# TOWARDS RESILIENT WIRELESS NETWORKS: MESH BACKHAULS FOR 5G ULTRA-DENSE NETWORKS AND RURAL INTERNET SERVICE PROVISION

GABRIELE GEMMI

Ph.D. in Computer Science
Department of Environmental Sciences, Informatics and Statistics
Ca' Foscari University of Venice

Ph.D. in Computer Architecture
Department of Computer Architecture
Polytechnic University of Catalunya

Gabriele Gemmi: *Towards Resilient Wireless Networks:*
*Mesh Backhauls for 5G Ultra-dense Networks and Rural Internet Service Provision,* Ph.D. in Computer Science and Computer Architecture, © Feb 2024

SUPERVISORS:
Leonardo Maccari
Llorenç Cerda-Alabern

# ABSTRACT

This thesis aims to investigate the efficacy and applicability of Wireless Backhaul Network (WBN) in two divergent contexts: ultra-dense urban networks for 5G and connectivity solutions for rural or digitally divided areas. WBN offer an adaptable and resilient framework for data transmission, making them an attractive option for the next generation of wireless networks, particularly 5G. In dense urban settings, where the demand for high data rates is pressing, wireless mesh backhauls can serve as a strategic asset in achieving the promised data throughput for 5G networks. These networks inherently require a dense deployment of nodes to deliver on their promise of high-speed, low-latency communication. Therefore, we analyze the role that wireless mesh backhauls can play in such densely populated areas, emphasizing their potential to meet or even exceed 5G's high data rate expectations.

Conversely, in rural or digitally divided areas, the economic feasibility of deploying traditional last-mile copper or fiber-optic networks often proves to be prohibitive. In such cases, WBN, deployed either as Wireless Community Network (WCN) or as part of a Wireless Internet Service Provider (WISP)'s infrastructure, can offer a viable alternative. This technology could bridge the digital divide by providing robust, cost-effective internet connectivity to underserved regions.

To substantiate my findings, I leveraged Geographic Information Systems (GIS) technology and GPU-based computational methods together with graph analysis and simulations. GIS technology, leveraging different open datasets such as Digital Surface Models (DSM) and vectorial maps, make a detailed understanding of the geographical landscape possible, crucial for building detailed feasibility models in various environments. GPU-based computational methods accelerate the model analysis, enabling an in-depth evaluation within a reasonable timeframe. As a result I was able to optimize various network aspects, such as node placement, network topology and energy consumption, which are critical parameters for the effective deployment of these networks.

This work integrates these computational methods and technologies to offer a comprehensive view of how wireless mesh backhauls can be efficiently deployed in both urban and rural settings. The results offer valuable insights into the network architecture best suited for each context, with particular attention to scalability and resiliency. In summary, this thesis contributes to a broader understanding of the potential that wireless mesh backhauls hold in addressing the connectivity requirements of diverse settings, ranging from the ultra-dense urban environment required for 5G to the unique challenges posed by rural and digitally divided regions.

# PUBLICATIONS


[1] G. Gemmi, M. Polese, T. Melodia, *et al.*, "Optimal and dynamic integrated access and backhaul topology adaptation for next-generation cellular networks," in *Submitted: 2024 IEEE International Conference on Computer Communications (INFOCOM)*, Core 2022 A*, 2024 (cit. on p. 103).

[2] G. Gemmi, M. Segata, and L. Maccari, "Estimating coverage and capacity of high frequency mobile networks in ultradense urban areas," *Submitted: Computer Communications*, 2024, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 38).

[3] G. Gemmi, L. Cerdà-Alabern, L. Navarro, *et al.*, "Toward smart community networks," *IEEE Network*, vol. 37, no. 2, pp. 128–134, 2023, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 145).

[4] G. Gemmi, L. Cerdà-Alabern, and L. Maccari, "A robust and cost-effective wireless backhaul design for rural areas," *Submitted: IEEE Transactions on Network and Service Management*, 2023, JCR 2022 Q2, SJR 2022 Q1 (cit. on p. 146).

[5] G. Gemmi, M. Elkael, M. Polese, *et al.*, "Joint routing and energy optimization for integrated access and backhaul with open ran," in *Accepted: 2023 IEEE Global Communications Conference (GLOBECOM)*, Core 2023 B, 2023 (cit. on p. 103).

[6] E. Moro, G. Gemmi, M. Polese, *et al.*, "Toward open integrated access and backhaul with O-RAN," in *21st Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2023 (cit. on pp. 103, 117, 131).

[7] G. Gemmi, M. Segata, and L. Maccari, "Vehicles or pedestrians: On the gNB placement in ultradense urban areas," in *2023 18th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2023, pp. 9–12 (cit. on p. 38).

[8] G. Gemmi, R. L. Cigno, and L. Maccari, "On the properties of next generation wireless backhaul," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 1, pp. 166–177, 2023, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 38).

[9] G. Gemmi, L. Cerdà-Alabern, and L. Maccari, "A realistic open-data-based cost model for wireless backhaul networks in rural areas," in *18th International Conference on Network and Service Management (CNSM)*, Core 2023 B, 2022, pp. 55–63 (cit. on p. 146).

[10] 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, JCR 2022 Q2, SJR 2022 Q1 (cit. on pp. 37, 118, 126, 161).

[11] G. Gemmi, R. L. Cigno, and L. Maccari, “WIP: Analysis of feasible topologies for backhaul mesh networks,” in *2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Core 2021 A/B, 2021, pp. 240–243 (cit. on p. 38).

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES



# ACRONYMS

| | |
|---|---|
| 1G | 1st generation |
| 3G | 3rd generation |
| 3GPP | 3rd Generation Partnership Project |
| 4G | 4th generation |
| 5G | 5th generation |

| | |
|---|---|
| AI | Artificial Intelligence |
| AMF | Access and Mobility Management Function |
| API | Application Programming Interface |
| BAP | Backhaul Adaptation Protocol |
| BLER | Block Error Rate |
| BS | Base Station |
| BTS | Base Transceiver Station |
| CapEx | Capital Expenditure |
| CDF | Cumulative Distribution Function |
| CN | Core Network |
| COTS | Commercial Off-the-Shelf |
| CU | Central Unit |
| DAG | Directed Acyclic Graph |
| DEM | Digital Elevation Model |
| DSL | Digital Subscriber Line |
| DSM | Digital Surface Models |
| DTM | Digital Terrain Model |
| DU | Distributed Unit |
| E2SM | E2 Service Model |
| ECDF | Empirical Cumulative Density Function |
| EPDF | Empirical Probability Density Function |
| ESPFP | Elementary SPFP |
| ETSI | European Telecommunications Standards Institute |
| FFT | Fast Fourier Transform |
| FPGA | Field Programmable Gate Array |
| FR2 | Frequency Range 2 |
| FSPL | Free Space Path Loss |
| FTTCab | Fiber-to-the-Cabinet |
| FTTH | Fiber-to-the-Home |
| FWA | Fixed Wireless Access |
| GIS | Geographic Information Systems |
| gNB | Next Generation Node Base |
| GPU | Graphics Processing Unit |
| GTP | GPRS Tunneling Protocol |
| IAB | Integrated Access and Backhaul |
| ICMP | Internet Control Message Protocol |
| ILP | Integer-Linear Problem |
| InH | Indoor Hotspot |
| IoT | Internet of Things |

| | |
|---|---|
| ISP | Internet Service Provider |
| ISTAT | Italian Institute of Statistics |
| ITU | International Telecommunication Union |
| IXP | Internet Exchange Point |
| KPI | Key Performance Indicator |
| KPM | Key Performance Measurement |
| LiDAR | Light Detection And Ranging |
| LoS | Line of Sight |
| LTE | Long Term Evolution |
| MAC | Medium Access Control |
| MCS | Modulation and Coding Scheme |
| MIMO | Multiple Input, Multiple Output |
| ML | Machine Learning |
| mmWave | millimeter wave |
| MNO | Mobile Network Operator |
| MnS | Management Services |
| MT | Mobile Termination |
| mttf | mean-time-to-failure |
| mttr | mean-time-to-repair |
| Near-RT RIC | Near Real-time RAN Intelligent Controller |
| NGWB | Next Generation Wireless Backhaul |
| NLoS | Non-LoS |
| Non-RT | Non-real-time |
| Non-RT RIC | Non-Real-Time Ran Intelligent Controller |
| NRF | Network Repository Function |
| O-RAN | Open Radio Access Network |
| O2I | Outdoor to Indoor |
| OAI | OpenAirInterface |
| OpEx | Operational Expenditure |
| OSM | Open Street Map |
| OTA | Over-The-Air |
| PDU | Packet Data Unit |
| PoP | Point of Presence |
| PRB | Physical Resource Block |
| RAN | Radio Access Network |
| RC | RAN Control |
| RF | Radio Frequency |
| RIC | RAN Intelligent Controller |
| RIS | Reconfigurable Intelligent Surface |

| | |
|---|---|
| RLC | Radio Link Control |
| RMa | Rural Macro |
| RoI | Return on Investment |
| RRC | Radio Resource Control |
| RTT | Round Trip Time |
| RU | Radio Unit |
| RWBDB | Reliable Wireless Backhaul Design Problem |
| SDR | Software-defined Radio |
| SM | Service Model |
| SMF | Slicing Magangement Framework |
| SMO | Service Management and Orchestration |
| SNR | Signal-to-Noise-Ratio |
| SPAL | Shortest Path without Attached Leaves |
| SPFP | Shortest Path with Forbidden Paths |
| SPT | Shortest Path Tree |
| SRN | Standard Radio Node |
| SRTM | Shuttle Radar Topography Mission |
| TDD | Time Division Duplexing |
| TDMA | Time Division Multiple Access |
| UE | User Equipment |
| UL-SCH | Uplink Shared Channel |
| UMa | Urban Macro |
| UMi | Urban Micro |
| UPF | User Plane Function |
| VDSL | Very high-speed Digital Subscriber Line |
| VES | VNF Event Stream |
| VNF | Virtual Network Function |
| WBN | Wireless Backhaul Network |
| WCN | Wireless Community Network |
| WISP | Wireless Internet Service Provider |

# 1 INTRODUCTION

Over the past decade, the increasing demand for wireless communication bandwidth has driven the shift towards higher frequency bands, primarily due to the exhaustion of the lower frequency spectrum. The Extremely High Frequency (EHF) band, spanning from 30 GHz to 300 GHz has recently gathered significant interest. Notably, its lower portion commonly called millimeter wave (mmWave) band, has been included in different technologies such as 802.11ay, standardized by IEEE for personal wireless networks [12], and 5G New Radio (5G-NR), standardized by 3GPP for 5th generation (5G) cellular networks under the name of Frequency Range 2 (FR2) [13]. At the same time, the research community pushing this boundary even further, researching radio communication in the THz band [14].

Even more recently, the unallocated portion of the radio spectrum between 7GHz and 24GHz, known as Frequency Range 3 (FR3), has gained a lot of attention and will be probably a focal point of future communication technologies such as 6G.

While these bands offer enhanced capacity, they also reduce its communication range and resilience to Non-LoS (NLoS) communication. Consequently, traditional network architecture and planning tools for Radio Access Network (RAN) are becoming obsolete. Innovative solutions are imperative for the effective deployment of 5G and subsequent cellular networks.

Addressing the propagation challenges of higher frequencies and augmenting capacity per user requires a significant increase in RAN density. Estimates from both academic [15] and industry [16] sources suggest up to a hundred Base Stations (BSs) per square kilometer. Furthermore, to ensure Line of Sight (LoS) communication, these stations will likely be positioned at street level, on fixtures like lampposts, building exteriors, or traffic lights.

Such an increase of BSs also demands a redeployment of the wired backhaul connecting these devices to the network core. However, deploying such a widespread fiber backhaul network to individual street-level devices is neither economically viable nor feasible for Mobile Network Operators (MNOs).

In this scenario, Wireless Backhaul Networks (WBNs) offer a more flexible approach to cellular deployments. They allow MNO to connect only a fraction of the BSs via fiber and then use wireless communication to link the remaining stations. Such technology has been widely used in the past, with Point-to-point radio links used as a *"wireless cables"* to connect remote BS. With 5G, however, WBN have been integrated and standardized in the RAN with the concept of Integrated Access and Backhaul (IAB) [13]. The

core idea of IAB is that BSs functions can be split into two parts: the majority of the Next Generation Node Bases (gNBs) are IAB-nodes, smaller and suitable for street-level placement without wired connections, while a minority, IAB-donors, resemble traditional base stations and connect to the network via fiber. IAB-nodes and IAB-relays then interconnect in a multi-hop wireless mesh which is used to relay the user traffic to the wired network. It's worth mentioning that since IAB is part of the 5G standard, it uses the same radio technology (5G-NR) used in the access, thus the same frequency used for the radio access can also be used for the wireless backhaul.

Beyond urban settings, WBNs have proven invaluable in rural areas, often overlooked by MNOs and Internet Service Providers (ISPs) due to low population density and reduced profit margins. Here, WBNs serve both mobile access and fixed broadband networks. Technologies like IAB, combined with Non-Terrestrial Networks, offer sustainable backhaul solutions for cellular BSs. Conversely, for fixed broadband, WBNs enable Wireless Internet Service Providers (WISPs) or Wireless Community Networks (WCNs) to bridge the digital divide, deploying wireless equipment on subscriber premises to provide and relay connectivity.

The deployment of such an infrastructure, made of hundreds of radio nodes that need to be interconnected, cannot be performed opportunistically as often done in the past and requires meticulous planning. Traditional modeling techniques, such as stochastic LoS probability and pathloss models, are still employed, but the shift to the EHF spectrum and ultra-dense deployments necessitates more precise models. Geographic Information Systems (GIS) technology can be helpful in this context, together with GPU-based algorithms, to evaluate the algorithms and models for WBNs. Leveraging open geographic datasets, such as Open Street Map (OSM), and high-definition morphological datasets from Light Detection And Ranging (LiDAR) technology, allows for accurate assessments of WBN behavior.

To the best of our knowledge, this thesis is the first comprehensive work on the planning and optimization of WBNs in ultra-dense and rural environments. Its open-source nature ensures that other researchers or network operators can utilize this work for network deployment planning, optimal node placement, topology determination, and real-time network optimization.

The rest of the manuscript is organized as follows:

- Chapter 2 provides an overview of the evolution of mobile and fixed broadband networks and introduces the models, techniques, and algorithms used in this thesis to model WBN.
- Chapter 3 focuses on ultra-dense networks, discussing both the placement of obstacle-free base stations and their interconnection to form WBN.

- Chapter 4 delves into the optimization of IAB network topologies, offering algorithms and techniques for designing robust and energy-efficient topologies.
- Chapter 5 shifts the focus to rural areas, analyzing the technical and economic feasibility of using WBNs to deliver broadband connectivity.

**OPEN-DATA AND REPRODUCIBILITY** Reproducibility stands as a cornerstone in academic research. It ensures that other scholars can validate our conclusions and utilize our resources to advance the field further. In line with this principle, throughout this thesis, we have exclusively employed datasets under open licenses. Additionally, we have made both the code that generated our results and the results themselves publicly available. The specifics of the datasets employed can be found in Section Sect. 2.3. Below, we provide references to all datasets and repositories associated with this work.

| Type | Description | Sect. | URL |
|---|---|---|---|
| Dataset | Optimal placement of the BSs | 3.2 | 1 |
| Code | Visibility analysis GPU algorithms and placement heuristics | 3.2 | 2 |
| Dataset | Traffic model and optimal placement of the BSs for vehicular networks | 3.3 | 3 |
| Code | Placement heuristics | 3.3 | 4 |
| Dataset | Visibility graphs and IAB network topologies | 3.5 | 5 |
| Code | Visibility Graph GPU Algorithms | 3.5 | 6 |
| Emulation Scenarios | Realistic RF Scenarios for IAB | 4.2 | 7 |
| Containers | Containers with OAI and IAB | 4.2 | 8 |
| Code | IAB experiments orchestrator | 4.2 | 9 |
| Code | Code to generate the scenarios | 4.2 | 10 |
| Code | Optimization models implementation | 4.3 | 11 |
| Dataset | IAB topologies | 4.3 | 12 |
| Code | Optimization models implementation | 4.4 | 13 |
| Dataset | IAB topologies | 4.4 | 14 |
| Code and Datasets | Topology design algorithm and results | 5.3 | 15 |

[1] https://zenodo.org/record/6813283
[2] https://github.com/UniVe-NeDS-Lab/TrueBS/
[3] https://github.com/UniVe-NeDS-Lab/TrueBS/tree/vehicular_mod
[4] https://github.com/UniVe-NeDS-Lab/TrueBS/tree/vehicular_mod
[5] https://zenodo.org/record/4905536
[6] https://github.com/UniVe-NeDS-Lab/TrueNets
[7] https://colosseumneu.freshdesk.com/support/solutions/articles/61000303373-integrated-access-and-backhaul-scenarios
[8] experiments.colosseum.net
[9] https://github.com/wineslab/iab-manager
[10] https://github.com/UniVe-NeDS-Lab/ColosseumScenarioGenerator
[11] https://github.com/UniVe-NeDS-Lab/backhaul-topology-optimization
[12] https://github.com/UniVe-NeDS-Lab/backhaul-topology-optimization
[13] https://github.com/wineslab/IABEnergyOptimization
[14] https://github.com/wineslab/IABEnergyOptimization/tree/main/simulator/results_globecom/milano_5_80
[15] https://github.com/UniVe-NeDS-Lab/ODCM

# 2 BACKGROUND

## 2.1 BROADBAND NETWORKS

Broadband networks have rapidly evolved over the last few decades, revolutionizing the way we communicate, work, and entertain ourselves. This technological evolution required provisioning multi-gigabit data rates to billions of people in heavily populated areas, both using wired and wireless technologies. However, the rise of these networks has unveiled a stark digital divide, with rural areas often left behind in the digital race. This has led to the deployment of new kinds of access networks using a mix of wireless and wired technology capable of being economically sustainable even in loosely populated areas.

In this section, we will analyze and explain the different kinds of broadband networks and their evolution in the past decades. We will make a first distinction between *Mobile Networks* which are meant to connect individuals, cars, and mobile devices; and *wired broadband networks* which are meant to connect households and businesses.

### 2.1.1 Mobile Networks

Mobile Access Networks, often simply called mobile networks or cellular networks, can be defined as a telecommunication network whose last-mile connection is provided by a wireless transceiver, also called Base Station (BS) which communicates with a Mobile Termination (MT), often called User Equipment (UE).

Typically, multiple BSs get distributed in *cells* over an area, and together they provide radio coverage. Moreover, thanks to signaling between the MTs and the UEs, terminals that are moving through multiple cells get automatically reconnected to the strongest cell in a process called *handover*. This allows users to be able to perform voice calls and maintain data connection even while walking, driving, or on high-speed vehicles such as trains.

#### 2.1.1.1 *From the First to the Fifth Generation*

Introduced in the late 1970s, the 1st generation (1G) of mobile network represents the pioneering era of mobile communications. Operating on analog signals, 1G primarily facilitated voice calls. With limited capacity and susceptibility to eavesdropping, these networks laid the foundation for future mobile telephony innovations.

Emerging in the early 1990s, the second generation, known as 2G, marked the transition from analog to digital communication. This dig-

ital shift, characterized by technologies such as the Global System for Mobile Communications (GSM), introduced features like text messaging (SMS) and basic data services, typically reaching speeds up to 144 kbps. Enhanced encryption in 2G also provided better security, minimizing call interception risks.

By the early 2000s, the demand for more advanced services rose and the 3rd generation (3G) emerged. Offering faster data transfer rates, often up to 14 Mbps, 3G networks enabled internet browsing, video calling, and multimedia streaming. This generation utilized a broader spectrum, essentially setting the stage for the ubiquity of mobile internet and applications. The following evolution led to the 4th generation (4G) around 2010. Designed for high-speed mobile internet and utilizing technologies like Long Term Evolution (LTE), 4G networks delivered data transfer speeds often reaching up to 150 Mbps, if not more. This surge in speed and capacity enabled seamless video streaming, online gaming, and a plethora of data-intensive applications on mobile devices.

As the need for higher data rates increased, fueled by applications like augmented reality, high-resolution video streaming, and the Internet of Things (IoT), mobile networks have had to adapt in innovative ways. In 2019 the 5th generation (5G), standardized by the 3rd Generation Partnership Project (3GPP) in Release 15 [17], introduced several new bands. In the sub-6Ghz range, several bands were introduced. Such as the n78 with a range from 3.3GHz to 3.8GHz and the band n79 with a range from 4.4GHz to 5GHz. Moreover, a whole new frequency range, called Frequency Range 2 (FR2), commonly called millimeter wave (mmWave) was introduced. This new band, with frequencies above 24GHz, allows for larger channel bandwidths up to 2GHz.

