq. (1).

□

III. Performance Evaluation Setup

This section presents the techniques used to synthetically generate the set of measurements graphs $\mathcal{G}(\mathcal{V},\mathcal{E})$, needed to evaluate the feasibility and effectiveness of our optimization model. In particular, we will be using datasets representing an area of 0.092km^2 in the center of Milan, Italy. To do so, in the first subsection, we will describe the state-of-the-art techniques used to place the IAB-nodes [17], and the UEs [18]. Then, in the second subsection, we present our channel model, based on 3GPP specifications combined with ray tracing. Finally, the third section presents our data-driven time-varying UE density model, which enables us to generate different instances depending on the time of the day and the day of the week.

A. Placement of gNBs and UEs

The set of nodes of our graph $\mathcal{V}$ is comprised of both IAB-nodes, and UEs, whose placement is done separately using two different techniques. IAB-nodes are placed on building facades with a given density λ_{gNB}. The exact position is computed by taking advantage of a state-of-the-art placement heuristic [17] that exploits highly precise 3D models to place the gNB such that the number of potential UEs' location in line of sight is maximized. UEs are then randomly distributed both in public areas, such as streets, and inside buildings. Specifically, given a density of λ_{UE}, indoor UEs are uniformly randomly distributed inside buildings with a density equal to $r_{i/o}\cdot\lambda_{\text{UE}}$ and outdoor UEs are uniformly randomly distributed inside buildings with

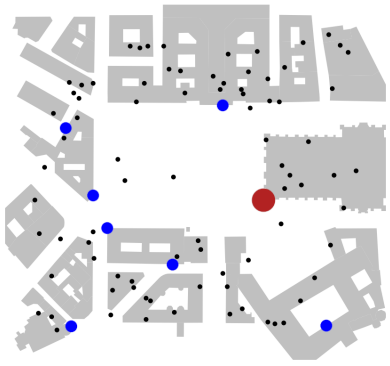


Fig. 2: Sample deployment of a network in the center of Milan, with 1 IAB-donor (in red), 7 IAB-nodes (in blue), and 83 UEs (in black). It corresponds to $\lambda_{\text{UE}}(9) = 900$ UE/km^2 (Mon 9am).

density $(1 - r_{i/o}) \cdot \lambda_{\text{UE}}$, where $r_{i/o}$ is a commonly used ratio of indoor to outdoor UE equal to 0.8 taken from 3rd Generation Partnership Project (3GPP) technical report [18]. In short, we consider that in our simulations 80% of the UEs are placed indoors. Figure 2 shows a deployment with $\lambda_{\text{gNB}} = 45$ and $\lambda_{\text{UE}} = 900$ UE/km^2.

B. Access and Backhaul channel models

Once the location of both UEs and IAB-nodes have been determined, we evaluate the path loss by applying the 3GPP Urban-Micro (UMi) stochastic channel model [19]. However, instead of using the stochastic Line of Sight (LoS) probability model provided by the same UMi model, we deterministically evaluate the LoS by employing ray tracing analysis on the same 3D models used to find the optimal locations, obtaining a more accurate estimation [20]. For indoor UEs, we always consider them to be Non-LoS (NLoS) and we add the additional Outdoor to Indoor (O2I) penetration loss. Since the buildings in the area we consider are mostly made out of concrete, we use the high-loss O2I model [19].

Finally, we compute the Signal-to-Noise-Ratio (SNR) using the thermal noise and by adding the receiver noise figure, then we calculate the Shannon capacity. Both access and backhaul are assumed to be using the same frequencies, but different values of antenna gain and numbers of MIMO layers are used. Tab. I details all the values used in our simulations, which are aligned with typical literature and 3GPP studies on this topic.

C. Time-varying UE density model

As explained in more detail in Sect. I, most studies dealing with topology optimization focus their analysis on a single, or a handful, value of λ_{UE}. Since the energy optimization technique we devise tunes the IAB-node activation on the basis of the number of UEs and their load, we need to evaluate our model on a large number of values of UE density, ideally following a realistic trend. Therefore, we employ a technique used in similar research [21] to devise a time-varying UE density model. First, we extract the cell load profile $p(t)$ related to our analysis area, in Milan, from openly available datasets [14]. We then normalize it in the range $(0, 1]$, and we model the UE density as a function

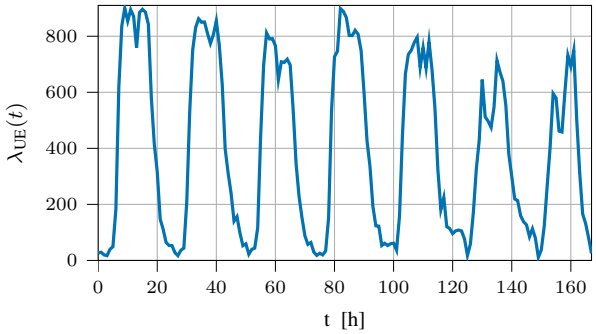


Fig. 3: Weekly profile for the UE density in central Milan.