Even more recently a new frequency range, called FR3, has gained the attention of the telecommunication industry, and will possibly be included in the next 3GPP release. This range lies between FR1 and FR2 specifically between 7GHz and 24 GHz. Among its advantages, there is a less challenging propagation than FR2 frequencies and enough unallocated frequencies to support the increasing demands for bandwidth.

These higher frequency bands can transmit vast amounts of data at incredible speeds. However, they come with a limitation: shorter transmission ranges compared to lower frequencies and less penetration. This necessitates a radical change in the network architecture, with different solutions on multiple levels: On the hardware level, antennas with higher gain are needed. This is made possible by taking advantage of massive Multiple Input, Multiple Output (MIMO) antenna arrays, with up to 256x256 antenna elements that enable the use of beamforming to increase the directionality of the antennas and reduce the interference from other transmissions.

On a network design level, the number of BS needs to grow dramatically and their planning needs to be optimized to guarantee Line of Sight (LoS) communication. While this poses infrastructural and economic challenges, the payoff in terms of data rates and reduced latency is immense, 5G in

fact aims to reach multi-gigabit data rates and an in-network delay down to 1ms.

#### 2.1.1.2 Current Architecture

The 5G network architecture is a complex arrangement of interconnected elements, designed to offer a multitude of services with varying requirements. Fig. 2.1 gives an overview of such architecture. At the user end, the UE (or MT) serves as the primary interface for 5G services. The UE connects to the network via the 5G New Radio (5G-NR), which operates on a broad spectrum ranging from sub-6 GHz frequencies to mmWave.

The next layer of the architecture is the Next Generation Node Base (gNB), the 5G equivalent of the BS. The UEs connect to the gNB and routes data to and from various network functions. The gNBs themselves are interconnected through the so-called Xn interface, allowing seamless user mobility and data exchange.

The 5G architecture introduces a more flexible and scalable approach to networking by decoupling the control plane and the user plane, unlike earlier cellular technologies. The control plane is responsible for setting up the connections, routing, and overall network management. The user plane deals with the actual data transmission.

The gNBs connect to the 5G core network through the NG interface, which carries both control-plane and user-plane data. In the core, several key functions take place, most notably:

- Access and Mobility Management Function: Handles UE's mobility and access authentication.
- Session Management Function: Manages the data sessions.
- User Plane Function: Responsible for packet routing and forwarding.
- Network Slice Selection Function: Assists in network slicing, directing UEs to the appropriate slice based on their service requirements.

These core functions are interconnected through various interfaces, like N2 and N3, facilitating robust, high-speed data services for UEs. By

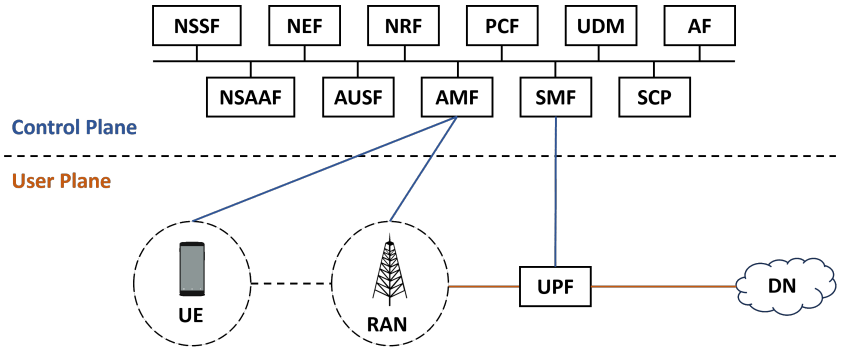



Figure 2.1: Architecture of a 5G network, detailing the main interfaces between the Radio Access Network (RAN) and the core.

segregating functions, adopting software-centric approaches, and through the use of containerization, the 5G core architecture aims for agility and scalability, enabling quick adaptation to varied service demands.

### 2.1.1.3 *Integrated Access and Backhaul (IAB)*

Traditional backhaul solutions, in which fiber is deployed to each gNB location, are limited in terms of scalability and costs. Already in pre-5G networks, Point-to-Point out-of-band wireless backhauls were used to connect remote or inaccessible locations. With the 3GPP Release 16th of the standard, native support for wireless backhaul has been introduced. Specifically, with IAB the same protocol and waveform of 5G are used to self-backhaul the BS through a multi-hop wireless network. IAB, in fact, defines two different kinds of gNB: IAB-donors and IAB-nodes. The formers are, like in traditional deployments, connected using fiber; the latter instead, can connect directly to the IAB-donors using the 5G protocol stack, or to other IAB-nodes forming a backhaul wireless mesh network. Then, over the air, the IAB traffic is encapsulated using the Backhaul Adaptation Protocol (BAP), a protocol standardized by 3GPP specifically for IAB [18].

Fig. 2.2 depicts an example with two UEs, two IAB-nodes and one IAB-donor, together with the protocol stack.

Note that while 3GPP standardized the most technical aspects of IAB, many system-level aspects are left under the control of the Mobile Network Operator (MNO). One important aspect that's subject to different interpretations and that radically affects our research, is whether both access and backhaul should share the same radios and spectrum or not.

Sharing the same radio resources among the backhaul and access networks is beneficial in economic terms, as less spectrum and radios are needed to run the network. However, this comes at the cost of reducing the capacity of the mesh backhaul and thus the depth of the IAB tree. This, in turn, requires more IAB-donors to be deployed and thus increases the

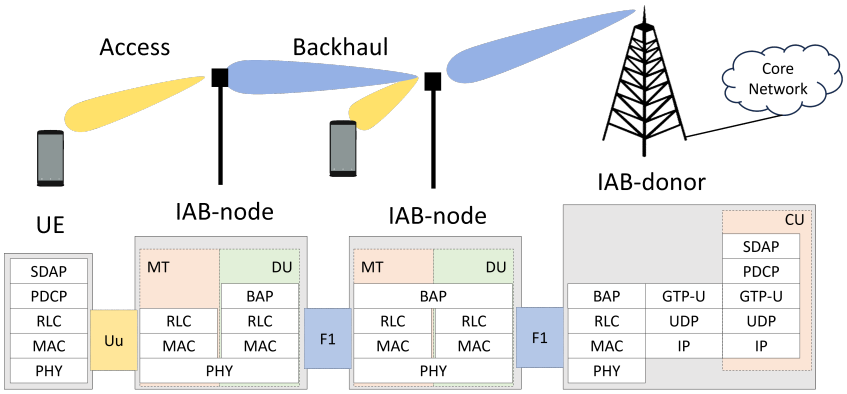



**Figure 2.2:** Architecture of an IAB network, with details on the IAB protocol stack.

upfront cost that the MNO has to undertake to deploy the network, also known as Capital Expenditure (CapEx).

#### 2.1.1.4 Open RAN

Open-RAN (or O-RAN) represents a significant shift in the design and deployment of mobile networks. Unlike traditional RAN architecture, which tightly couples the hardware and software components, O-RAN aims to decouple these elements. This allows for more modular networks with interoperable components, reducing dependency on a single vendor and potentially lowering costs. [19] The initiative is backed by the O-RAN Alliance and the Telecom Infra Project, among other industry groups, and it has received considerable attention as an enabler for more flexible, efficient, and scalable network architectures.

In an O-RAN environment, the functionalities of the radio access network are virtualized and run on general-purpose hardware. This opens the door to employing software-based, cloud-native technologies that can be more easily updated and scaled. It enhances the network's ability to support features like network slicing, real-time analytics, and machine learning algorithms for network optimization. One of the significant advantages of O-RAN is the facilitation of multi-vendor environments. It allows operators to mix and match products from different vendors at a modular level—something nearly impossible in traditional RAN settings since components from different vendors often had compatibility issues. Following the disaggregation paradigm proposed in the 3GPP Rel. 17 [13], the traditional monolithic BS, is split in three different functional units, the Radio Unit (RU), the Distributed Unit (DU), and the Central Unit (CU) (O-RU, O-DU, and O-CU in the O-RAN terminology).

Fig. 2.3 exemplifies the network architecture, with the main building blocks and interfaces.

The higher-level component of the stack is the CU, which takes care of the higher level of the 3GPP stack. This component can be virtualized on general-purpose servers and deployed at the edge.

Below the CU, we have the DU, which takes care of high-level physical layer functionalities and of the Medium Access Control (MAC) and Radio Link Control (RLC) layers. This component requires more specialized hardware components and is generally virtualized on general-purpose

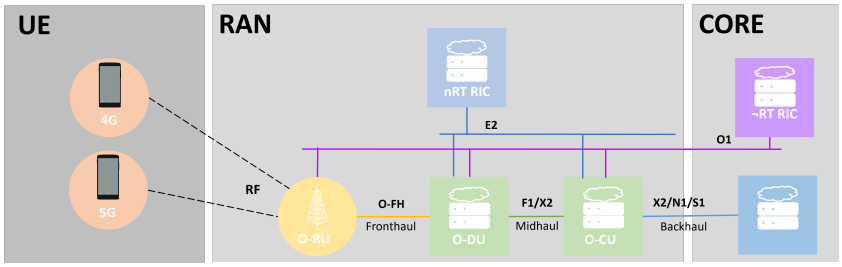



**Figure 2.3:** Disaggregated O-RAN architecture.

machines with hardware accelerators such as Graphics Processing Units (GPUs). It's worth noting that different companies are working on developing solutions in this field, such as NVIDIA with the Aerial platform [20] and Intel with FlexRAN [21].

Finally, deployed in physical proximity of the antenna and the Radio Frequency (RF) circuitry we have the RU, this component is in charge of lower-phy functionalities, such as Fast Fourier Transform (FFT), equalization, cyclic prefix addition/removal, etc. It is often generally implemented using Field Programmable Gate Arrays (FPGAs) or Application-Specific Integrated Circuits (ASICs), making it a relatively inexpensive and simple device.

One of the additional innovations in O-RAN is the RAN Intelligent Controller (RIC). The RIC is an intelligent software management layer that operates in conjunction with the RAN. Its primary role is to optimize the RAN performance by using realtime analytics and machine learning algorithms. It enhances the adaptability and automation of the RAN, enabling dynamic resource allocation, automated network slicing, and other advanced functionalities that make the network more resilient and efficient.

Two different RICs have been defined in O-RAN: the Near-realtime RIC and the Non-realtime RIC. The first one is deployed on the network's edge and interacts with the RAN with a time-scale between 10ms and 1s; the second one operates on longer time scales (higher than 1s) and can be deployed in the core of the network. On the two RICs, custom optimization algorithms, called *apps*, are run. In the case of the Near-realtime RIC, these are called xApps and interact with the RAN using the E2 interface, while in the case of the Non-realtime RIC, these are called rApps and use the O1 interface.

As the standardized interfaces and the concept of *apps* have enabled third-party algorithms to interact with the RAN, a thriving area of research has emerged with different apps concepts proposed and evaluated. These apps' goal ranges from optimizing the allocation of resources through slicing [22] to optimizing spectrum sharing among different operators [23].

### 2.1.2 Fixed Broadband Networks

The evolution of wired broadband networks has been marked by progressive advancements to accommodate increasing data transmission requirements. Initially, dial-up technology was prevalent, utilizing telephone lines to provide speeds up to 56 kbps. Subsequently, Digital Subscriber Line (DSL) technology was introduced to fulfill the demand for enhanced bandwidth. Depending on its variant and line quality, DSL could provide speeds ranging from 128 kbps to 24 Mbps. The pursuit for even higher speeds led to the integration of Fiber-to-the-Cabinet (FTTCab). In this configuration, fiber-optic cables reach street-level cabinets, after which Very high-speed Digital Subscriber Line (VDSL) technology, using the

traditional copper lines is employed, with data-rates ranging between 50 Mbps to 100 Mbps. Advancing further, Fiber-to-the-Home (FTTH) became a benchmark in broadband technology. With fiber-optic cables running directly to individual residences, FTTH can achieve data rates exceeding 1 Gbps, often reaching up to 10 Gbps in some implementations.

### 2.1.3 Digital Divide

The evolution of fixed broadband technology highlights an associated issue: the digital divide. This refers to the unequal distribution of advanced digital technologies, including high-speed internet, across different regions and demographics. Urban areas, due to infrastructural investments and demand concentrations, often benefit from the latest broadband technologies. Meanwhile, rural and economically disadvantaged regions may remain reliant on older technologies or lack internet access entirely. This discrepancy poses significant implications for education, healthcare, and economic opportunities in underserved areas. Addressing the digital divide is crucial to ensure that the benefits of technological advancements are accessible to all segments of the population.

In the following sections, we will analyze different network solutions to overcome the digital divide in rural areas. First, we will explain the phenomena of Wireless Internet Service Providers (WISPs), then Wireless Community Networks (WCNs), and finally Satellite Broadband.

#### *2.1.3.1 WISPs*

WISPs serve as a crucial bridge in mitigating the broadband connectivity gap prevalent in rural and remote areas. They offer a cost-effective alternative by utilizing wireless technologies to provide broadband services. These providers typically employ a network of strategically placed wireless antennas, usually on mountains or places with high visibility, to provide broadband connectivity to individual homes or businesses.

The initial CapEx is considerably lower than traditional wired networks, and the modular nature of wireless technology allows for incremental expansion as demand increases.

Albeit in some cases WISPs use licensed bands or proprietary wireless technology, generally, standard 802.11 Wi-Fi equipment is preferred, as it does not require any license and the devices are available at competitive prices. Similarly to the other kinds of broadband networks seen before, data rates for these networks have grown systematically following the different Wi-Fi generations. Reaching hundreds of Mbps with the latest generations.

#### *2.1.3.2 WCNs*

WCNs further complement the efforts to address the digital divide. Rooted in community-driven initiatives, WCNs are set up, operated, and maintained by local communities, often without significant commercial intent.

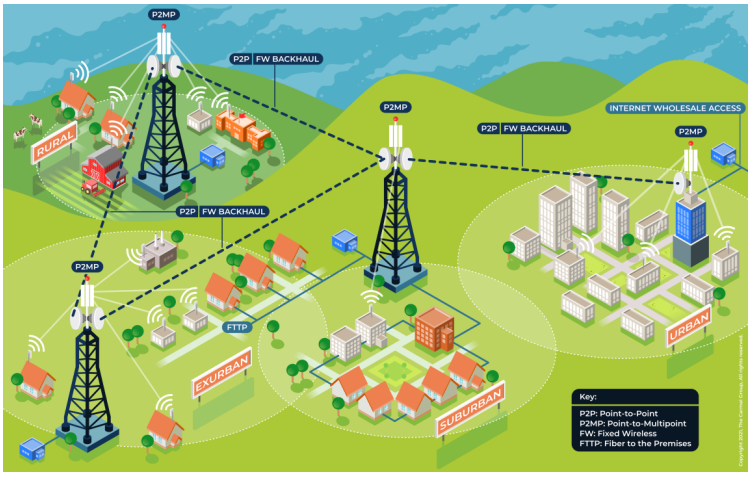



**Figure 2.4:** Graphical depiction of the architecture of a typical Wireless Internet Service Provider (WISP). Source `wispa.org`

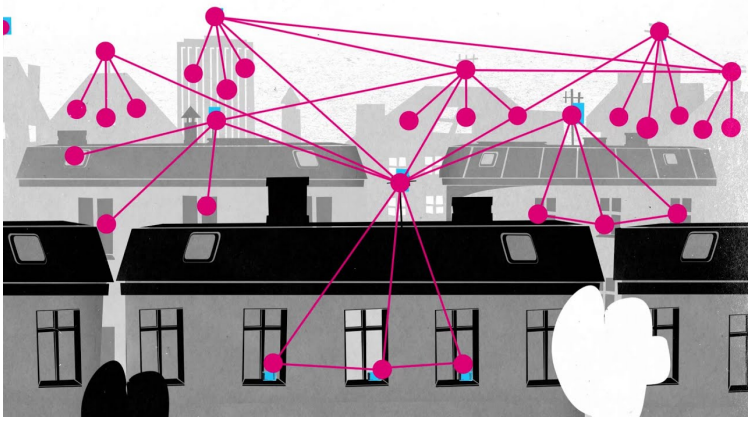


**Figure 2.5:** Graphical depiction of a Wireless Community Network (WCN). Source `freifunk.net`

These networks empower communities to take charge of their connectivity needs, ensuring that local requirements and challenges are at the forefront of network design and operation. By leveraging shared resources, open technologies, and collaborative approaches, WCNs can provide affordable and locally tailored internet solutions. They not only serve as a means to provide connectivity but also foster community engagement, digital literacy, and local innovation. In essence, while Fixed Wireless Access (FWA) offers a top-down approach by extending established networks to underserved areas, WCNs provide a bottom-up solution, cultivating connectivity from within the community itself.

Differently from WISPs, where a centralized architecture is deployed, WCNs follows a more organic development, where older nodes get upgraded with multiple radio devices to allow new nodes to be added to the network. This architecture results in a more robust network, which can sustain the failure of multiple nodes before affecting the users' performance.

The most notable examples of WCN are Guifi.net and Freifunk two community-led initiatives with more than 20 years of history. Guifi.net[1] is based in Catalunya and its network made of both fiber and wireless backhauls connects more than 37.000 nodes across the whole region of Catalunya. Freifunk[2] is based in Germany and is composed of various dozen of smaller network, roughly one for each major German city.

#### *2.1.3.3 Satellite Broadband Connectivity*

Satellite broadband connectivity has been an existing solution in remote areas since the early 2000s. This technology can be broadly categorized based on satellite altitude: Geostationary (GEO) and Low Earth Orbit (LEO).

**GEO** Early satellite solutions utilized a parabolic dish mounted on the user's rooftop, aligned with a Geostationary satellite, accompanied by a ground station connected to the internet. The primary limitation of GEO-based systems is high latency, a consequence of the significant propagation delay. Given that GEO satellites are stationed approximately 36,000 km above Earth, the one-way propagation delay is about 120 ms, resulting in a round-trip delay of around 250 ms.

This technology branched into two main models: one-way and two-way satellite connectivity. In the one-way model, the satellite link was dedicated to the downlink channel, with the uplink managed through a DSL or dial-up connection via the telephone line. This method enabled the use of existing satellite dishes, often already installed for television, with the ISP broadcasting the downlink channel via a commercial satellite.

The two-way model required a more sophisticated and costly setup, including a satellite modem and transceiver, as it involved both transmitting

[1] https://guifi.net
[2] https://freifunk.net/en/

and receiving data via the satellite link. This approach, while more expensive due to the complexity of the equipment and increased bandwidth usage, offered connectivity in areas devoid of DSL infrastructure.

**LEO** LEO satellites have emerged as an alternative to GEO satellites in providing broadband connectivity in the late 2010s. Due to the reduced orbit, which is usually around 500km over the earth's surface, the latency greatly improves. However, the satellites move over the user with great speed and thus traditional satellite dishes cannot be used. Moreover, to guarantee uninterrupted service a large constellation of satellites is needed.

Prominent companies in this domain include SpaceX with its Starlink project, Amazon with Project Kuiper, and OneWeb. Specifically, Starlink has deployed over 5000 LEO satellites, providing extensive coverage across most of Europe and North America. The user equipment for Starlink involves a complex MIMO array with 1464 antenna elements, operating in the mmWave band (12-90GHz), to facilitate high-speed data transmission, which reaches 250Mbps in downlink.

## 2.2 CHANNEL MODELING

In the upcoming chapters of this Thesis, different contributions will be evaluated in simulation using a wide array of channel models. This section aims to detail all those different channel models, ranging from the simplest to the most complex and accurate ones.

In the first section, we detail the different pathloss models, from Free Space Path Loss (FSPL) to more complex stochastic models. In the following section, we move to describe two LoS probability models. Then we move to link budgeting and finally, we detail three capacity models, both technology agnostic and specific.

### 2.2.1 Propagation Models

#### 2.2.1.1 FSPL

The FSPL model is used to predict the loss of power over the distance in free space, without considering any reflection or refraction. This model is particularly useful in the context of wireless communications, especially when analyzing line-of-sight paths between the transmitter and receiver in an open environment. The formula to calculate the FSPL, given in decibels (dB), is represented as:

$$PL(d) = 20\log_{10}\left(\frac{4\pi d f_c}{c}\right) \tag{2.1}$$

where $d$ is the distance between the transmitter and receiver in meters, $f_c$ is the carrier frequency of the signal (in Hz), and $c$ is the speed of light.

#### 2.2.1.2 Log-Normal Pathloss Model

One of the most used models in signal propagation is the *Log-distance* model [24], which assumes an exponential relationship between pathloss and distance. The formula of the model is the following:

$$PL(d) = L_0 + 10\alpha log(d) + \mathcal{X} \tag{2.2}$$

Where $L_0$ is the, obtained by measuring the pathloss at a reference distance (usually 1m); $\alpha$ is the pathloss exponent, which indicates the rate at which the pathloss increases with distance; and $\mathcal{X}$ is a zero-mean Gaussian random variable (in dB) which accounts for shadowing effects. The estimation of the pathloss exponent $\alpha$ is crucial for the accuracy of the model and many studies have been made in this regard [24] [25]. Common values for it are 2, in the case of a Free Space LoS link, between 2.7 and 3.5 for Non-LoS (NLoS) links.