Parameter	Value
Area size	0.092 km^2
UE density range	[0-900] UE/km^2
Indoor/Outdoor UE ratio	80/20
Carrier frequency	28 GHz
Bandwidth	100 MHz
Noise Figure	5 dB
O2I Loss	14.15 dB
Reception gain (Access/Backhaul)	3 / 10 dBi
MIMO layers (Access/Backhaul)	2 / 4
Backhaul transmission power	30 dBm
Minimum Capacity per UE	80Mb/s
Number of independent simulation runs	10

TABLE I: Simulation Parameters

of time $\lambda_{\text{UE}}(t) = p(t) l \lambda_{\text{gNB}}$, where $l = 10$ is the number of UEs per gNBs taken from the 3GPP technical report [18]. Finally, we generate a set of 168 graphs spanning an average week with a one-hour granularity.

Figure 3 reports the hourly trend of $\lambda_{UE}(t)$ corresponding to the area of our analysis, showing how for several hours every night the network has to serve almost no UEs and how in the weekends, even at peak hours, the density of UEs never exceed 80% of the weekday peak hours.

IV. Results

As already mentioned in Sect. I, we evaluate our model on a 0,092km^2 area in the center of the city of Milan (Italy), for which we computed the UE density trend λ_{UE} of an average week. For each hour of the week (168 in total), we generate the measurements graph as described in the previous section and then we run our optimization algorithm on it. We compare the trees found by our solution with 4 strategies:

- **All donors**, a dense deployment without IAB, where all the gNB are wired. It is an upper bound in terms of energy consumption and capacity. Additionally, no re-distribution of the UEs is performed as they are always attached to the gNB with the lowest SNR.
- **No relays**, a deployment where all the IAB-nodes are not active. It is a lower bound in terms of energy and capacity.
- **Widest Tree**, a strategy that employs the well-known widest path algorithm to find the path of maximum capacity (*e.g.,* with the largest bottleneck in terms of capacity) from each UE towards the donor and deactivates all the IAB-nodes that are not part of any path.

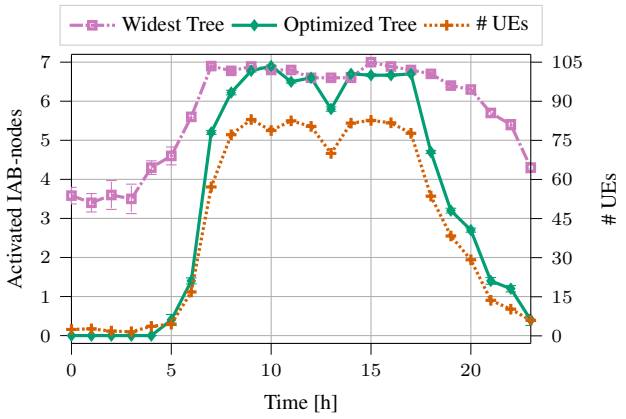


Fig. 4: Number of IAB-nodes activated in the first 24 of the week (Monday).

- **Optimized Tree**, our optimization model.

In the first part of this section, we compare the energy consumption (both in terms of the number of nodes activated and of the overall number of gNB-hours) of the different algorithms. Then, in the second part, we evaluated the topologies in terms of bottlenecks of the downlink capacity.

A. Energy Consumption

To evaluate the energy consumption of the IAB networks we first show the hourly number of active IAB-nodes, then we introduce a metric that measures the total number of hours each IAB-node has been active. The IAB-donor is not taken into consideration as we always need at least one node to be active to provide a minimum service to the users.

Figure 4 shows the number of IAB-nodes activated, on the left axis, and the number of UEs connected to the network, on the right axis. To improve the readability only the values for the first day of the week have been reported. **Optimized Tree**, shows that it is possible to fully deactivate the IAB-nodes at nighttime (from 12 pm to 4 am) and that also during daytime several IAB-nodes can be deactivated. By comparing its trend with the number of UEs, we can also see that it gets perfectly followed, highlighting the effectiveness of our optimization model. **Widest Tree**, on the other hand, never manages to deactivate more than 3 IAB-nodes, highlighting that a specific algorithm is needed to fully implement energy-saving policies.