#### 2.2.1.3 ITU Channel Model

More recently, in an endeavor led by 3GPP and European Telecommunications Standards Institute (ETSI), a model to predict the propagation of

wireless links between 0.5 and 100GHz has been proposed [26] [27]. This model encompasses four distinct environmental scenarios: Urban Micro (UMi), Urban Macro (UMa), Rural Macro (RMa), and Indoor Hotspot (InH). Furthermore, it is parametrized based on the LoS condition of the wireless connection.

For the scope of subsequent chapters, we focus on the UMi environment and present the associated formulas, beginning with the LoS scenario.

$$PL_{LoS}^{UMi}(d) = \begin{cases} PL_1(d) & 10m \leq d_{2D} \leq d_{BP} \\ PL_2(d) & d_{BP} \leq d_{2D} \leq 5km \end{cases} \tag{2.3}$$

where $d_{BP}$ is the Breakpoint distance and is defined as: $d_{BP} = 4h_{\text{BS}}h_{\text{UT}}\frac{f_c}{c}$ and $h_{\text{BS}}$ and $h_{\text{UT}}$ are respectively the height of the Base Transceiver Station (BTS) and the height of the UE above the ground level (in meters).

Then, $PL_1$ and $PL_2$ can be defined as follows:

$$PL_1(d) = 32.4 + 21\log_{10}(d_{3D}) + 20\log_{10}(f_c) \tag{2.4}$$

$$\begin{aligned} PL_2(d) =& 32.4 + 40\log_{10}(d_{3D}) + 20\log_{10}(f_c) \\ & - 9.5\log_{10}((d_{BP})^2 + (h_{\text{BS}} - h_{\text{UT}})^2) \end{aligned} \tag{2.5}$$

For the NLoS case, the formulas are the following:

$$PL_{NLoS}^{UMi}(d) = \max(PL_{LoS}^{UMi}(d), PL_3(d)) \tag{2.6}$$

$$\begin{aligned} PL_3(d) =& 22.4 + 35.3\log_{10}(d_{3D}) + 21.3\log_{10}(f_c) \\ & - 0.3(h_{\text{UT}} - 1.5) \end{aligned} \tag{2.7}$$

Additionally, this model includes a zero-mean Gaussian random variable $\mathcal{X}$ which accounts for the shadow fading effects. The std of the random variable (in dB) depends on the LoS state of the link. For LoS $\sigma_{\mathcal{X}} = 4$, while for NLoS $\sigma_{\mathcal{X}} = 7.82$.

Fig. 2.6 offers a comparative analysis of various channel models for transmissions at $3.8GHz$ over distances less than $1km$. Notably, the FSPL model serves as a lower bound, exhibiting the lowest pathloss with the UMi model for LoS links being the closest to it. Then, by increasing the pathloss exponent, or by using a NLoS model the pathloss dramatically increases with differences up to 50dB.

### 2.2.2 LoS Probability Models

As shown in the models above, the state of the link can change quite significantly the parameters of the propagation models. For this reason, estimating the state of the link correctly is crucial to effectively model the propagation. Over time, different models have been proposed to predict the LoS probability. Here we detail two models that have been employed in this thesis.

#### 2.2.2.1 WINNER II / ITU Model

One of the most widely recognized and utilized models originated from the Winner II project [28], which was later expanded upon and endorsed by both International Telecommunication Union (ITU) [27] and ETSI [29]. These models depict the LoS probability across various environments as a function of distance, exhibiting an exponential decay.

In this thesis, we predominantly reference and employ the ITU model for the Urban-Micro environment (UMi), as well as the Winner II models for suburban (SU) and rural macro (RMa) contexts. For additional formulas and detailed models, the original documentation serves as a comprehensive resource.

$$P_{LoS}^{UMi} = \begin{cases} 1 & d < 18m \\ \frac{18}{d} + e^{-\frac{d}{36}}(1 - \frac{18}{d}) & d > 18m \end{cases} \tag{2.8}$$

$$P_{LoS}^{\text{SU}} = e^{-\frac{d}{200}} \tag{2.9}$$

$$P_{LoS}^{RMa} = e^{-\frac{d}{1000}} \tag{2.10}$$

#### 2.2.2.2 Al-Hourani Model

The second model we consider is provided by Al-Hourani [30] and is intended to estimate the LoS probability between two flying objects at height $h_1$ and $h_2$ respectively, at a certain distance $d$. Distinct from the earlier mentioned model, Al-Hourani's approach takes advantage on terrain morphology data sourced from Digital Elevation Model (DEM) or Open Street Map (OSM), specifically utilizing the building height's Cumulative Distribution Function (CDF) to tailor the model for specific environments.

The probability of LoS is articulated as:

$$P_{\text{LoS}}^{\text{AL-H}}(d, h_1, h_2) = \exp\left(-2r_0\lambda_0 \int_0^{d-\frac{\pi}{2}r_0} G(h)dx\right) \tag{2.11}$$

Herein, $r_0$ denotes the average building's radius, approximated as cylinders, $\lambda_0$ represents the average building density, and $G(h)$ the building height's complementary CDF. The variable $h$ represents the height of the LoS at the integration partition and can be obtained as follows:

$$h = \frac{x}{d}(h_2 - h_1) + h_1 \tag{2.12}$$

It is worth noting that, unlike the Winner II / ITU model, the Al-Hoourani model requires detailed knowledge of the properties of the area where it is applied. Without such precise data the model can not be effectively used.

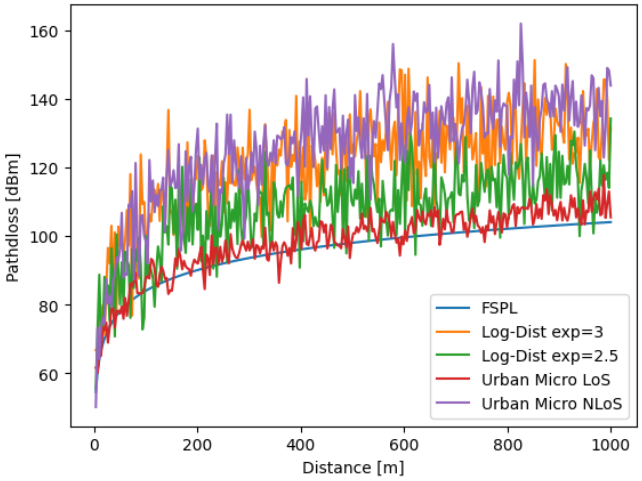



**Figure 2.6:** Comparison of the different pathloss models from 0 to 1000m.

### 2.2.3 Raytracing

Ray tracing is a method used to model RF signal propagation in a given environment [31]. It maps how RF signals, emitted from a transmitter, interact with surrounding structures and elements. The signals may reflect, refract, or diffract based on the encountered obstacles before reaching a receiver. Ray tracing is particularly suited at capturing multi-path propagation, where signals can take multiple trajectories to the receiver, each with its own attenuations. Its applications span cellular network design, indoor wireless network planning, and broadcasting coverage prediction. The technique is computationally demanding, especially for intricate environments or high-frequency signals, but recent advancements in parallel computing and the broad availability of GPUs have made it possible to perform raytracing analysis on large environments.

Various software applications utilize ray tracing algorithms for RF propagation. Among the most renowned is Wireless InSite[3], though its

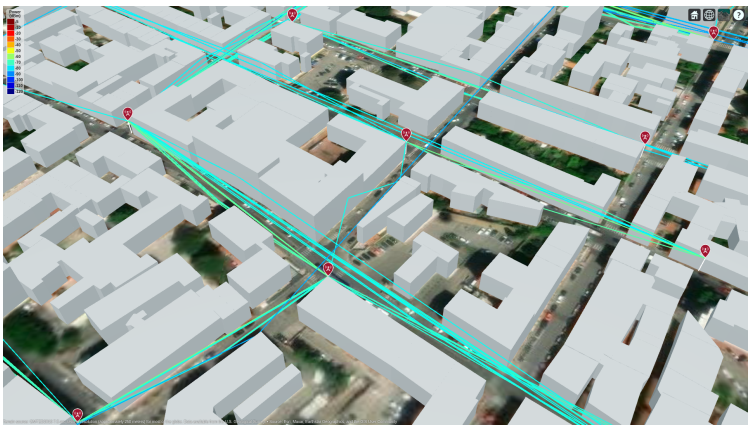


**Figure 2.7:** Raytracing output using Matlab.

[3]https://www.remcom.com/wireless-insite-em-propagation-software

licensing costs are prohibitively high. Recently, Matlab introduced its ray tracing toolkit[4], which we adopted for this thesis. Additionally, Nvidia has recently released their open-source ray tracer called Sionna RT[5], which is optimized for their GPUs [32].

Given a pair of transceivers $(s, d)$, and a 3D model of the environment we can perform raytracing to obtain a set $\Gamma_{s,d}$ of rays. Associated to each ray $r \in \Gamma_{s,d}$ is a pathloss $P_r$, a phase $\phi_r$, a delay $\delta_r$, an angle of arrival $AoA_r$, and an angle of departure $AoD_r$.

By combining these rays, as shown in the following equation, it is possible obtain an overall pathloss, that takes into account the delay and the phase offsets, with the following formulae:

$$PL(s, r) = \sum_{r \in \Gamma_{s,r}} \sqrt{P_r} e^{j(-2\pi\delta_r f_c + \sigma_r)} \tag{2.13}$$

#### 2.2.3.1 MIMO Channel Modeling

A more generalized model for the channel can be adopted by computing the channel matrix $\mathbf{H}$ from the set of rays $\Gamma_{s,d}$ [33]. This allows us to model MIMO channels with arrays of arbitrary size.

$$\mathbf{H} = \sum_{r \in \Gamma_{s,r}} \sqrt{P_r} e^{j(-2\pi\delta_r f_c + \sigma_r)} \mathbf{a}_{rx}^{\star}(AoA_r) \mathbf{a}_{rx}^{H}(AoD_r) \tag{2.14}$$

Where $\mathbf{a}_{rx}$ and $\mathbf{a}_{tx}$ are respectively the receiver and transmitter array responses, $^\star$ is the conjugate operator and $^H$ is the hermitian operator.

Once obtained the channel matrix $\mathbf{H}$ we use it to compute the pathloss $PL(\mathbf{H}, s, d)$ as:

$$PL(\mathbf{H}, s, d) = \mathbf{w}_{s,d}^{T} \mathbf{H}_{s,d} \mathbf{w}_{d,s} \tag{2.15}$$

where $w_{s,d}$ ($w_{d,s}$) is the beamforming vector used by the device $s$ ($d$) to communicate with the device $d$ ($s$), obtained by applying singular value decomposition on the channel matrix $\mathbf{H}_{s,d}$ ($\mathbf{H}_{d,s}$).

### 2.2.4 Link budget and Signal-to-Noise-Ratio (SNR) computation

The Signal-to-Noise-Ratio (SNR) is a measure often used in engineering to compare the level of the desired signal to the level of the background noise. It is a ratio between the power of a signal and the power of the background noise.

$$SNR = \frac{S_r}{N} \tag{2.16}$$

In order to compute the SNR we need to first obtain the signal level $S$. This is achieved by conducting a link budget analysis, which incorporates all the various gains and losses, ultimately determining the received signal power $S_r$. Mathematically, this is expressed as:

[4] https://it.mathworks.com/help/antenna/ref/rfprop.raytracing.html
[5] https://nvlabs.github.io/sionna/api/rt.html

$$S_r = \Pi G_t G_r PL() \tag{2.17}$$

Here, $\Pi$ stands for the transmitter output power (in mW), $G_t$ for the transmitter antenna gain, $PL()$ for a generic pathloss function[6], and $G_r$ for the receiver antenna gain. While here we report the formula in using linear unit of measurements (W), it's often used an reported in logarithmic units of measurements (dB).

Subsequently, to determine the noise power level $N$ we generally compute the thermal noise at a given frequency and temperature, to then add the so-called *noise figure*, which represents the additional noise introduced by the receiver. The formula to calculate the noise is:

$$N = N_0 B N_f \tag{2.18}$$

where $N_0$ represents the spectral noise density (in W/Hz) at a given temperature (commonly set at 290K), $B$ defines the transmission bandwidth, and $N_f$ denotes the receiver's noise figure.

### 2.2.5 Capacity Models

Upon determining the received power between a transceiver pair and their SNR, the next logical step typically involves calculating a metric that encapsulates the prospective capacity of the wireless link in terms of bits/s. Numerous capacity models have been formulated over time, with some specifically tailored to certain technologies, while others maintain a broader applicability.

#### *2.2.5.1 Shannon Limit*

As a first approximation, the Shannon capacity model is often used. This model, named after Claude Shannon, provides a theoretical maximum data rate for a communication channel with a specified bandwidth and noise level [34]. This limit, often referred to as the Shannon limit, delineates the highest possible data rate at which information can be transmitted error-free over the channel assuming gaussian noise. The capacity $C$ between a pair of tranceivers $(s, d)$ can be expressed as a function of the SNR using the formula:

$$\boldsymbol{C}_{\mathrm{SH}}(d) = \mu B \log_2 \left(1 + \frac{S}{N}\right) \tag{2.19}$$

Herin, $\frac{S}{N}$ is the SNR, $B$ is the transmission bandwidth in hertz, and $\mu$ is the number of parallel MIMO streams.

It is worth noting that achieving rates close to the Shannon limit requires sophisticated modulation and coding schemes.

[6] Whenever $PL()$ is expressed in $dB$ and not in linear terms such as in the first equations of the section, it has to be converted in linear term as: $10^{(-\frac{PL()}{10})}$

Despite the Shannon channel model being very often used in this area of research to evaluate the performances of wireless networks, in practice, the capacity of such networks is deeply affected by the technology and the hardware used to implement it. For this reason, in this thesis we have often adopted more sophisticated models derived from real-world hardware.

#### 2.2.5.2 802.11ac capacity

For the 802.11ac technology, we have relied on the datasheets of commercial devices from well-known manufacturers [35]. These datasheets usually contain a table indicating the expected data rate and Modulation and Coding Scheme (MCS) for a given signal quality, expressed either in terms of SNR or directly in received power. Let the table $T_{Wi-Fi}$ be made of tuples $(S_i, M_i)$.

We then compute the maximum MCS $M$ as:

$$M = \max M_i \ \text{ s.t. } \ S_r \geq S_i. \tag{2.20}$$

Having obtained the MCS of the transmission we can look up at the VHT-MCS table [36] to obtain the capacity $C_{Wi-Fi}(M, \mu, B)$ as a function of the MCS, the number of spatial streams ($\mu$) and the bandwidth ($B$).

#### 2.2.5.3 5G-NR capacity

For 5G-NR, finding such tables from the datasheets is not possible, as the standard allows many more combinations of parameters, and the devices are not intended to be purchased by end users. However, thanks to the open-source implementation of both the gNB and UE released by the OpenAirInterface project [37] we were able to have access to the code containing a similar table that matches the SNR to a given MCS. Let the table $T_{5G}$ be made of triplets $(S_i, M_i, E_i)$ that matches a certain SNR $S_i$ with a given MCS $M_i$ and a desired Block Error Rate (BLER) $E_i$ [38]. In our analysis we have used a maximum BLER equal to 0.1, but the value can be adapted to the reliability requirements of any specific study.

Then we can compute the maximum MCS $M_max$ as follows:

$$M_{max} = \max M_i \ \text{ s.t. } \ E_i < 0.1 \ \text{ and } \ S_{s,r} \geq S_i. \tag{2.21}$$

Finally, we compute the downlink capacity $C$ using the following formula:

$$C_{5G} = \mu Q(M_{max}) R(M_{max}) \frac{12RB}{T_u} (1 - Oh) R_{slot}, \tag{2.22}$$

where $\mu$ is the number of MIMO layers, $Q(M)$ and $R(M)$ are two functions associating the MCS to the modulation order and the code rate, $RB$ is the number of Resource Blocks used, $Oh$ is the control channel overhead, $R_{slot}$ is the ratio of Downlink to Uplink slots used, and $T_u$ is the average duration of an OFDM symbol. Further details regarding the formula and the different values can be found in the 3GPP technical specifications [39].

### 2.2.6 Conclusions

In this section, we have shown an array of models and techniques to estimate the capacity of a wireless link, with different degree of accuracy and applied to different technologies. Besides ray tracing techniques, which often require expensive hardware and licenses, most techniques to estimate the pathloss between two transceivers are stochastic models that rely on the LoS conditions, either through a specific parameter, such as in the ITU model, or by tweaking the pathloss exponent, such as in the log-distance model. For this reason, achieving a precise estimation of the LoS is crucial to reach a realistic estimation of the pathloss and thus of the capacity.

In support of this claim, we report in Fig. 2.8 the estimation of the capacity of three different kinds of wireless technologies: sub6, with $f_c = 3.8\,\text{GHz}$ and $B = 100\,\text{MHz}$; mmWave, with $f_c = 27\,\text{GHz}$ and $B = 1\,\text{GHz}$; and subTHz $f_c = 100\,\text{GHz}$ and $B = 3\,\text{GHz}$. The capacity is then further differentiated for the two different conditions: LoS and NLoS. The capacity has been computed using the Shannon capacity model and the ITU Pathloss model.

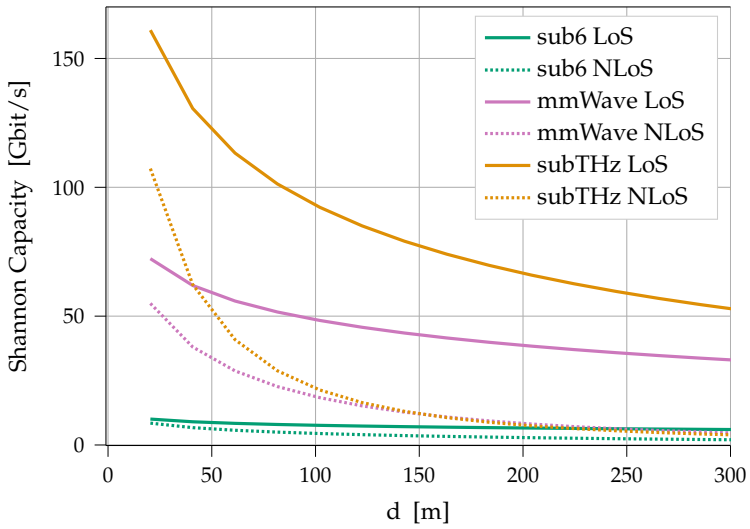



**Figure 2.8:** Shannon channel capacity as a function of the distance between two points $d$ and the LoS/NLoS conditions, for different kinds of technologies.

From this analysis, we can clearly see how by increasing the carrier frequency the difference in terms of channel capacity for LoS and NLoS channels varies extremely. Specifically at a distance of 100 m the ratio of the capacity between NLoS and LoS is equal to 0.59 for the sub6 technology, 0.38 for the mmWave technology, and 0.24 for the subTHz technology.

In conclusion, it's clear that a realistic estimation of LoS conditions, which could have been neglected for traditional technologies, is going to become of utmost importance with the advancements towards THz communications. For this reason, in the next section, we will detail how

we employed real-data to obtain a realistic estimation of LoS and NLoS conditions.

## 2.3 TOWARDS MORE REALISTIC STUDIES: THE USE OF GIS AND GEODATASETS

As we have shown in the previous section, channel modeling accuracy is highly influenced by the data we feed into the models. In this chapter, we explore the use of Geographic Information Systems (GIS) techniques and open Geodatasets to achieve accurate and realistic estimations. First, we will introduce the concept of GIS, then the different datasets and their sources used throughout the thesis, and finally the visibility analysis techniques we have employed in our studies.

### 2.3.1 Introduction to GIS

GIS constitute a computational framework for the acquisition, storage, manipulation, and presentation of spatially referenced data. Rooted in cartographic science and geography, GIS has expanded its applicability across diverse academic and industrial fields. It provides a structured environment for complex spatial analyses, thereby influencing decision-making processes in areas such as urban development, environmental management, and geopolitical studies. As such, GIS stands as a fundamental mechanism in the contemporary methodologies for spatial inquiry.

#### *2.3.1.1 Projection systems*

A geographic coordinate system (GCS) provides means to represent locations on the Earth's surface using a set of numerical values. At its essence, a GCS defines positions using latitude, longitude, and often altitude. Latitude measures the angular distance north or south from the equator, with values ranging from -90° to 90°. Longitude, on the other hand, measures the angular distance east or west of the prime meridian, with values spanning from -180° to 180°. Altitude or elevation represents the vertical distance above or below a reference surface, typically mean sea level. To ensure accurate representations, a GCS relies on a datum, which is a model of the Earth's shape, and a reference ellipsoid, a mathematical figure that approximates the Earth's form. Different datums and ellipsoids are employed depending on the region and purpose of the study, and understanding the nuances of these systems is crucial for accurate spatial analysis and data integration in GIS applications.

In this research, different planar datums will be used depending on the region considered, so that the distortion is always minimized.

#### *2.3.1.2 PostGIS and QGIS*

Within GIS, PostGIS[7] and QGIS[8] have established themselves as fundamental tools for the management and visualization of spatial data.

[7] https://postgis.net/
[8] https://qgis.org

PostGIS is an extension of the PostgreSQL relational database that allows for the storage and management of spatial data. Its strength lies in its ability to handle and query large datasets efficiently. PostGIS supports a vast array of spatial data types, including points, lines, and polygons, and offers a rich set of functions to perform complex spatial queries, making it a preferred choice for professionals requiring robust spatial database capabilities.

QGIS, on the other hand, is an open-source GIS desktop application that provides an intuitive interface for visualizing, editing, and analyzing geospatial data. Its versatility is evidenced by its wide range of plugins and extensions, enabling users to customize the platform to their specific needs. QGIS can connect directly to PostGIS databases, allowing for seamless integration between spatial data storage and visual analysis.

Throughout this thesis, both software have been extensively used to store, interrogate, display, and analyze geographic datasets.

#### *2.3.1.3 Raster and Vectorial Data*

In GIS, data representation primarily takes on two distinct formats: raster and vector. Raster data comprises grid structures, where each cell or pixel contains a specific value representing information such as elevation, temperature, or land cover. This format is particularly suited for continuous spatial phenomena and is commonly used in remote sensing and DEMs. In contrast, vector data represents geographical features using points, lines, and polygons. Points might denote features like wells or cities; lines can represent roads or rivers, and polygons can depict areas like buildings, lakes, or administrative boundaries.

### 2.3.2 Vectorial Datasets - Roads, Buildings, and Landuse

#### *2.3.2.1 OpenStreetMap*

OSM is a collaborative project that provides freely accessible geographic data and mapping to users around the world. Unlike traditional mapping services, OSM harnesses the collective input of a vast number of users, from professional cartographers to general public enthusiasts, all contributing, and refining the map's data. At its core, OpenStreetMap operates predominantly as a vector dataset. It uses points to represent specific locations like lamp posts or mailboxes, lines for linear features such as roads, footpaths, and rivers, and polygons to depict areas like forests, buildings, or administrative boundaries. This vectorial representation allows for detailed, scalable, and dynamic visualizations, making OSM a versatile tool suitable for various applications, from route planning to spatial analysis in professional GIS platforms.

The complete raw OSM dataset is freely available for download from multiple sources. Yet, its substantial volume (exceeding 100GB even when compressed) combined with its format (a markup language crafted for generating OSM tiles) often tilts the preference towards acquiring pre-

processed datasets from intermediaries. Throughout this research, we have relied on the shapefiles processed by Geofabrik GmbH, which are available on their website [9]. Fig. 2.9 shows a portion of one of the vectorial datasets used throughout this thesis.

#### 2.3.2.2 *Technical Regional Charts*

In certain areas, usually in rural environments, OSM may have insufficient data to effectively perform our simulations. In such cases, we have resorted to different datasets, made available by public administrations under different types of licenses (both open and closed). Specifically, we have used the so-called Technical Regional Charts, which represent large areas in vectorial datasets containing different kinds of information such as the layout of roads, buildings, or land use. Tab. 2.1 details all the different datasets used for Technical Regional Charts.

### 2.3.3 Raster Datasets - Digital Elevation Models

DEMs are specialized datasets that capture the terrain's altitude information over a geographic area. Represented typically as a raster grid, each cell in a DEM contains a single elevation value, representing the height above a specific reference point, often sea level. DEMs serve as a fundamental tool in a wide array of applications, from hydrological modeling, where they help in predicting water flow patterns, to urban planning, where they assist in determining suitable construction sites. Additionally, they play a crucial role in environmental research, facilitating studies on soil erosion, habitat distribution, or landform processes. The accuracy and resolution of DEMs can vary, with some sourced from satellite imagery while others derive from airborne LiDAR data, which can offer finer detail.

**Figure 2.9:** The buildings of portion of the city of Florence IT, as represented by OSM.

[9] https://download.geofabrik.de/

#### 2.3.3.1 Shuttle Radar Topography Mission

The Shuttle Radar Topography Mission (SRTM) is a collaborative initiative spearheaded by NASA and the National Geospatial-Intelligence Agency (NGA) to generate high-resolution topographical data of the Earth's land surface. Conducted in February 2000 using the Space Shuttle Endeavour, SRTM utilized a specially modified radar system to obtain digital elevation data for over 80% of the globe, covering latitudes from 56°S to 60°N. The primary output of this mission is a world-wide DEMs with a spatial resolution of 1 arc-second (30m) meters. I.e. a raster representation of the morphology of the earth with cells of size 30x30m. These DEMs have been extensively used in large-scale RF propagation analyses, where natural obstacles, such as hills and mountains affect the propagation. However, their use is limited in urban areas where, due to their resolution, the morphology of individual buildings can not be represented.

#### 2.3.3.2 Light Detection and Ranging

Light Detection And Ranging (LiDAR) is a remote sensing method that employs laser pulses to measure distances to the Earth's surface. A LiDAR system typically consists of a laser, a scanner, and a specialized GPS receiver, often mounted on an aircraft or drone. As the system flies over an area, it emits rapid laser pulses towards the ground. By measuring the time taken for each pulse to bounce back after reflecting off the surface, LiDAR calculates precise distance measurements.

These accumulated measurements result in a dense set of elevation data points known as a *point cloud*. After post-processing, this point cloud can be converted into highly accurate DEMs. More specifically, DEM can either be a Digital Surface Models (DSM) or a Digital Terrain Model (DTM). The main difference between the two is that the former includes any object above the ground, such as trees, buildings, cars, etc.; while the latter is obtained by filtering only the ground points, which are then interpolated into a surface that represents the terrain, without any artificial obstruction on top. Fig. 2.10 shows a rendering of the DSM for a portion of the city of Florence, with a precision of 1 point per meter. As expected both individual trees and buildings are clearly identifiable.

The accuracy of LiDAR-derived DEMs is unique, often achieving vertical accuracy within a few centimeters. This level of detail allows for

| Area | Country | License | Source |
|---|---|---|---|
| Tuscany | IT | CC-BY-SA | Tuscany Region [10] |
| Campania | IT | Closed | Campania Region |
| Trentino | IT | CC-BY 2.5 | Province of Trento [11] |

**Table 2.1:** Table detailing the different Technical Regional Charts.

[10] http://www502.regione.toscana.it/geoscopio/cartoteca.html

[11] https://patn.maps.arcgis.com/apps/webappviewer/index.html?id=75c536fa3dd0463599818475ca06be6e

clear topographical analyses, making LiDAR an invaluable tool for applications such as flood modeling, forest canopy analysis, and urban planning. LiDAR measurement campaigns are often performed by public administrations, which then release the dataset to the public with open data licenses. Throughout this thesis, different DEM have been used, as detailed in Tab. 2.2.

More formally, let the DEM $\mathcal{D}$ be a set of triplets $(x, y, z)$, where $(x, y)$ represents the coordinate of a point on the surface and $z$ its height above sea level. We then model the DEM as a bi-dimensional matrix of real numbers $\mathbf{E} \in \mathbb{R}^{m_x \times m_y}$, $m_x$ and $m_y$ are the number of samples in the $x$ and $y$ dimensions respectively[12].

$$\mathbf{E}_{x,y} = z \quad \forall (x, y, z) \in \mathcal{D} \tag{2.23}$$

### 2.3.4 Other Datasets

#### 2.3.4.1 *Censuary Data*

Every ten years, the Italian National Institute of Statistics (ISTAT) makes a new population census, which is then partially released in the form of open data. Different granularity are available, with many reports released for large areas such as the whole country, individual regions, or cities. However, a more fine-grained open dataset called the 'censuary variables' dataset is also released which contains a portion of these variables divided into extremely small areas called *censuary sections*. These areas are the smaller units of measurement in the censuary process and can be as small as a single block in urban areas. These variables include the number of residents, divided by sex and age, educational level, etc. In this thesis, we have used the number of households for each censuary section to devise a demand model for broadband connectivity.

The dataset is available on the ISTAT website [13]

#### 2.3.4.2 *Broadband Coverage*

In Italy, the Regulator for Communications Guarantees (AGCOM) releases different datasets on broadband coverage in Italy, both for fixed and mobile access.

The dataset is divided into two categories: metrics associated with each censuary section and gridded maps. In the first case, variables such as the number of households connected with each technology (DSL, VDSL, FTTH), the average download speed for each technology, and the number of households with download speed into different ranges (0-2, 2-30, 30-100, 100-500, 500-1000). Regarding mobile networks, the dataset provides the

[12] In the whole thesis we always use a precision of 1 point per meter so that the number of sample on a given axis equals its size in meters. For DEMs with a precision higher than 1 point per meter downsampling has been performed

[11] https://www.icgc.cat/en/Downloads/Elevations/Lidar-data

[13] https://www.istat.it/it/archivio/104317

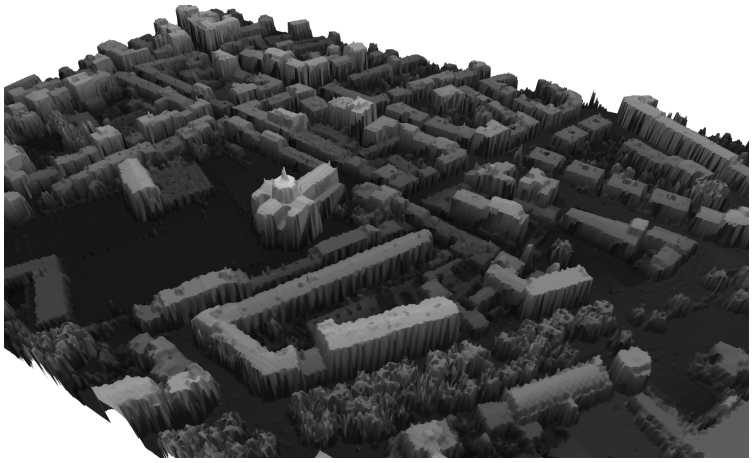


**Figure 2.10:** 3D rendering of the DSM of a portion of the city of Florence (IT). The DSM has been obtained from a LiDAR measurement campaign and has a precision of 1 point per meter.

number of different MNO available for each technology for each censuary section. In the second case, similar data are mapped over a gridded square with a cell size of 1x1 km.

Fig. 2.11 shows a coverage map from the AGCOM dataset. All datasets are available for download on the AGCOM website[20]

### 2.3.5 Visibility Analysis

Visibility analysis represents a pivotal GIS technique, finding applications across a myriad of disciplines such as telecommunications [40], [41], urban planning [42], and archaeology [43].

It is of particular interest in the context of wireless networks where, as we have seen in Sect. 2.2.1, LoS and NLoS conditions significantly impact performance. This technique allows network designers to assess

| Area | Country | Density | Source |
|---|---|---|---|
| Milan | IT | $1p/m^2$ | Italian Ministry of Environment [14] |
| Tuscany | IT | $1p/m^2$ | Tuscany Region [15] |
| Naples | IT | $1p/m^2$ | Province of Naples[16] |
| Trentino | IT | $1p/m^2$ | Province of Trento [17] |
| Luxemburg | LU | $4p/m^2$ | Luxemburg National Geoportal [18] |
| Catalunya | ES | $9p/m^2$ | Catalunyan Cartographic Institute [19] |

**Table 2.2:** Table detailing the different LiDAR datasets.

[14] http://www.pcn.minambiente.it/mattm/

[15] http://www502.regione.toscana.it/geoscopio/cartoteca.html

[16] https://sit.cittametropolitana.na.it/layer.php?id=sit:quadro_unione_lidar_dsm

[17] https://siat.provincia.tn.it/stem/

[19] https://geoportail.lu/en/

[20] https://maps.agcom.it/

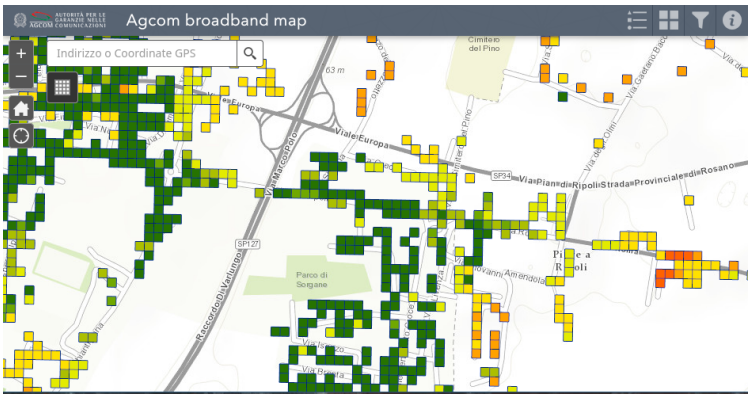



**Figure 2.11:** Coverage map showing the download speed in the suburbs of Florence (IT).

how terrain, buildings, and other physical obstructions affect the propagation of RF signals between the transceivers. By applying GIS tools to high-resolution topographical and architectural data, visibility analysis provides a comprehensive assessment of where the signal will be strong enough to meet performance criteria and where it may be degraded or lost entirely.

These obstructions can cause signal reflection, diffraction, and absorption, affecting the network's reliability and performance.

Several computational methods and algorithms can aid in conducting visibility analysis, ranging from simple LoS calculations to more complex models that consider various types of signal interference and attenuation. The integration of these tools with GIS technology allows for a more accurate and efficient design process.

#### 2.3.5.1 LoS determination

The basic block in visibility analysis is the determination of the LoS between two points. Given two points in the 3D space $p_i = (x_i, y_i, z_i)$ and $p_j = (x_j, y_j, z_j)$, and a DEM matrix $\mathbf{E}$, the objective is to establish whether the direct line joining these points intersects with the DEM. Specifically, this means determining if the elevation of the DEM at any point lies below the height of the line. Formally:

$$\mathrm{Y}(\mathbf{E}, p_i, p_j) = \mathbf{E}[x_k, y_k] < p_k \quad \forall p_k \in f(p_i, p_j) \tag{2.24}$$

Where $f(p_i, p_j)$ is a function that computes the points of a raster (in our case $\mathbf{E}$) that approximate the line between the two points $p_i$ and $p_j$. Such a function can be implemented using the well-known Bresenham's line algorithm [44], which chooses the raster's cell whose center is the closest to the line. Fig. 2.12 depict the result of the Bresenham algorithm.

In practice, implementing the LoS determination can be more difficult, as we need to take into account aspects such as the earth's curvature, and it has to be parallelizable on a large number of points. During this thesis, the work by Osterman [45] has been adapted and implemented

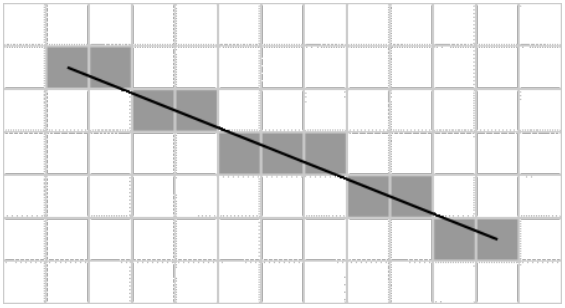


**Figure 2.12:** Illustration of the result of the Breshenam's line algorithm for $p_i = (1,1)$ and $p_j = (11,5)$.

.

in Numba [46], a Python framework for parallel and GPU computing and run on NVIDIA GPUs. Algorithm 2.1 shows the pseudocode of the implementation. We can see that its complexity is linear in the distance between the two points.

#### 2.3.5.2 *Viewsheds*

We can now move to the 2D extension of the LoS algorithm, which is the viewshed algorithm. We are interested in determining the LoS between a single point, usually called observer, and a large set of other points on the DEM ($\Lambda$). We generally refer to the result of this operation as the *viewshed* $\sigma^i$, binary matrix, representing the visibility from the observer point $p_i$ to all the other points $p_j = (x, y)$ of the DEM $\mathbf{E}$, defined as:

$$\sigma^i_{x,y} = \begin{cases} 0 & (x,y) \notin \Lambda \\ 0 & d(p_i, p_j) \geq d_{\max} \\ \Upsilon(\mathbf{E}, p_i, p_j) & \text{otherwise} \end{cases} \tag{2.25}$$

where $\Upsilon$ is the function defined in Eq. (2.24), $d(p_i, p_j)$ is the euclidean distance between two point, and $d_{max}$ is a technology-dependent arbitrary maximum communication distance.

Fig. 2.13 shows the viewshed computed over set $\Lambda$ representing the roads, from an observer point on the corner of a building.

A naive implementation could take advantage of the algorithm shown in the previous section and call it iteratively for any point on $\mathbf{E}$. However, this would be highly inefficient, with a complexity of $O(n^3)$ with respect to the width (or height) of the area.

More efficient approaches with lower complexity have been studied [47], with the most notable one being the sweep algorithm from van Kreveld [48] with a complexity of $O(nlog(n))$. In this thesis however, we focused on the advantages brought by GPU computing, which can dramatically accelerate such algorithms and we have adapted an algorithm from Osterman et Al. [45] that runs with $O(n^2)$ time complexity. We don't report the algorithm as it is available on the original manuscript.

**Algorithm 2.1** Algorithm for the computation of a single Line-of-Sight on a CUDA kernel.

**Require:** **E** (DEM of the area), $p_i$ (source node), $p_j$ (target node)

**procedure** Y(**E**, $p_i$, $p_j$)
  $(x_i, y_i, z_i) = p_i$ ▷ Read coordinates of source node
  $(x_j, y_j, z_j) = p_j$
  $x^s = x_j - x_i$
  $y^s = y_j - y_i$
  $d = ||p_j - p_i||_2$ ▷ Calculate the euclidean distance
  $h_c = \sqrt{d^2 + Re^2} - Re$ ▷ Calculate earth curvature height correction
  $z_j = z_j - h_c$ ▷ Correct the height w.r.t. the earth's curvature
  **if** $|x^s| > |y^s|$ **then** ▷ Find dominating axis and compute the step values
    $\text{step}_x = 1 * sgn(x^s)$
    $\text{step}_y = y^s / |x^s|$
    $\text{steps} = |x^s|$
  **else**
    $\text{step}_y = 1 * sgn(y^s)$
    $\text{step}_x = x^s / |y^s|)$
    $\text{steps} = |y^s|$
  **end if**
  $t = (-z_i - 2 + z_j + 2)/d$ ▷ Calculate the slope of the Line of Sight
  **for** $j \leftarrow 0$ to steps **do** ▷ Iterate over the dominating axis
    $x' = \text{step}_x * j + x_i$
    $y' = \text{step}_y * j + x_j$
    $d^{x'} = x' - x_i$
    $d^{y'} = y' - y_i$
    $d' = \sqrt{d^{x'} * d^{x'} + d^{y'} * d^{y'}}$
    $h'_c = \sqrt{d'^2 + Re^2} - Re$
    $z' = E[int(x'), int(y')] - h'_c$ ▷ Read the height of the current point and correct
    $l = t * d' + z' + 2$ ▷ Calculate the height of the Line of Sight
    **if** $z' > l$ **then**
      **return** 0
    **end if**
  **end for**
  **return** 1
**end procedure**

Figure 2.13: Representation of a viewshed computed in an urban area from the corner of a building.

#### 2.3.5.3 Set of viewsheds

A further extension of the viewshed analysis is the analysis from multiple points (a set $\mathcal{P}$). By doing so, we obtain a set of viewsheds ($\Omega$) that can be used in several ways. Formally, let define $\Omega$ as:

$$\Omega = \{\sigma^i \; \forall \; p_i \in \mathcal{P}\} \tag{2.26}$$

**CUMULATIVE VIEWSHEDS** One first application of this set is the computation of the cumulative viewshed ($\sigma^\star$). In fact, by summing together the cells of each viewshed, for each viewshed of the set, we obtain a single viewshed, with values ranging from 0 to $\mathcal{P}$ that represents the cumulative visibility from the set of points $\mathcal{P}$. In other words, if an element $\sigma^\star_{x,y}$ is equal to 0 it means that it's not in visibility with any of the points in $\mathcal{P}$, and if it is equal to $n$ it means that $n$ points in $\mathcal{P}$ are in LoS with $(x, y)$.

This type of analysis is often used in archeological studies [43] and in landscape planning [42].