Additionally, by integrating the number of activated gNB at each hour for the span of the week we obtain the total number of gNB-hours for each strategy. For **No relays**, the value is $168h$, as only one gNB is always active. For **All donors**, on the other hand, the total number of gNB-hours is equals to $168h \cdot 8 = 1344h$, since 8 IAB-donors are active at all times. More interestingly, the values for **Widest Tree** and **Optimized Tree** respectively activate the RAN for $1141h$ and $709h$, which means our method improves the power consumption of 47% over **All Donors** and 38% over **Widest Tree**.

B. Capacity

To evaluate the performance of the topology we analyze the capacity served to each UEs with three different capacity metrics, which are shown in Figure 5 and detailed below. First, let us define some functions used throughout the section. Let $p(u,t)$ be the function returning the set of edges forming the path from u to t over our topology tree. Let $N_{in}(t)$ the number of edges directed towards t in the tree and $c(i,j)$ the capacity of the edge (i,j).

The first metric, called Average Idle Capacity measures the average theoretical capacity of UEs, *e.g.* the capacity that would be attainable if the network resources were completely unused. As detailed in Eq. (14) below, it is computed as the minimum capacity (the bottleneck) of the edges over the path between each UE (u) and the donor (t). Which is then averaged across all the UEs $u \in \mathcal{U}$. This metric represents an upper bound on the capacity per UE.

$$\hat{c}_I = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} c(i,j) \tag{14}$$

Figure 5a shows the Average Idle Capacity for the four different strategies on the first 24 hours of our week. The first insight provided by this figure is that the maximum capacity per UE does not depend on the load of the network. Moreover, as we were expecting **All donors** and **No relay** are respectively the upper and lower bounds in terms of capacity. **Widest Tree**, the strategy that maximizes the bottleneck between each UE and the donor, manages to achieve a capacity very close to the upper bound (8% lower). **Optimized Tree** instead shows a more significant drop with a loss of 35%. The drop can be explained by the minimum capacity constraint that, instead of letting each UEs reach the IAB-donor through the widest path, in certain cases picks paths worse in terms of maximum capacity that instead guarantee the minimum capacity.

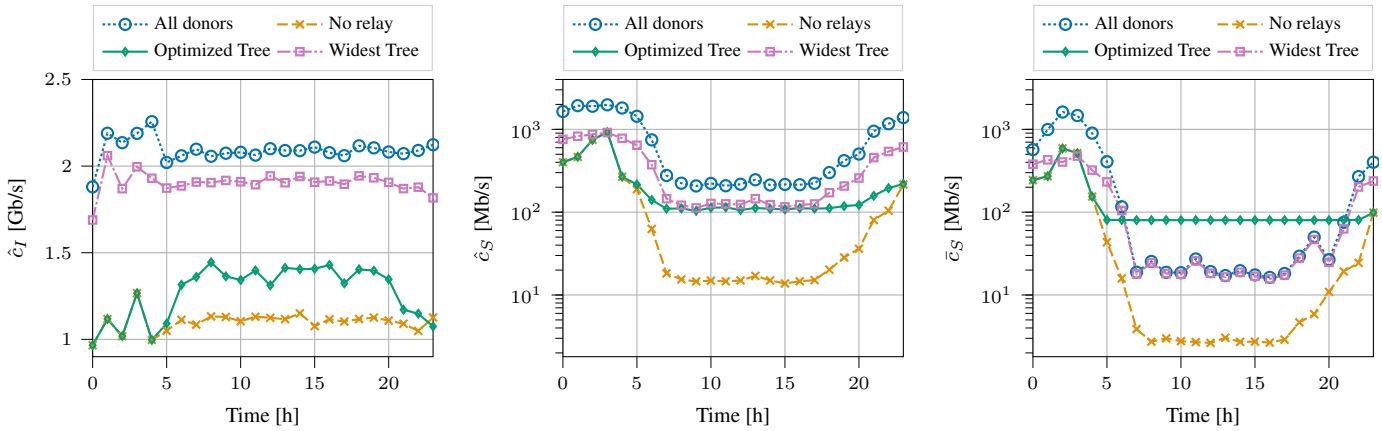
The second metric, called Average Saturation Capacity and detailed in Eq. (15), is formulated in a very similar way as Eq. (14). However, here we assume that all the UEs try to access the network at the same time, thus we divide the capacity of each edge $c_{s,t}$ by the number of inner neighbors of the node t, since those edges share the same resources through the scheduler.

$$\hat{c}_S = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} \frac{c(i,j)}{N_{in}(j)} \tag{15}$$

As in the previous metric, also here **All donors** and **No relays** behaves respectively as upper and lower bound. The difference between the two other strategies, and their distance from the upper bound drops sharply. In fact, at peak time **All donors** is capable of delivering roughly 200Mb/s per UE, while **Optimized Tree** and **Widest Tree** respectively deliver 115 and 130 Mb/s per UE.