Formally each element $(x, y)$ of the cumulative viewshed $\sigma^\star$ can be defined as:

$$\sigma^\star_{x,y} = \sum_{\sigma^i \in \Omega} \sigma^i_{x,y} \tag{2.27}$$

By changing the operation performed on the matrixes, such as by using the bitwise and ($\bigwedge$) or the bitwise or ($\bigvee$) instead of the sum ($\Sigma$), we can obtain different results, as done in the following researches, specifically in Chapter 3.

#### 2.3.5.4 Visibility Graphs

A visibility graph is a graph of locations that are mutually visible to each other. Their use ranges from robotics, where they are used to navigate a space filled with obstacles [49], to physics where they are used as a tool to analyze time-series [50]. In the context of telecommunication they have

been widely used as a tool to optimize the placement of radio antennas [41] and more broadly for telecommunication network planning [40].

More formally, given a set of points $\mathcal{N}_v$, and a DEM matrix $\mathbf{E}$. The visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ can be defined as follows:

$$\mathcal{E}_v = \{(v_i, v_j) \quad s.t.\ \Upsilon(\mathbf{E}, v_i, v_j) = 1 \quad \forall (v_i, v_j) \in \mathcal{N}_v \times \mathcal{N}_v\} \tag{2.28}$$

In other words, it is a graph whose vertices are points on the space and whose edges represent mutual LoS between those points. Algorithm 2.2 shows the algorithm used for the computation of visibility graphs, taking advantage of the function Υ introduced in Eq. (2.24) and implemented in Algorithm 2.1. Its complexity is cubic, or more precisely it is equal to $O(|\mathcal{N}_v|^2 \cdot d_{max})$, where $d_{max}$ represents the maximum distance between two points in $\mathcal{N}_v$.

**Algorithm 2.2** Algorithm for the computation of $G(\mathcal{N}_v, \mathcal{E}_v)$.

**Require:** $\mathbf{E}$ (DEM of the area), $\mathcal{N}_v$ (Set of points),
1: **procedure** GENERATE $G(\mathcal{N}_v, \mathcal{E}_v)(\mathbf{E}, \mathcal{N}_v)$
2: $G_v = \mathbf{0}$
3: **for** $p_i \in \mathcal{N}_v$ **do**
4: **for** $p_j \in \mathcal{N}_v \setminus \{p_i\}$ **do**
5: $G(\mathcal{N}_v, \mathcal{E}_v)[i, j] = \Upsilon(\mathbf{E}, p_i, p_j)$
6: **end for**
7: **end for**
8: **return** $G(\mathcal{N}_v, \mathcal{E}_v)$
9: **end procedure**

### 2.3.6 Set covering algorithms

An additional application of viewshed analysis, specifically using the set of viewshed $\Omega$, that we explore thoroughly in Sect. 3.2 and Sect. 3.3, is the application of set covering algorithm to the set of viewsheds $\Omega$. By employing these techniques is possible to develop heuristics to optimize the placement of gNBs for LoS communications.

Let briefly recall the general set cover problem: Given a set of elements $\mathcal{U}$ and a collection $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$, such that $s_i \subset U \quad \forall \quad s_i \in U$ and $\bigcup \mathcal{S} = U$ the objective is to find a sub-collection $\mathcal{S}' \subset \mathcal{S}$ with minimal size $(min \quad |\mathcal{S}'|)$, that covers all the elements of $U$ $(\bigcup \mathcal{S}' = U)$. The problem is known to be NP-complete, but polynomial-time heuristic exists with bounded error [51], [52].

We can then adapt its definition to our problem: Given a set $\Lambda$ and a collection of viewsheds $\Omega$, where each $\sigma^i$ is a subset of elements of $\Lambda$, we need to find a minimal subset $\Omega' \subset \Omega$ that covers the entire $\Lambda$.

In our placement problems, $\Lambda$ is the set of points on a public street, and $\sigma^i$ is the set of points covered by a specific gNB. However, we can introduce some differences compared to the classical problem obtaining different problems that we will explore in the rest of this thesis. First,

we want to achieve reliability by covering each element multiple times; second, we do not want to fully cover each set, but we tolerate a small percentage of uncovered points; and third, we want to weight each element of the set by following a vehicular traffic model. The first problem is generally referred to as a multi-cover problem and the second one is a partial-cover problem. Their intersection yields a less studied, yet more difficult problem [53], which in recent works has been solved through the use of heuristics and integer linear problems [54], [55]. These solutions are unfortunately computationally impossible to solve in our context of large-scale optimization.

Additionally, we might want to associate a demand to each element of $\Lambda$, so that we can skew the placement algorithm to cover some areas that we know have an higher density of devices. This problem is generally referred to as the weighted maximum coverage problem and is a generalization of the set-cover problem [21]. It retains its NP-completeness and is approximable by a greedy heuristic with bounded error [56].

The intersection of the multi-coverage with the weighted variant, i.e. providing reliable coverage by following a vehicular demand model, has not been studied in this thesis and is left for future analyses.

[21] By equally weighting all the elements of the set, the problem converges to the classic set-cover problem.

# 3 NETWORK DENSIFICATION AND WIRELESS BACKHAULING

The advancement of 5G networks has highlighted numerous challenges and factors to consider, with network densification standing out as a vital aspect. This chapter specifically addresses network densification in the context of 5G and Wireless Backhaul Network (WBN), focusing on the role of Integrated Access and Backhaul (IAB).

In telecommunications, densification typically involves increasing the number of Next Generation Node Bases (gNBs) in a specified area to enhance both capacity and coverage. The promise of 5G, with its high data rates, reliability, and low latency, requires an effectively dense network configuration. However, simply adding more gNBs without methodical planning can lead to suboptimal performance. It's therefore essential to have methodologies that guide both the strategic placement of gNBs and their efficient interconnection.

One of the fundamental challenges in wireless transmission is the accurate assessment of Line of Sight (LoS) and Non-LoS (NLoS) conditions. The quality of the wireless channel is significantly influenced by these conditions and traditional methods of estimating it often fall short of providing a realistic representation, especially in dense urban settings where obstructions and interferences are common. Thus, there is a need for innovative techniques that can provide more accurate estimations.

In this context, this chapter introduces the application of visibility analysis techniques, drawing from the field of Geographic Information Systems (GIS), in Radio Frequency (RF) planning. These techniques allow for the creation of models and datasets that offer realistic estimations of LoS and NLoS conditions, which are crucial for optimizing the performance of 5G networks in dense environments.

This chapter comprises four key studies that delve into this subject. These works, published on journals or in the progress of being published, provide a comprehensive understanding and solution framework for the challenges posed by network densification.

The first study, published in 2022 by IEEE Transaction of Network and Service Management [10], explores the optimal placement of gNBs in ultra-dense access networks. In fact, the ongoing transition towards high-frequency communication is challenged by both fixed obstructions, such as buildings, and mobile obstructions, such as vehicles, which can significantly impact signal propagation. By leveraging visibility analysis, this study investigates their placement based on LoS conditions across 3 different urban areas in Italy by leveraging different optimization targets, so that each point on the ground is covered by multiple gNBs. Through extensive simulations, the study also identifies lower bounds on the required density of gNBs to achieve such coverage.

The second study presented in this chapter, published initially at The 18th Annual Conference on Wireless On-demand Network Systems and Services [7] and then invited for an extension on the Computer Communication Journal [2] where it's undergoing the review process, extends the aforementioned placement optimization algorithms in the context of vehicular networks. Here we analyze how the need for densification differs between a vehicular-centric and a pedestrian-centric network. Moreover, by applying channel models for millimeter wave (mmWave) communication, we evaluate the capacity of such networks at different gNB densities.

The subsequent two studies shift focus to the interconnection of these nodes, a process called backhauling. These studies introduce the concept of visibility graphs in the research area of telecommunications, where performance evaluations are often carried out with simplistic assumptions that do not resemble the real world or by leveraging on computational-intense methods, such as ray-tracing, which however constrain the evaluation on small areas.

The third, published as a Work In Progress paper at the 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks [11], introduces the idea of exploiting visibility analysis to backhaul design and compares Its applications to more traditional approaches. It also implements a state of the art algorithm for reliable topology design and tests its application on the visibility graph.

The fourth, and last, research, published in 2023 by IEEE Transaction of Network Science and Engineering [8], further extends this, by analyzing 9 different Italian municipalities for which the visibility graph is built between all the buildings in order to characterize topologically the connectivity of those areas. Visibility graphs are then compared to less realistic – yet widely-used – topological models that leverage the simplistic assumption, such as LoS probability models, in terms of their topological properties. Moreover, by applying state-of-the-art backhaul topology design algorithms on both the visibility graph and the less-realistic graphs, further evaluation are carried out to demonstrate the need for more realistic topological models.

## 3.1 RELATED WORKS

The background of these studies is rooted in different, but related research areas: cellular networks, visibility analysis, set covering algorithms, wireless backhaul topology design, and channel models. Some of these topics have already been covered previously in Sect. 2.2.1 and Sect. 2.3.5. Here below we detail the state of the art related to the one specific to this chapter.

### 3.1.1 Cellular Networks

The placement of gNBs has been widely studied since the deployment of the first mobile networks [57]. The topic has gotten more focus lately, due to the densification of gNBs, but the specific topic of LoS coverage has not received much attention. Even very recent works as [58] base their analysis on antenna directionality, but disregard the 3D geometry of the problem. Anjinappa et al. [59] investigate the optimal placement for gNBs and passive reflectors in two urban areas; however, they do not consider any cost model or try to account for the Capital Expenditure (CapEx) to deploy such a network. Zhang et al. [60] tackle the problem of gNBs placement by trying to minimize the outage probability by studying a regular, Manhattan-style urban topology, while Haile et al. [61] use an approach similar to ours, for NLoS communications. Another branch of research focuses on the placement of Unmanned Aerial Vehicles (UAVs) base stations [62] but has different requirements compared to our problem.

Research investigating placement strategies to optimize mobile coverage by taking advantage of realistic traffic data seems, to the best of our knowledge, to be missing. The most similar research, from Jaquet et Al. [63] is focused on enhancing vehicular networks by taking advantage of unmanned aerial vehicles. Another research, from Anjinappa et Al. [64], optimizes the placement of mmWave gNBs and Passive Metallic Reflectors to provide LoS coverage, lacking however a validation on realistic traffic models. Devoti et Al. [65], proposes a network planning framework for mmWave networks, which takes into account angular separation and links length to optimize the placement of gNBs in an urban scenario. Fiore et Al. [66], introduces Reconfigurable Intelligent Surfaces (RISs) and optimize the placement of both RISs and gNBs to increase the reliability of IAB networks.

Similar problems have also been addressed for the placement of sensors and cameras. Sensors are, in fact, assumed to be able to perceive events in a certain area around them, and one problem is finding the minimal deployment of sensors that cover each point multiple times. This problem is not new [67] and received a lot of attention [68], but the constraints and the solutions are very different from network coverage: The visibility regions of urban public spaces are not unit circles nor stochastic regions; furthermore, there is no reason to keep points connected because they are not communicating with each other, and our points are contiguous and not randomly distributed on a 2D space. Regarding camera placement, a common problem is placing a minimal set of cameras to monitor a certain area, or a certain number of objects in space [69]. The problem is only partly similar because cameras can be oriented, they can pan and zoom, and deployments are in the order of tens of cameras, while we have tens of thousands of potential locations. However, given the similarity to network coverage, we adapt the heuristic proposed in [70] that tries to achieve a fair coverage of selected targets from multiple cameras.

### 3.1.2 Wireless backhaul topology design

Several works have been proposed in the past to design a wireless backhaul for 5G networks with different optimization targets: reliability [71], energy efficiency [72], and cost [73]. These studies rely on mmWave links or free-space-optical links [74], both requiring LoS between the endpoints. The introduction of IAB in the latest revision of 5G provided a concrete application and reinforced the interest in the creation of efficient backhaul topologies [75]–[78].

To the best of our knowledge, works that focus on the design of wireless backhaul topologies take a very simplistic approach in modeling the characteristics of the visibility graph, due to the total absence of literature that characterizes the realistic properties of visibility graphs. In the same way, works on mobile networks often leverage unrealistic gNBs placement strategies to evaluate their contributions. For these reasons, we stress the need for realistic topological models, as well as gNB placement algorithms to be published. So that future researchers can use them to evaluate their contributions more effectively.

## 3.2 RELIABLE LoS COVERAGE FOR URBAN AREAS

In this research, we used a novel 3D approach to estimate the $w$-coverage, i.e., the coverage with multiplicity $w$, and the cost of its deployment, by extending existing heuristics developed for two NP-complete problems: set coverage and minimum partial set multi-cover problems. The solution to these problems returns the optimal placement of gNB for LoS communication in urban areas, with different objectives depending on the different metrics we proposed and evaluated.

The analysis was carried out on 3 Italian Cities (Trento, Firenze, and Napoli) where we selected 5 central areas with an average size of 0.7 $km^2$ for a total of 15 different locations. As both the code and datasets had been released with open-source licenses, the extension to other areas is straightforward.

We considered placing antennas on buildings' facades, and we used state-of-the-art GIS techniques implemented on Graphics Processing Unit (GPU), as described in Sect. 2.3.5.2, to estimate the LoS coverage with its multiplicity on the ground area identified by public streets.

We then evaluated the coverage, and the cost necessary to achieve it, offering significant insight on the problem of robust coverage for LoS communication, ranging from the effective number of gNBs required to achieve a given coverage (e.g. 90 or 95%) to building selection strategies to reduce the coverage cost.

The contribution of this research addressed two different aspects of coverage planning related to mmWave and in general to ultra-high frequency technologies that require LoS for good performance. The first one relates to the algorithms available to solve the multi-coverage problem to increase the robustness and resilience of the network. We considered state-of-the-art algorithms derived from visibility analysis problems and adapted them to tackle communication networks' planning, showing how to modify them and how they affect the network design.

The second contribution is the first quantitative analysis of real urban data on the required density of gNBs to achieve a robust and reliable coverage with mmWave and THz communication. The results we present show that in general between 60 and 120 gNBs per $km^2$ are sufficient for LoS coverage in public streets when the maximum distance between the User Equipment (UE) and the gNB is set to 300m. This is consistent with the estimate of 100 gNBs per $km^2$ often mentioned by operators [16] and academics [58].

### 3.2.1 Problem constraints and Optimization Goals

As LoS communications prevent placing gNBs on roofs, we assume their deployment on the facades of buildings, a location that is already used by telecommunication companies to install fiber-to-the-home splitters and 4G small cells.

| Symbol | Type | Definition |
|---|---|---|
| $\mathcal{D}$ | Set | Digital Elevation Model (DEM) |
| $\mathbf{E}$ | Matrix | Matrix Representation of DEM |
| $p_i$, $p_j$ | Vector | Generic points in a 3D space |
| $s^r$ | Set | The points of the perimeter of building $r$ |
| $A(s^r)$ | Function | Function that returns the points inside $s^r$ |
| $R$ | Scalar | Number of buildings |
| $\mathbf{S}^r$ | Matrix | Rasterized building shape |
| $\delta()$ | Function | Dilation |
| $\mathbf{P}^r$ | Matrix | Rasterized perimeter of the building $r$ |
| $\mathcal{C}^r$ | Set | Set of potential coordinates on the border of building $r$ |
| $\Lambda$ | Set | $(x, y)$ coordinates of street points to cover |
| $\sigma^i$ | Matrix | Ground visibility from the point $p_i$ |
| $\Omega$ | Set | Ground visibility matrices |
| $\Gamma()$ | Function | Maximum coverage heuristic |

**Table 3.1:** Symbols and notation used. In general scalars and vectors are standard math characters and matrices' names are bold. Sets are calligraphic letters, elements of a matrix use the same name of the matrix and appropriate pedices, e.g., $\sigma^i_{x,y}$ is an element of matrix $\sigma^i$.

Let $A(s^r)$ be the area contained in a polygon $s^r$ that represents the 2D perimeter of a building, which we obtain from open datasets, and let $R$ be the total number of buildings in the area. Given a building $r$ whose 2D perimeter is represented by $s^r$, we implement a function that transforms the polygon into a binary matrix $\mathbf{S}^r$ of the same dimensions of $\mathbf{E}$, so that:

$$S^r_{x,y} = 1 \iff (x, y) \in A(s^r) \tag{3.1}$$

In order to isolate the border points of the building, we use a morphological operation $\delta()$ called dilation, which dilates the binary matrix of the polygon by one unit in every direction [79] and sets to 1 the corresponding coordinates. Then given the matrix $\mathbf{S}^r$ we define its perimeter using a matrix $\mathbf{P}^r$ as:

$$\mathbf{P}^r = \delta(\mathbf{S}^r) - \mathbf{S}^r \tag{3.2}$$

Figure 3.1 shows the graphical representation of the matrices $\mathbf{S}^r$ and $\mathbf{P}^r$.

We assign to every point on the perimeter a height from the ground $z_r$ that is the minimum between the height of the building minus one meter and 10 m from the street level, and finally we define the set of coordinates $\mathcal{C}^r$ as:

$$\mathcal{C}^r = \{(x, y, z_r) \mid P^r_{x,y} == 1\} \tag{3.3}$$

Eq. (3.3) identifies a set of potential positions for gNBs on the facade of building $r$. Each point lies on the border of the buildings and it is placed at the height of the building, or at 10 m from the ground if the building is taller than 10 m (as the ITU recommendations for micro cells in urban areas suggest [27]). In every area we have roughly 500 to 700 buildings, the set of points $p_i \in \{\cup_r \mathcal{C}^r\}$ is the set of potential locations where to

place our gNBs. As the average number of points per building perimeter is roughly 100 we have a total of 50,000 to 70,000 potential gNB locations.

#### 3.2.1.1 Defining the Ground Visibility Matrix

We use street maps provided by Openstreetmap, where each road is identified by a mono-dimensional line that represents the center of the street. We expand the line to make it a 2D surface and we call $\Lambda$ the set of all $(x, y)$ coordinates of the points on the street, obviously $(x, y) \in \Lambda \rightarrow (x, y) \notin A(s^r)\ \forall r < R$. Let us now consider all points $p_j$ of coordinates $(x, y, E_{x,y} + 1.5)$ where $(x, y) \in \Lambda$. The point $p_j$ is a point in the street elevated by 1.5 m from the ground, a common assumption of the position of a mobile terminal. Assume $p_i$ is the position of a gNB, we call $\sigma^i$ the viewshed matrix computed from such point, as defined in Eq. (2.25). Figure 3.2 shows the graphical rendering of a matrix $\sigma^i$.

Given all the potential locations of gNBs, i.e., the union of all the sets $\mathcal{C}^r$, let $\Omega$ be the collection of all potential visibility matrices:

$$\Omega = \{\sigma^i\ \forall\ p_i \in \{\cup_r \mathcal{C}^r\}\} \tag{3.4}$$

and let $k$ be the number of gNB we want to install in that specific area. The problem can be expressed as the search of a subset $\Omega' \subseteq \Omega$ that maximizes $w$-coverage, with $|\Omega'| \leq k$. We start by defining 1-coverage, which is intuitive, and we then extend the approach to arbitrary values of $w$.

#### 3.2.1.2 Maximizing 1-coverage

Let $|\cdot|_0$ be the $L_0$ norm of a matrix (the number of non zero elements); the problem of maximizing 1-coverage can be formalized as the search of a subset $\Omega' \subseteq \Omega$, such that

$$|\Omega'| \leq k \quad \text{and} \quad \left|\sum \sigma^i \in \Omega'\right|_0 \quad \text{is maximised.} \tag{3.5}$$

This is a classical maximum coverage problem, in which we have a collection $\Omega$ of sets, each set has elements in $\Lambda$ and given $k$, we need to find the $k$ sets that cover the largest number of elements of $\Lambda$. Being a set covering problems, it is NP-complete and can not be solved exactly when $|\Omega|$ and $|\Lambda|$ are in the order of tens of thousands. We leverage a polynomial heuristic with bounded error [80] that allows finding a semi-optimal solution in a reasonable time, extending it to fit our problem. Since the maximum coverage problem can be seen as a special case of maximization of submodular functions with a cardinality constraint, we can state that the error of this polynomial heuristic is bounded and it matches the theoretical bound [51] [81].

#### 3.2.1.3 Maximizing $w$-coverage

The interpretation of Eq. (3.5) is straightforward: given a number $k$ of gNBs that can be deployed to cover a certain area, we want to have visibility

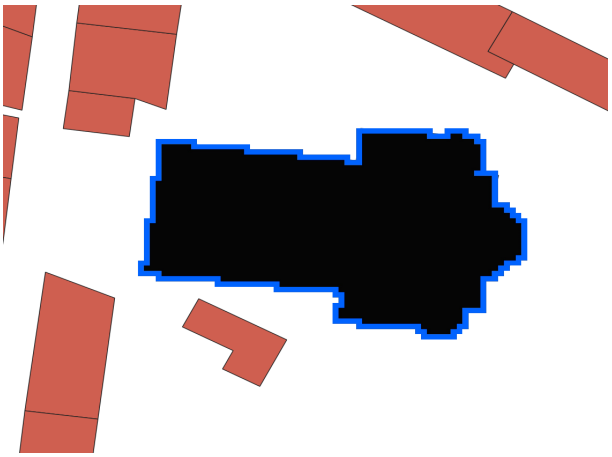


Figure 3.1: Dilation of a building $r$ with its perimeter points $\mathbf{P}^r$ in blue (each pixel represents a point) and the rasterized building shape $\mathbf{S}^r$ in black. © 2022 IEEE

Figure 3.2: Ground coverage $\sigma^0$ (light green) from a gNB in point $p_0$ (black) in a portion of the city of Florence. Figure shows also the buildings' shapes $s^r$ (red), the set of points on the street $\Lambda$ (grey) and other gNB locations (purple). © 2022 IEEE

from at least 1 gNB to the largest possible number of points in the public streets.

The 1-coverage is sufficient to claim to be covering an area, but does not provide reliability because there can be obstacles that obstruct the LoS with the gNB, the most obvious one being the person that holds the terminal, but also other people, cars, trucks, and so forth. The goal of a good coverage plan must be to provide 1-coverage to the largest area, but also a reliable service to mobile users, thus the problem is better formulated as a $w$-coverage, with a suitable $w$.

Maximizing $w$-coverage can be formulated as finding $\Omega'$ such that:

$$\begin{aligned} |\Omega'| \leq k \quad \text{and, given} \quad \mathbf{\Phi} = \sum \sigma^i \in \Omega' \\ \left|\{\Phi_{x,y} \geq w\}\right| \quad \text{is maximised,} \end{aligned} \tag{3.6}$$

where $\mathbf{\Phi}$ is a matrix populated with the multiplicity of the coverage for each point.

This is an extension of the set covering problem known as *minimum partial set multi-cover problem*, it is NP-complete [55], and no efficient heuristics are yet known. To solve this problem, we start from a heuristic proposed for multi-camera visibility [70] and, after analyzing the specific problem and goal, we propose a simpler one with comparable performance.

### 3.2.2 A generic heuristic for $w$-coverage

The goal of $w$-coverage is to guarantee a robust and efficient coverage, but in communications it has many subtleties that need to be considered, in particular, the problem that a complete $w$-coverage in real cities is probably impossible or simply too costly to pursue, so one has to take into account also the points where the coverage is smaller than $w$, solving dilemmas as, for instance, *is it better to add an antenna that improves the 1-coverage in n points or one that improves the w coverage in m other points?*

To achieve this goal we wish to have a single approximation algorithm where different metrics (or score functions) can be applied to explore different balances of the coverage. Algorithm 3.1 describes the heuristic we propose to tackle the $w$-coverage problem, extending the heuristic algorithm for set coverage proposed in [80].

Let us first introduce the notation used in the algorithm. Let $\mathbf{0}$ be the matrix of dimension $m_x \times m_y$ of all zeros. We also introduce the *sum-capped-to-w* operator with symbol $+_w$ so that given matrices $\mathbf{B}$, $\mathbf{D}$, then $\mathbf{A} = \mathbf{B} +_w \mathbf{D}$ is a matrix whose elements are: $\mathbf{A}_{i,j} = \min(w, \mathbf{B}_{i,j} + \mathbf{D}_{i,j})$.

Algorithm 3.1 works as follows. It takes in input a set of ground visibility matrices $\Omega$, each one corresponding to a point $p_j$ where a gNB can be placed, a number $k$ of gNBs to be placed, and the desired value for $w$. It returns a subset $\Omega^\star \subset \Omega$ corresponding to the set of the viewsheds (1-t-1 mapped to gNB positions) that provide a close-to-optimal $w$-coverage. Line 2 initializes the matrix $\mathbf{C}$ at $\mathbf{0}$. The points on public streets are considered by the algorithm, the others remain untouched at zero. Line 3 initializes the set $\Omega^\star$ to an empty set. Every iteration of the loop beginning at Line 4 chooses a new viewshed $\sigma^j$ (corresponding to a possible gNB location) to be added to $\Omega^\star$. To select the viewshed above, every iteration of the loop beginning at Line 6 evaluates the additional coverage offered by every possible viewshed, i.e., every possible candidate position for a gNB. $\mathbf{C}^\star$ is defined at Line 7 as the matrix of the points covered if the candidate viewshed $\sigma^j$ is added to the set. Every element of $\mathbf{C}^\star$ is thus the result of the element by element sum capped to $w$ of the already existing coverage plus the viewshed $\sigma^j$. The score function $\rho(\mathbf{C}^\star, k)$ at Line 8 is the core of the algorithm that actually defines the metric by which candidate gNB locations are ranked. Sect. 3.2.2.1 discusses three different score functions we propose and evaluate.

The score functions are defined so that a better coverage obtained by adding $\sigma^j$ corresponds to an algebraically higher value of $h_j$. Line 9 checks

**Algorithm 3.1** Computation of approximate solution to the maximum coverage and minimum partial set multi-cover problems

**Require:** $\Omega$ (Set of viewsheds), $k$ (number of gNB), $w$ ($w$-coverage)
**Ensure:** $\Omega^\star$ (Set of the viewsheds from optimal locations)

1: **procedure** $\Gamma(\Omega, k, w)$
2: $\quad \mathbf{C} = \mathbf{0}$
3: $\quad \Omega^\star = \{\}$
4: $\quad$ **for** $i \leftarrow 0$ to $k$ **do**
5: $\quad\quad h^\star = -\infty$
6: $\quad\quad$ **for** $\sigma^j \in \Omega$ **do**
7: $\quad\quad\quad \mathbf{C}^\star = \mathbf{C} +_w \sigma^j$
8: $\quad\quad\quad h_j = \rho(\mathbf{C}^\star, w)$ ▷ Calculate score
9: $\quad\quad\quad$ **if** $h_j > h^\star$ and $\sigma^j \notin \Omega^\star$ **then**
10: $\quad\quad\quad\quad \sigma^\star = \sigma^j; h^\star = h_j$
11: $\quad\quad\quad$ **end if**
12: $\quad\quad$ **end for**
13: $\quad\quad \mathbf{C} = \mathbf{C} +_w \sigma^\star$ ▷ Update covered elements
14: $\quad\quad \Omega^\star = \Omega^\star \cup \{\sigma^\star\}$
15: $\quad$ **end for**
16: $\quad$ **return** $\Omega^\star$
17: **end procedure**

the score and, if the score for candidate $\sigma^j$ is greater than the largest found so far, the candidate $\sigma^\star$ and its corresponding score $h^\star$ are updated. Finally, Lines 13 and 14 update the coverage $\mathbf{C}$ and the list of selected viewsheds $\Omega^\star$; at the end of the algorithm, $\Omega^\star$ is returned, also yielding the selected gNB positions.

#### 3.2.2.1 Score Functions

The final outcome of Algorithm 3.1 depends on the score function $\rho()$, which specifies, step-by-step, the metric for the best $\sigma^j$ selection. These functions express how the algorithm will push towards 1-, 2-, ... or $w$-coverage at each iteration. We introduce here three functions that assign a different importance to different levels of coverage; Sect. 3.2.5 analyzes their impact on the final outcome.

#### 3.2.2.2 $w$-coverage Maximization ($w$-CM)

The first score we consider is a simple 1-norm (sum of all the elements) on the candidate coverage $\mathbf{C}^*$ computed at Line 7 in Algorithm 3.1:

$$\rho(\mathbf{C}^*, w) = |\mathbf{C}^*|_1 \tag{3.7}$$

For $w = 1$ applying Eq. (3.7) in Algorithm 3.1 solves the canonical maximum set coverage problem. For $w > 1$ it returns the global weight of $\mathbf{C}^*$ elements, thus Algorithm 3.1 ends up in maximizing the weighted coverage, without any attempt to prioritize the points that have zero coverage

over the ones that have coverage between 1 and $w-1$. Notice that $\mathbf{C}^*$ are capped to $w$, so that points whose coverage is larger than $w$ do not influence the score. Besides its simplicity, this naive function completely disregards the difference between a point that is not served at all and one that already has coverage, but whose coverage is improved by adding the considered viewshed.

**$w$-COVERAGE FAIRNESS ($w$-CF)** Reducing the difference between different levels of coverage resembles a problem of fairness, thus we recall the well-known fairness index proposed by Ray Jain [82], that applied to our problem can be written in terms of matrix norms as

$$\mathrm{FI}(\mathbf{C}^*) = \frac{(|\mathbf{C}^*|_1)^2}{|\Lambda|(|\mathbf{C}^*|_2)^2} \tag{3.8}$$

the 2-norm is the square root of the sum of the squared elements of the matrix.

The fairness pushes all elements of $\mathbf{C}^*$ to a similar value, but alone does not provide a useful score, since we are interested in a wide, fair coverage and not only in a fair one (all 0-coverage is perfectly fair). This problem has been analyzed in the context of visual coverage in sensor networks in [70], and we adopt the same cost function proposed there, which, adapted to the problem and with the notation of this paper can be formalized as

$$\rho(\mathbf{C}^*, w) = \mathrm{FI}(\mathbf{C}^*) \times \frac{|\mathbf{C}^*|_1}{|\Lambda|w} = \frac{(|\mathbf{C}^*|_1)^2}{|\Lambda|(|\mathbf{C}^*|_2)^2} \times \frac{|\mathbf{C}^*|_1}{|\Lambda|w} \tag{3.9}$$

$\mathrm{FI}(\mathbf{C}^*)$ is weighted by the ratio between the 1-norm of the 1-coverage and the target $w$-coverage ($|\Lambda|w$), so that the first factor of Eq. (3.9) tries to balance the coverage and the second tries to extend it.

**$w$-COVERAGE GAP ($w$-CG)** The score defined by Eq. (3.9) is composite (a multiplication of two factors aiming at different goals), and its interpretation not always straightforward, thus we propose a third, simple metric whose goal is to weight the gap between the actual coverage and the target coverage with a quadratic function. Let $\mathbb{1}$ be the $m_x \times m_y$ matrix with 1 in all the positions that need to be covered and 0 otherwise. The target coverage can be expressed simply as $\mathbb{1} \cdot w$, and the score becomes

$$\rho(\mathbf{C}^*, w) = -(|\mathbb{1} \cdot w - \mathbf{C}^*|_2)^2 \tag{3.10}$$

We use the squared value of $|\cdot|_2$ for computational efficiency. Since this score measures a gap from the intended coverage, we use its negative value so that Algorithm 3.1 remains consistent in selecting the algebraic maximum.

When $w = 1$ Eq. (3.7) and Eq. (3.10) yield exactly the same result, while Eq. (3.9) returns a different numeric result (values are divided by $|\Lambda|^2$) but the obtained ranking is the same of the other two. For this reason, Sect. 3.2.5 presents only 1-CM results, while for other values of $w$ all metrics are reported.

#### 3.2.2.3 *Complexity*

Algorithm 3.1 is composed of two nested loops, the first one iterates over the number $k$ of gNBs that are to be deployed and the second one iterates over the number of potential locations $|\Omega|$. At each inner loop the algorithm computes a score $h_j$ by calling the function $\rho$. All score functions loop iteratively over all the values of $|\mathbf{C}^*|$, so they have time complexity $O(|\mathbf{C}^*|)$. This leads to a worst case time complexity of $O(k\,|\Omega|\,|\mathbf{C}^*|)$.

The memory required to execute the algorithm is bounded by the number of ground points ($\Lambda$), as the algorithm allocates a 2-dimensional matrix of size $(|\Lambda|, |\Omega|)$ (Line 2). At Line 7 we allocate a new temporary matrix $\mathbf{C}^*$ of the same size. Since we only need to store a boolean value for each cell of the matrix, we can use the smallest datatype available, which is `uint8`[1]. Thus the memory footprint is $2\,|\Lambda|\,|\Omega|$ bytes. The memory footprint of $\rho$ is constant.

### 3.2.3 Three-step heuristic

The heuristic presented in the previous section finds what we call the semi-optimal placement for the gNBs. Given a certain number $k$ of gNBs that the operator can afford to place, it provides the best locations where these gNBs should be placed. Iterating on values of $k$, we can answer the question: *what is the minimum number of gNBs and their positions to obtain w-coverage of a certain fraction of points on the street, e.g. 95%?* However, Algorithm 3.1 treats every $\sigma^j$ in $\Omega$ the same way, irrespectively of the building it corresponds to. This means that if convenient, the algorithm can place one gNB per building, or even all gNBs in the same building. Albeit theoretically possible, these configurations would be practically impossible or not cost-effective. To produce realistic results we need to introduce two practical constraints based on the cost estimation of a gNB deployment.

In a study done in 2020, Oughton et al. calculated the cost of a 5G small cell to be 20100$, of which only 3380$ were accountable to the radio[2] while the rest includes several fixed costs for constructions, hardware, backhaul fiber network, etc. that are needed to realize the gNB [83]. We assume that a multi-radio small cell can be realized keeping the fixed cost constant, and increasing the number of radios up to a maximum of 5 per gNB. This is a free parameter of our algorithm and can be adjusted to any other value.

We designed a heuristic that applies Algorithm 3.1 in three steps to limit the number of buildings on which to place gNBs and the number of radios per gNB: First, for every building, it finds the best 5 points that maximize the building visibility; Second, it finds the top $g$ buildings ordered by their

[1] Theoretically a single bit could be used, but the Numba architecture does not support it.

[2] These costs have been converted to Dollars ($) for the sake of readability, in the original research they were expressed in British Pounds (£).

cumulative visibility provided by the best 5 points; Third, among the best 5 points of the top $g$ buildings, it finds the overall best $k$ locations for the gNBs. The choice of $g$ is key for the final result, in fact, if the target of the operator is to provide $w$-coverage to a certain percentage of the points on the ground, if $g$ is too small such objective may not be reachable. On the other hand, the larger is $g$, the higher is the cost of the deployment. The optimal $g$ is of course scenario-dependent and can be found iterating on $g$ with our heuristic. Since reducing the number of buildings (and thus the cardinality of $\Omega$) makes the 3-step heuristic faster than the original one, the optimal $g$ for an area of roughly 0.7km$^2$ with 135 gNB/km$^2$ can be computed in around 1 minute.

Note that the formulation of the $\Gamma()$ function in Algorithm 3.1 is extremely generic as it takes a set $\Omega$ of viewsheds as an input, and it returns another set of viewsheds. We can thus change the input $\Omega$ and $k$ and apply the algorithm three times obtaining semantically different results, without the need to modify its internals.

#### 3.2.3.1 First step: Point selection

At the first step we limit $\Omega$ to the set of all points from a single building. Let $\Omega^r$ be the set of all the visibility matrices from points on the (dilated) perimeter of builing $r$, then:

$$\Omega^r = \{\sigma^j \; \forall \; p_j \in \; \mathcal{C}^r\} \tag{3.11}$$

We apply $\Gamma()$ using $\Omega^r$ as the first argument, and limiting the choice to 5 points. We obtain a list of viewsheds from the best 5 points in the building.

$$\Omega_5^r = \Gamma(\Omega^r, 5, w) \tag{3.12}$$

We repeat this step for every building $r$.

#### 3.2.3.2 Second step: Building Selection

Let $\mathbf{\Phi}^r$ be the ground visibility matrix of building $r$, i.e., the visibility matrix from the best 5 points on the facade of building $r$. $\mathbf{\Phi}^r$ is defined as the logical OR ($\bigvee$) of all the binary visibility matrices:

$$\mathbf{\Phi}^r = \bigvee_{\sigma^j \in \Omega_5^r} \sigma^j \tag{3.13}$$

and let $\Omega^R = \{\mathbf{\Phi}^r \; \forall \; r < R\}$ be the set of all the visibility matrices from all buildings. Invoking the function $\Gamma()$ using $\Omega^R$ and $g$ as an input, we obtain the cumulative viewsheds from the best $g$ buildings. Since the areas in which we run the experiments are different and contain a different number $R$ of buildings, in practical terms it is more convenient to refer to the percentage $X$ of buildings to be used, where $g = \lceil \frac{X}{100} * R \rceil$. We then define:

$$\Omega_X^R = \Gamma\left(\Omega^R, \left\lceil \frac{X}{100} * R \right\rceil, w\right) \tag{3.14}$$

that is the set of the best cumulative viewsheds from a percentage $X$ of the buildings of the area.

#### 3.2.3.3 *Third step: gNB Selection*

Step two returns a set of cumulative viewsheds obtained considering the best 5 points on $X$% of the buildings, and we use it to select the set $B_X$ of the indices of the best buildings:

$$B_X = \{r \,\forall\, r \,|\, \mathbf{\Phi}^r \in \Omega_X^R\} \tag{3.15}$$

and thus we can define the input to the third and last step of the heuristic:

$$\Omega_X = \{\sigma^j \in \Omega_5^r \,\forall\, r \in B_X\} \tag{3.16}$$

$\Omega_X$ contains all the possible viewsheds from each of the best 5 points (selected in step 1), from $X$% of the buildings that have the best cumulative viewshed (selected at step 2). As an example consider an area with 600 buildings each of which has an average of 100 points on its border. If $X = 10$ then $g = 60$, and the third step of our heuristic would explore at most $|\Omega_X| = 60 \times 5 = 300$ locations out of the 60.000 available.

Finally, we apply $\Gamma()$ again and obtain:

$$\Omega_{X,k}^{\star} = \Gamma(\Omega_X, k, w) \tag{3.17}$$

$\Omega_X^{\star}$ contains exactly $k$ viewsheds that correspond to $k$ points in space where to place gNBs. These places represent the semi-optimal choice among $\Omega_X$, with a maximum of 5 radios per building and $X$% of buildings.

Similarly to what we did passing from $g$ to $X$, it is more practical to target a desired density $\lambda$ of gNBs per squared km, rather than a number $k$. So, given a certain area and $\lambda$ we have:

$$\Omega_{X,\lambda}^{\star} = \Gamma(\Omega_X, \lambda * \text{area}, w) \tag{3.18}$$

We are now able to compare the coverage obtained in heterogeneous areas with a target density of gNBs per squared km, using $X$ as a tuning parameter to obtain the most favorable cost/performance trade-off.

Note that while the original heuristic has a bounded error, our three steps heuristic loses this theoretical property. However, when used with $X = 100$% the effect is to pick the best 5 points for each building, and then apply the bounded heuristic to this restricted set of points. This reduces the dimension of the problem, but also corresponds to a practical constraint in real deployments, since the number of gNBs per building cannot grow arbitrarily. The application of the three-step heuristic with $X = 100$% thus maintains the bounded error, and Sect. 3.2.5 shows that results with $X = 4$% are very close with a strong reduction of the cost of the infrastructure. Finally, since Algorithm 3.1 has polynomial complexity, its three-step application is again of polynomial complexity.

### 3.2.4 Experiments Set Up and Metrics

As already mentioned we evaluate 5 areas for each of the three densely populated Italian cities for a total of 15 scenarios. Trento, Firenze, and

Figure 3.3: Graphical rendering of the area choosen for the experiment (in dark grey), its buffer of radius $d_{\max}/2$ (in light grey), and the set of points on the street $\Lambda$ in black. © 2022 IEEE

Napoli, however, have quite different urban textures, so they represent a valid sample of different urban areas. The average size of the areas is roughly 0.7 km$^2$ and is limited by the computational power we had access to, as obtaining $\Omega^{\star}_{X,\lambda}$ is a computationally intensive task that requires handling tens of thousands of large integer matrices. To speed-up the algebra on large matrices we implemented the algorithms on an NVIDIA V100 GPU equipped with 5120 CUDA cores, which enables parallelization of matrix operations. The limit is given by the 16 GB of RAM of the GPU which limits computation to areas of up to 1 km$^2$. Results can be extended to larger areas straightforwardly with more resources, and with some effort by improving the memory space optimization.

Placing the gNBs in a constrained area causes a border effect: Points on the border of the area can be covered only by the gNBs placed inside the area, but not from the ones outside the area, which would be present in the real world coverage. For this reason, we enlarge the area where gNBs can be placed with a guard band of width $d_{\max}/2$ m as shown in Fig. 3.3 (light gray area), but we measure coverage only in the inner area (dark gray area).

#### 3.2.4.1 Evaluation Metrics

To evaluate the performance of different algorithms and their parameters we calculate the effective coverage, that is the sum of all the viewsheds $\sigma^i \in \Omega^{\star}_{X,\lambda}$:

$$\mathbf{\Phi}_{X,\lambda} = \sum_{\sigma^j \in \Omega^{\star}_{X,\lambda}} \sigma^j \tag{3.19}$$

This matrix, with the same dimensions of $\mathbf{E}$, contains values ranging from 0 (no LoS with any gNB) to $k = \lambda * \text{area}$ (LoS with all gNBs). Fig. 3.4 shows the effective coverage on a sample area of Napoli.