The third metric, called Minimum Saturation Capacity and detailed in Eq. (16), measures the capacity delivered to worst UE while the network is under saturation by all the UEs, *e.g.* it defines the minimum level of Quality-of-Service provided by the topology. It is defined similarly to the previous one, but instead of averaging over the UEs we take the worst value.

$$\bar{c}_S = \min_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} \frac{c(i,j)}{N_{in}(j)} \tag{16}$$



(a) Average idle capacity. (b) Average saturation capacity. (c) Minimum saturation capacity.

Fig. 5: Capacity Metrics for the first 24h of the week (Mon).

Figure 5b shows that **Optimized Tree** is the only strategy that manages to guarantee the minimum level of service, equal to 80Mb/s during peak hours (7 am-16 pm), while also minimizing the excessive capacity at night time. In comparison, with **No relays** we measure a minimum level of service that, at peak time, is one order of magnitude lower than the minimum level of service (between 2 and 7 Mb/s) while **Widest Tree** and **All donors** behave similarly in terms of minimum capacity, as they can both take advantage of all the IAB-nodes available. However, since the UEs are not load-balanced across all the available gNB, the minimum level of service is not met. We also note that with **All donors** there is also an excess of capacity at night time; when energy-saving policies could deactivate several IAB-nodes, moreover despite being in a more favorable position where no routing is to be performed beyond the first link between the UEs and the BS, **All donors** still has less capacity than the Optimized Tree. This emphasizes the importance of balancing the load of UEs between base-stations

V. Conclusions

In this paper, we have modeled and solved the problem of finding an IAB topology optimized for energy efficiency. We find optimal solutions to the problem, which enables us to save up to 47% of energy compared to the baseline, while still respecting capacity constraints, which other approaches cannot do. Our results hence show the importance of considering energy-efficiency as a core feature of IAB topology design. In the future, we plan to improve the speed of our algorithm and to propose fast heuristics inspired by it. We also plan on evaluating the energy savings and capacity on a real testbed.

References

[1] D. López-Pérez *et al.*, "Towards 1 Gbps/UE in Cellular Systems: Understanding Ultra-Dense Small Cell Deployments," *IEEE Communications Surveys & Tutorials*, no. 4, pp. 2078–2101, 2015.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, pp. 335–349, 2013.

[3] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, no. 6, pp. 1164–1179, 2014.

[4] M. Polese *et al.*, "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Communications Magazine*, no. 3, pp. 62–68, 2020.

[5] GSMA, "5G Energy efficiencies: Green is the new black," Nov 2020.

[6] M. Polese *et al.*, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, 2023.

[7] S. D'Oro *et al.*, "Orchestran: Network automation through orchestrated intelligence in the open ran," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 270–279.

[8] E. Moro *et al.*, "Toward Open Integrated Access and Backhaul with O-RAN," in *21st Mediterranean Communication and Computer Networking Conference (MedComNet 2023)*. IEEE, 2023.

[9] M. N. Islam *et al.*, "Integrated access backhaul in millimeter wave networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[10] ——, "Investigation of performance in integrated access and backhaul networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2018, pp. 597–602.

[11] D. Meng *et al.*, "An energy-saving scheme with multi-hop transmission for mmwave backhaul networks," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–5.

[12] L. Lei *et al.*, "Noma aided interference management for full-duplex self-backhauling hetnets," *IEEE Communications Letters*, no. 8, pp. 1696–1699, 2018.

[13] T. Rault *et al.*, "Energy efficiency in wireless sensor networks: A top-down survey," *Computer networks*, vol. 67, pp. 104–122, 2014.

[14] G. Barlacchi *et al.*, "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," *Scientific data*, no. 1, pp. 1–15, 2015.

[15] N. Piovesan *et al.*, "Machine learning and analytical power consumption models for 5G Base Stations," *IEEE Communications Magazine*, no. 10, pp. 56–62, 2022.

[16] 3GPP, "Study on integrated access and backhaul," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.874, 01 2019, version 16.0.0.

[17] G. Gemmi *et al.*, "On cost-effective, reliable coverage for los communications in urban areas," *IEEE Transactions on Network and Service Management*, 2022.

[18] 3GPP, "Study on Scenarios and Requirements for Next Generation Access Technologies," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, 04 2022, version 17.0.0.

[19] ——, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.901, 01 2020, version 16.1.0.

[20] G. Gemmi *et al.*, "On the properties of next generation wireless backhaul," *IEEE Transactions on Network Science and Engineering*, 2022.

[21] A. Baiocchi *et al.*, "Joint management of energy consumption, maintenance costs, and user revenues in cellular networks with sleep modes," *IEEE Transactions on Green Communications and Networking*, no. 2, pp. 167–181, 2017.