**Figure 3.4:** Graphical rendering of the coverage $\mathbf{\Phi}_{[4,45]}$ (shades of green), and the set of optimal points (black dots) in a portion of one of the considered areas, in Napoli.

Every experiment is repeated varying $\lambda$, $X$ and $w$ as reported in Tab. 3.2; in graphs we report the confidence interval on the 15 areas with confidence level $\alpha = 0.95$ as error bars.

**COVERAGE** The coverage metric, analyzed as a function of $X$, counts the number of points with $w$-coverage normalized by the total number of points on the public street;

$$c_w(\lambda) = \frac{\gamma_w(\mathbf{\Phi}_{X,\lambda})}{|\Lambda|} \tag{3.20}$$

where $\gamma_w(\mathbf{\Phi}_{X,\lambda})$ is a function that counts the number of elements of $\mathbf{\Phi}_{X,\lambda}$ that are greater or equal than $w$. In the numerical results we show $c_1(\lambda)$, which expresses the portion of streets covered by at least one gNB, and $c_3(\lambda)$ the portion of streets covered with higher reliability (at least 3 gNBs). We also show the coverage distribution on the points of $\Lambda$, i.e., the Empirical Probability Density Function (EPDF) of the coverage values.

**RESISTANCE TO OBSTRUCTION** Fig. 3.5 depicts the ground projection of the rays that connect a point $p$ to three gNBs in two different cases. In both

| Parameter | Value | Descriptioon |
|---|---|---|
| $\lambda$ | {15, 45, 75, 105, 135, 165, 195 } | gNB density |
| $X$ | $\{2, 4, 100\}$% | Percentage of buildings |
| $d_{max}$ | 300 m | Maximum distance for the links [84] |
| $c_{gNB}$ | 16720 $ | Cost of a gNB without any radio [83] |
| $c_{radio}$ | 3380 $ | Cost of each radio of a gNB [83] |
| $w$ | {1,3} | Target coverage |

**Table 3.2:** Parameters used for the numerical results.

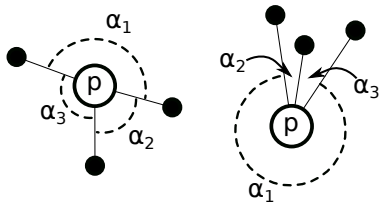



**Figure 3.5:** Ground projection of two examples of 3-coverage. Black dots are gNBs. © 2022 IEEE

scenarios, an obstacle that obstructs an angle $\alpha_2 + \alpha_3$ may totally shadow $p$, but in the right figure this is way more likely than in the left one, so $w$-coverage alone does not necessarily imply resistance to obstruction. We define $\alpha_c(p) = 360° - \max(\alpha_1, \alpha_2, \alpha_3)$ to be a measure of robustness against shadowing. It is easy to show that one single object that obstructs an angle smaller than $\alpha_c(p)$ cannot totally prevent LoS with some gNB, so the larger is $\alpha_c(p)$ the better it is. Let $\Lambda_2$ be the set of points with coverage larger or equal than 2, then

$$O_R = \frac{|\{p \in \Lambda_2 \,|\, \alpha_c(p) > 45°\}|}{|\Lambda_2|} \tag{3.21}$$

provides the fraction of points for which one obstacle that obstructs 45° does not prevent LoS. In the absence of finer metrics, $O_R$ provides a heuristic measure of how well the coverage can resist to shadowing.

**TOTAL COST** The cost function we use to evaluate the CapEx for the deployment of the network follows a cost model conceived for 5G networks [83]. From the model we extract a fixed part $c_{\text{gNB}}$ for the deployment of the gNB in a building, and a variable part for each radio interface of the node ($c_{\text{radio}}$):

$$\text{cost}(\lambda) = c_{\text{gNB}} * N_{\text{gNB}} + c_{\text{radio}} * k \tag{3.22}$$

where $N_{\text{gNB}}$ counts the number of buildings on which at least one gNB has been deployed and $k = \lambda * \text{area}$. The cost() value depends on the desired density of gNBs per km$^2$, so it is expressed as a function of $\lambda$, but it is also influenced by $X$. We also define upper and lower bounds of the cost function, as:

$$\begin{aligned} \text{UB} &= (c_{\text{radio}} + c_{\text{gNB}}) \cdot k \\ \text{LB} &= c_{\text{radio}} \cdot k + c_{\text{gNB}} \end{aligned} \tag{3.23}$$

The upper bound UB is given by the worst case where every coverage point is a single antenna gNB on a different building, while the lower bound LB is given by the (practically impossible) network where all radios are placed on a single building.

**MARGINAL COST** To estimate the Return on Investment (RoI) of deploying a robust network, we also evaluate the cost-effectiveness of adding new gNBs to further improve the coverage. We evaluate the cost of a deployment with growing densities $\lambda_i$. Since the heuristic is deterministic, if

$\lambda_{i-1} < \lambda_i$ then $\Omega^{\star}_{X,\lambda_{i-1}} \subset \Omega^{\star}_{X,\lambda_i}$, and the difference $\text{cost}(\lambda_i) - \text{cost}(\lambda_{i+1})$ is exactly the marginal cost of the added gNBs. Similarly, $c_w(\lambda_i) - c_w(\lambda_{i-1})$ is the marginal increase of relative coverage. Thus we can define the incremental cost metric $m_c$:

$$m_c(\lambda_i) = \frac{\text{cost}(\lambda_i) - \text{cost}(\lambda_{i-1})}{c_1(\lambda_i) - c_1(\lambda_{i-1})} \tag{3.24}$$

that provides an efficiency metric ($\$/m^2$) to estimate how cost-effective it is to increment the density of gNBs to improve coverage.

### 3.2.5 Results

We present the results following the same order we introduced for the metrics: coverage, total cost, and marginal cost.

#### *3.2.5.1 1-Coverage*

Fig. 3.6 reports $c_w(\lambda)$ for $X = 2, 4, 100\%$ and $w = 1, 3$. Each point is the average of all 15 areas we consider; vertical bars are the 95% confidence intervals. Since areas are never overlapping and thus they are independent, we can safely consider their average coverage as the outcome of an i.i.d. random variable, thus the confidence interval can be computed as a Student-t distribution with 15 degrees of freedom. The confidence intervals are reasonably compact, thus we deem that the results are reliable.

If we focus on the first row (1-coverage) the plots show one key point of our analysis, that it is possible to obtain a very large $c_1$ in urban areas with $\lambda$ below, or very close to 100. This confirms that in urban areas a very high LoS coverage is achievable with a number of gNBs that is close to what is expected (roughly ten times what is used for 4G, so $\lambda \simeq 100$). We see in fact that with $X = 100$ and with $X = 4$ all the curves reach $c_1 = 80\%$ ($\lambda = 45$), $c_1 = 90\%$ ($\lambda = 75$), and 1-CM and 3-CG reach $c_1 = 95\%$ ($\lambda = 105$). We stress that these results were obtained with realistic assumptions: a limited, realistic number of devices per building and a precise ray-tracing model based on real data. All curves follow a trend expected in a set covering problem, with a steep initial rise followed by a saturation phase.

1-CM provides the highest 1-coverage, which is consistent with its design, while it is important to note that 3-CG, when evaluated on $c_1$, dominates the other 3-coverage strategies on values of $c_1 > 80\%$. The difference is not very large, but it is consistent in the saturation phase, in which it is extremely expensive to gain even a single percentage point, thus this very simple score function serves well both coverage and robustness. We also see that with $X = 2\%$ all strategies stay below $c_1 = 90\%$, while as said, with $X = 4\%$ we can achieve results that are very close to $X = 100$.

**N-COVERAGE AND RELIABILITY** Let us now focus on the 3-coverage row, which confirms that $X = 2\%$ is outperformed by the other two configurations (and thus, we do not comment on results with $X = 2\%$

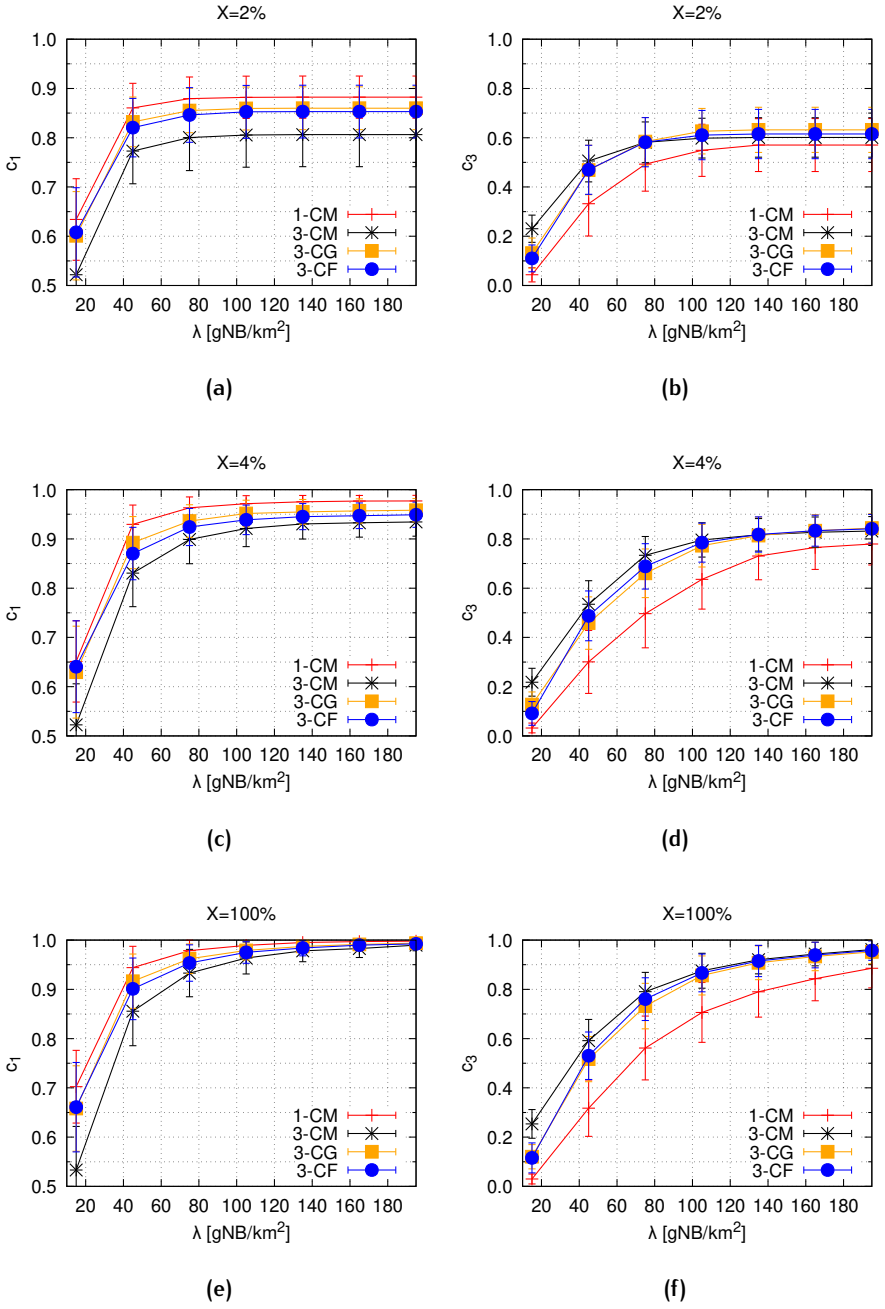



**Figure 3.6:** Coverage metrics $c_1$ and $c_3$ for different values of $X$ and different score functions evaluated on all the 15 areas considered. © 2022 IEEE

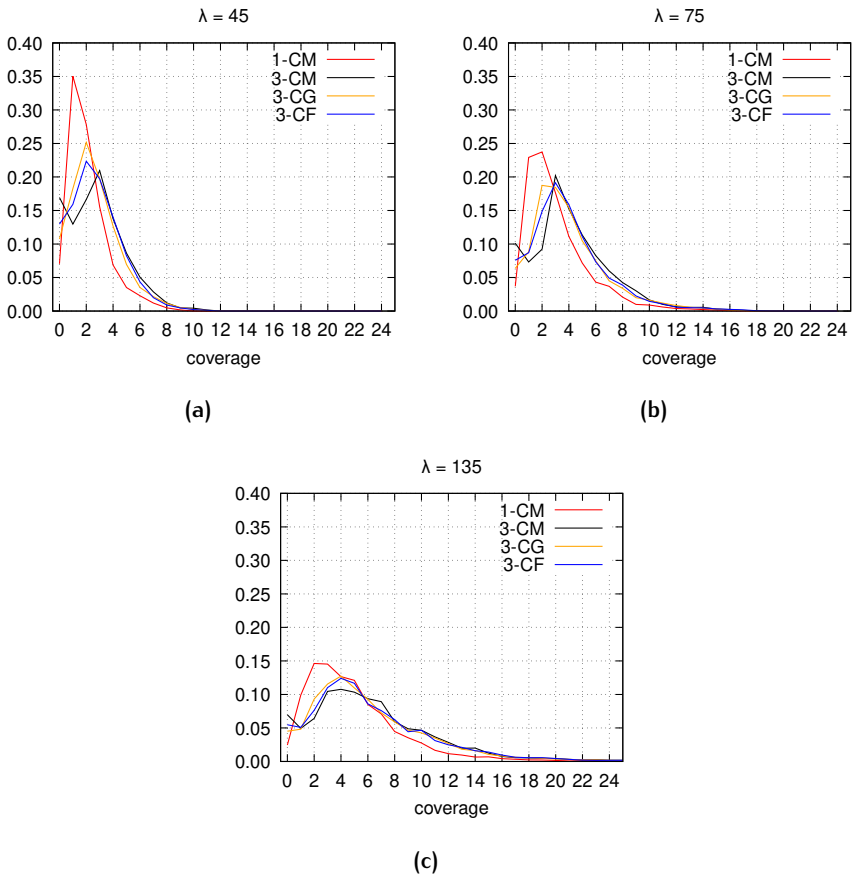



**Figure 3.7**: EPDF of the coverage for $X = 4$ and different $\lambda$. © 2022 IEEE

anymore). It is evident that a remarkable difference between 1-CM and the 3-coverage strategies exists. Among the latter ones, 3-CM is the one that performs slightly better than others, as it focuses only on 3-coverage and disregards other possible goals. In this case, the difference in the saturation point between $X = 4$ and $X = 100$ is more evident, as with $X = 4$ the highest 3-coverage is 84%.

The take-away from this set of results is that if an operator wants just to provide 1-coverage, it should use 1-CM, and can reach 95% coverage with $\lambda = 75$ and $X = 4\%$. This is the most cost-effective solution, as shown later, but obviously is also the most fragile. If the operator is interested only in 3-coverage, because its target customers need extremely high reliability, then the best choice is 3-CM, calibrating $X$ on the desired coverage. Most likely the operator is interested primarily in 1-coverage (as it provides basic connectivity) and with lower priority in 3-coverage (for users that need high reliability). In this case, 3-CF and 3-CG are the best choices, as they perform very close to 1-CM in terms of 1-coverage while being very close to 3-CM, outperforming 1-CM, in terms of 3-coverage.

To better understand the behavior of the coverage as a function of the gNB density, and why it tends to saturate before reaching 100% coverage, Fig. 3.7 shows the EPDF of the coverage multiplicity for $X = 4\%$ and three values of $\lambda$. First, it is clear that to reach a good 1-coverage some areas are covered many times, with the tail of the coverage distribution growing significantly. Second, 3-CM penalizes the 1-coverage to the point that the probability of 0-coverage is higher than the probability of strictly 1-coverage. The mass of the distribution is different between 1-CM and the 3- score functions, but the tails remain very close to one another suggesting that there are geometric properties that force a very high degree of coverage in some areas when trying to cover the areas that have remained uncovered.

Fig. 3.8 shows the robustness to obstruction $O_R$ for all points $p \in \Lambda_2$ for different $\lambda$ values. We report also the bar for 1-CM for completeness, but previous results show that 2-coverage for 1-CM is bad (for each reported $\lambda$ slightly more than 50% of 2-coverage compared to the worst of the other strategies), thus $O_R$ should be weighted by its lower 2-coverage and it is not really meaningful. The other strategies perform similarly, with a fairly high fraction of points (between 63% and 77%) that resists to shadowing from an obstacle that obstructs a 45 degree angle, and 3-CG better than the others in three cases. On the other hand, increasing the density of gNBs by a factor of 3 improves $O_R$ only about 10%, which indicates that to improve this result a different heuristic that takes into account this specific objective should be considered.

Finally, Fig. 3.9 presents the Empirical Cumulative Density Function (ECDF) of the length of the link to the closest gNB for every covered point. We report this metric to show that with any value of $\lambda$ the largest majority of the best links are below 100 meters of length, and the 95% of them are below 214 meters (less than 200 meters for higher density), so our choice

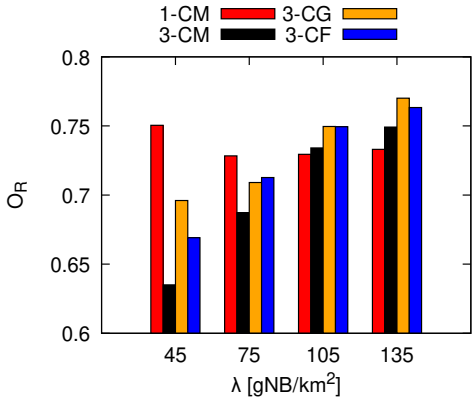



**Figure 3.8:** $O_R$ for all points with coverage larger or equal 2, $X = 4\%$. © 2022 IEEE

of setting the maximum link length to 300 meters only marginally impacts the results.

#### 3.2.5.2 *Total Cost*

Fig. 3.10 shows the cost of each deployment as a function of $\lambda$ for $X = 4\%$; and $X = 100\%$. Changing $X$ gives rise to very different trends, with the cost for $X = 4\%$ that grows remarkably more slowly than the cost for $X = 100\%$. This is the beneficial effect of limiting the number of buildings. Fig. 3.11 reports the trends of the cost using as a free variable the coverage, with vertical lines to highlight key coverage values (90% and 95% in $c_1$ and 84% in $c_3$). The cost for $X = 100\%$ is always higher than the cost for $X = 4\%$ at the same coverage: our heuristic with $X = 4\%$ may need a higher $\lambda$ compared to $X = 100\%$ (as Fig. 3.6 shows), however, the total cost of for $X = 100\%$ is higher because it uses more buildings. This results highlight two facts. First, the cost to ensure nearly complete $w$-coverage increases very fast with a sort of asymptotic behavior independently from

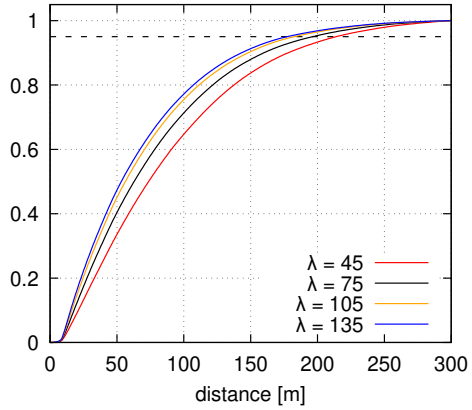



**Figure 3.9:** ECDF of the length of the link to the nearest BS. The dashed lines marks the 95th percentile. © 2022 IEEE

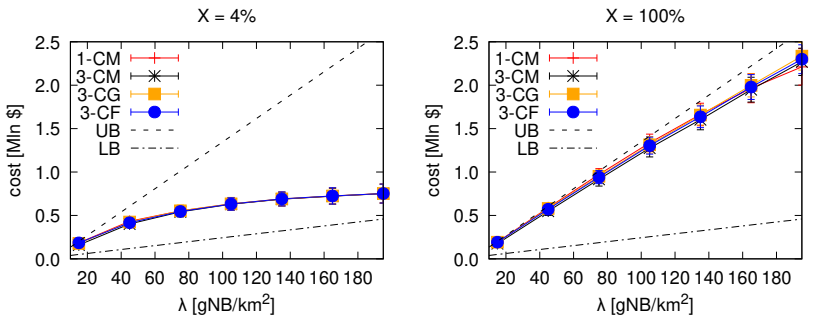



**Figure 3.10:** Deployment cost of the network with the four score functions and for $X = 4$ and 100%. © 2022 IEEE

$w$. Second, and rather obvious, including all buildings in the selection allows a better coverage again independently from $w$.

We quantify this increase in Tab. 3.3 reporting the relative cost difference between $X = 4\%$ and $X = 100\%$ with the 1-CM and the 3-CG strategies at the same coverage value (the intercept of the vertical lines and the curves in Fig. 3.11). Not limiting the number of buildings produces an increased cost of 19% and 36% for $c_1$ and a probably unacceptable 70% for $c_3$.

### 3.2.5.3 *Marginal Cost*

Fig. 3.12 reports the marginal cost per m$^2$ $m_c(\lambda)$. Each point is the ratio between the increased cost and the increased covered area (Eq. (3.24)). The figure shows that $c_1$ (left plots) monotonically increases for the 1-CM strategy, with very high costs when the gNB density becomes very high ($\lambda = 160, 180$). This is consistent with the design of the metric as it is trying to cover every single uncovered point, at the cost of placing new gNBs that cover just a few squared meters. The other metrics instead grow up to a certain coverage and then oscillate with a more noisy trend. This is

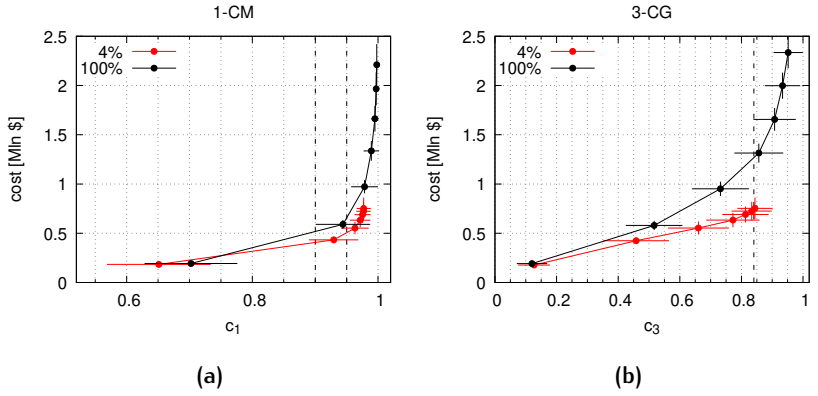



**Figure 3.11:** Deployment cost of the network as a function of coverage for 1-CM and 3-CG. © 2022 IEEE

| | coverage | 1-CM | 3-CG |
|---|---|---|---|
| | 0.84 | 19% | 24% |
| $c_1$ | 0.90 | 27% | 24% |
| | 0.95 | 30% | 36% |
| $c_3$ | 0.84 | – | 70% |

**Table 3.3:** Cost increase passing from $X = 4\%$ to $X = 100\%$.

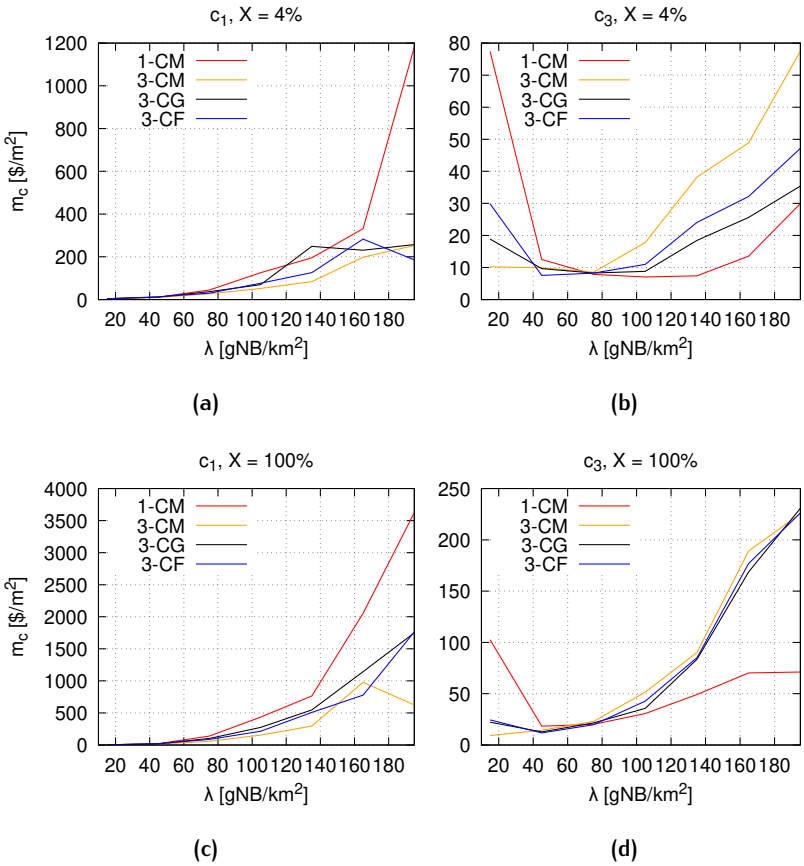



**Figure 3.12:** Marginal cost per covered $m^2$ for different $w$-coverage and strategies. © 2022 IEEE

an artifact of the metric, as the incremental cost is monotonously growing as in Fig. 3.10, but the incremental covered area is not monotonously growing for strategies that use $w = 3$ that do not simply try to maximize $c_1$.

Looking at the right column, we see that the marginal cost per $m^2$ of $c_3$ has a different trend. Initially, the cost for all strategies decreases, as the 3-coverage is very low (see Fig. 3.6) and it is easy to improve it. When the coverage curves start to saturate, then $m_c$ increases. The 1-CM strategy has a lower $m_c$ simply because it has a lower $c_3$ and thus, it is not saturating even at high values of $\lambda$.

Our approach makes it possible to obtain these data that are vital for operators, because they help to identify a realistic maximum value for coverage, beyond which it is not cost-effective to add more gNBs.

## 3.3 OPTIMAL GNB PLACEMENT FOR VEHICULAR NETWORKS

In this research, we extended the heuristics and the algorithms used in the previous section, focusing on the placement of gNBs for vehicular communication. This is accomplished by devising a vehicular demand model that, based on simulated traffic patterns, weights each point on the public roads. Given this demand model, we developed a new heuristic that exploits it to find the optimal locations for gNBs. Additionally, we evaluated the quality of the coverage in terms of capacity, using the International Telecommunication Union (ITU) Pathloss model, seen in Sect. 2.2.1.3, together with the Shannon capacity formula, seen in Sect. 2.2.5 for a mmWave access network at 28Ghz with 400MHz bandwidth.

The analysis has been performed on open-data from the city of Luxemburg, together with a realistic traffic model of the same city [85] leads us to the following findings:

- A "*reasonably low*" density (15 gNBs/km$^2$) can be used to provide 95% coverage to vehicles, while pedestrian areas require a higher density (35 gNBs/km$^2$);
- NLoS links can be useful to fill the gap but the performance penalty is extremely high;
- Providing connectivity to vehicles and pedestrians are two competing objectives, as a small part of the walkable area is on the streets. If the operator needs to guarantee service continuity also to pedestrians, it needs to adopt a different covering strategy.

### 3.3.1 Problem Formulation and Solution

Similarly as in Sect. 3.2, we start from the DEM $\mathcal{D}$ and its associated matrix $\mathbf{E}$. We define a set $\Lambda$ of points in the ground that can be potentially covered with a LoS connection from a gNB. Each point corresponds to an $(x, y, z)$ triplet, in which the $(x, y)$ coordinates are quantized using one point per squared meter. Points are selected to be outside any building shape and only in public areas (streets, roundabouts, street parking, sidewalks) and not in private areas. Each point in $\Lambda$ is uniquely identified by the $(x, y)$ couple, as we set the $z$ value to a height of 1.5m, as such, with a little abuse of notation we may use $(x, y) \in \Lambda$ when it simplifies the description.

The problem we tackle can be summarized in three steps described in the next two sections:

1. For each point in $\Lambda$ define a weight, the higher the weight the higher the probability the point will be covered. The matrix that associates $(x, y)$ to a generic weight is called $\boldsymbol{\tau}$;
2. Identify the set $\mathcal{P}$ of points in space in which a gNB could potentially be placed. Each position $p_i \in \mathcal{P}$ is defined by an $(x, y, z)$ triplet, we consider only points on the facades of buildings;

3. Find an algorithm that chooses the minimal number of gNBs so that the coverage is maximized according to some metric.

Step one is a novel contribution of this research. The solution to the second step comes from the work described in Sect. 3.2, while the third step modifies the solution proposed therein to take into account the weights introduced in step one. In the remainder of this section, we will formalize point 3 (the gNB placement problem) and provide an algorithmic solution for a generic weights matrix $\tau$. Then, in Sect. 3.3.2 we detail the different strategies to obtain 2 realistic weights matrices for vehicles and for pedestrians. For convenience, we list all the mathematical symbols in Tab. 3.4.

It is important to note that all our results are to be interpreted as a base-line on which to produce fine-grained further results, as we are considering only the obstruction due to buildings.

#### 3.3.1.1 gNB Placement

We briefly recall what was proposed in Sect. 3.2.2 that is at the base of this work. Let $\mathcal{B} = \{b_i\}$ be the set of buildings extracted from the OSM dataset. Let also $\phi(b_i)$ be a function that extracts a set of coordinates that compose the perimeter of the building $b_i$, with points spaced on average one meter away from each other, placed 1 m below the height of the roof. We can then define the set of candidate locations $\mathcal{P}$ as:

$$\mathcal{P} = \bigcup_{b_i \in \mathcal{B}} \phi(b_i) \tag{3.25}$$

Once the set of candidate locations is determined, we need to evaluate the coverage from each one of them. In order to do so, we take advantage of the viewshed algorithm mentioned in Sect. 2.3.5.2. This algorithm, applied on a highly precise Digital Surface Models (DSM) computes the presence of LoS from the candidate location $p_i$ to each point in $\Lambda$ given a maximum distance ($d_{max} = 300\,\text{m}$) from $p_i$. We consider this value of distance as it has been used as an upper bound in other works [84].

Let $\sigma^i$ be the *viewshed* matrix (or ground visibility matrix) from point $p_i$, defined as in Eq. (2.25). If we compute the *viewshed* matrix from all the points in $\mathcal{P}$, we obtain a collection of matrices that represent all the possible viewsheds from all the potential positions of gNBs, which we call $\Omega = \{\sigma^1, \sigma^2, \ldots, \sigma^m\}$. In which $\sigma^i_{x,y} = 1$ means that a terminal in position $(x, y)$ has LoS with a gNB placed in the point $p_i$.

Remember that we extend the classic definition of LoS with the additional constraint of being within a maximum distance $d_{max}$.

#### 3.3.1.2 Quasi-optimal gNB placement

We want to find a subset $\Omega^\star \subseteq \Omega$ whose size is lower than a parameter $k$ ($|\Omega^\star| \leq k$) that maximizes the elements coverage $\left|\bigvee_{\sigma^i \in \Omega^\star} \sigma^i\right|$, where $\bigvee$ is the OR operator between binary matrices and $|\cdot|$ is the norm-1 operator

**Table 3.4:** List of symbols

| Symbol | Meaning |
|---|---|
| $\Lambda$ | Set of ground points potentially to be covered |
| $\boldsymbol{\tau}$ | matrix of weights associated to points in $\Lambda$ |
| $\mathcal{P}$ | Set of all candidate locations for gNBs |
| $\mathcal{B} = \{b_i\}$ | Set of buildings extracted from the OSM dataset |
| $\phi(b_i)$ | Set of coordinates that compose the perimeter of the building $b_i$ |
| $\boldsymbol{\sigma}^i$ | binary matrix representing the viewshed from the point $p_i$ |
| $\Omega$ | Set of all viewsheds corresponding to all possible gNB positions |
| $\Omega^\star$ | Quasi-optimal set of viewshed computed by the optimization algorithm |
| $\boldsymbol{\pi}, \boldsymbol{\gamma}$ | Weight matrices for the pedestrian and vehicular strategies, respectively |
| $\Lambda_\pi, \Lambda_\gamma$ | Set of ground points to be covered by the pedestrian and vehicular strategies, respectively. Named as walkable and drivable areas in the text |

(the sum of all elements). In order to take into account the traffic patterns, consider a generic non-negative weight matrix $\boldsymbol{\tau}$ with the same shape of $\boldsymbol{\sigma}^i$. We can formulate the maximization objective as follows:¸

$$\max_{\Omega^\star} \left| \boldsymbol{\tau} \odot \left( \bigvee_{\boldsymbol{\sigma}^i \in \Omega^\star} \boldsymbol{\sigma}^i \right) \right| \quad \text{with } |\boldsymbol{\sigma}^i| \leq k \tag{3.26}$$

Where $\odot$ is the Hadamard product (the element-by-element multiplication between two matrices). This will lead to a choice of the optimal $k$ viewsheds in $\Omega^\star$ to cover the roads with the highest traffic. The problem is a so-called weighted maximum coverage problem.

Note that if we call $\mathbb{1}$ the matrix made of all one elements, and we set $\boldsymbol{\tau} = \mathbb{1}$ then the problem converges to the classical unweighted maximum coverage problem, in which we try to cover the largest portion of the points in $\Lambda$ treating all of them equally.

#### 3.3.1.3 *Heuristic solution*

The above-described coverage problem is NP-Hard, as the one previosuly described in Sect. 3.2.1.2. Here we modify the greedy heuristic as described in detail in Algorithm 3.2 to take into account the weight matrix. The heuristic proceeds as follows: we start by defining a coverage matrix $\mathbf{C}$ of the same size of $\boldsymbol{\tau}$, initialized with zeros (Line 2). Each iteration of the loop in Line 4 will choose the position of one gNB. For each candidate location $p_i$ and the corresponding viewshed $\boldsymbol{\sigma}^i$ we derive the so-far uncovered elements $\bar{\boldsymbol{C}}$ as the negation of the coverage matrix (Line 7). The idea is, at each step, to progressively add the gNB that provides the largest additional coverage with respect to the already covered area.

We define $\mathbf{C}^\star$ that represents the so-far uncovered elements that would be covered by the candidate location, with their weight given by $\boldsymbol{\tau}$ (Line 8). Note that $bool()$ is a function that makes an integer matrix a boolean one,

$\neg$ is the boolean NOT operand. We then provide a score for $p_i$ as the norm-1 of the coverage matrix (in Line 9).

Then, the element with the maximal ranking that is not already in $\Omega^\star$ ($\sigma^j \notin \Omega^\star$) is chosen and the corresponding values of the viewshed matrix are added to $\mathbf{C}$ (Line 15). Note that this makes $\mathbf{C}$ a non-boolean matrix. Finally, the viewshed with the maximal ranking $\sigma^\star$ is added to the set of optimal viewsheds (Line 16). The loop is repeated till the number of desired locations is reached. The operation at line Line 8 has complexity $|\Lambda|$, and is repeated at most $k \times |\mathcal{P}|$ times, so the overall complexity is $O(k|\mathcal{P}||\Lambda|)$.

This algorithm is referred to as $\Gamma(\Omega, k, \boldsymbol{\tau})$ and in the next section we use it with two weights matrices $\boldsymbol{\pi}$ or $\boldsymbol{\gamma}$, producing two sets of quasi-optimal viewsheds optimized for pedestrians or vehicles respectively.

### 3.3.2 A Demand Model for Vehicles and Pedestrians

Obtaining realistic traffic data, for both pedestrians and vehicles, is always a challenging task, as data collected by cities is rarely released to the public. One possibility, which is the one we consider in this work, is to generate traffic data using microscopic traffic simulators and realistic scenarios. We use the urban traffic simulator SUMO [86] to generate realistic mobility traces of the city of Luxembourg. In particular, we make use of the Luxembourg SUMO Traffic (LuST) scenario [85], a publicly available scenario generated from traffic data provided by the Luxembourg government which includes both public and private transportation over a period of 24 h. Unfortunately, since the model does not include pedestrian mobility, we had to resort to a different approach to model it. In the next sections, we detail both models.

#### *3.3.2.1 Vehicular Model*

To obtain traffic traces, we run the scenario over the full 24 h for a total of 286 215 vehicles moving on the streets. The simulation step is set to 1 s and, at the same frequency, we log the positions of the vehicles in the

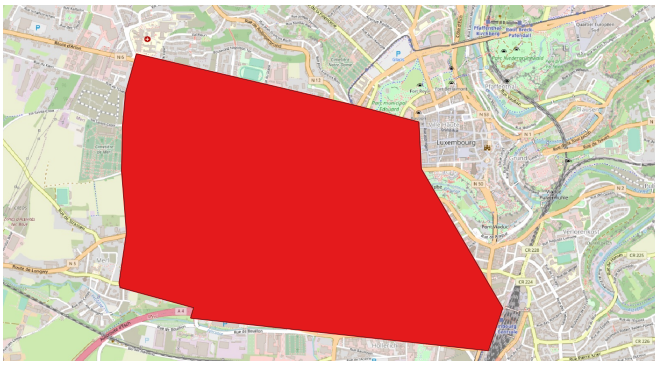


**Figure 3.13:** Area of the city of Luxembourg over which traces are collected. ©2023 IFIP

**Algorithm 3.2** Greedy algorithm for the weighted maximum coverage problem.

**Require:** $\Omega$ (Set of viewsheds), $k$ (number of gNBs), $\boldsymbol{\tau}$(weighted traffic matrix)
**Ensure:** $\Omega^\star$ (Set of the viewsheds from optimal locations)

**procedure** $\Gamma(\Omega, k, \boldsymbol{\tau})$
  $\mathbf{C} = \mathbf{0}$
  $\Omega^\star = \{\}$
  **for** $i \leftarrow 1$ to $k$ **do**
    $h^\star = -\infty$
    **for** $\sigma^j \in \Omega$ **do**
      $\bar{\mathbf{C}} = \neg bool(\mathbf{C})$
      $\mathbf{C}^\star = \bar{\mathbf{C}} \odot \sigma^j \odot \boldsymbol{\tau}$
      $h_j = |\mathbf{C}^\star|$
      **if** $h_j > h^\star$ and $\sigma^j \notin \Omega^\star$ **then**
        $\sigma^{i^\star} = \sigma^j$
        $h^\star = h_j$
      **end if**
    **end for**
    $\mathbf{C} = \mathbf{C} + \sigma^{i^\star}$
    $\Omega^\star = \Omega^\star \cup \{\sigma^{i^\star}\}$
  **end for**
  **return** $\Omega^\star$
**end procedure**

area of the city shown in Fig. 3.13, corresponding to an area $S$ of roughly 4 km$^2$. We collect traces using GPS (latitude/longitude) coordinates and then convert them to a `.gpx` file for later processing. We map each point in a trace to a cell in the discretized space given by $\Lambda$. We obtain a matrix $\gamma$ with the same shape as $\boldsymbol{\tau}$, where $\gamma_{x,y} = n$ means that $n$ vehicles have passed in cell $(x, y)$ during the whole simulation[3].

For the sake of readability we rescale it to the number of passages per minute, Fig. 3.17 shows the EPDF of the values of the cells with non zero value, binned with bins of size 0.125 passages/minute. It can be seen that the majority of the cells have less than one passage per minute, with the 95th percentile roughly at 0.55. The distribution is pretty skewed, with about 5 orders of magnitude between the largest and the lowest frequency.

We call $\Lambda_\gamma$ the set of points $(x, y, z) \in \Lambda$ in which $\gamma_{x,y} \neq 0$, and we refer to it as the *drivable* area.

[3] In our previous work [87] due to memory limitations of our GPU we had to re-scale the frequency values in $\boldsymbol{\tau}$ to be within the allowed range of one byte: $[0, 255]$. In this work we resort to a GPU with a larger memory, enabling us to work with 16 bit integers and get rid of such re-scaling.

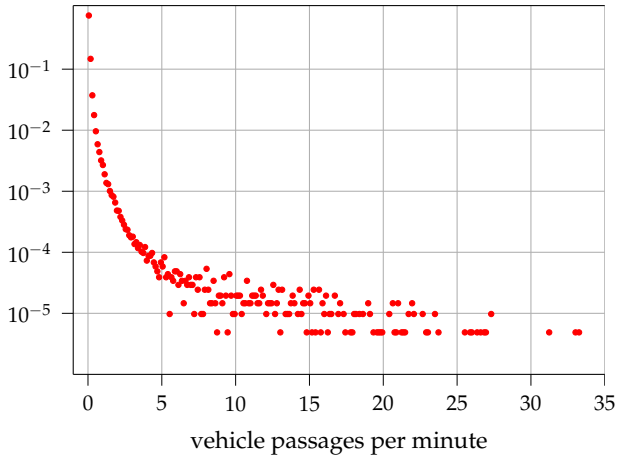



**Figure 3.14:** EPDF of the vehicles passages per minute per cell. ©2023 IFIP

#### 3.3.2.2 *Pedestrian Model*

Due to the lack of a pedestrian mobility model in the LuST scenario, we had to manually define certain public areas where the pedestrians would transit. We decided to focus on sidewalks and public open-air areas such as parks, gardens, squares, etc.

To model the sidewalks we started from the lines characterizing the OSM roads, which were used to center two symmetrical 2-meter wide sidewalks with a distance from the center of the road depending on the type of road. To model the public open-air areas, on the other hand, we relied on the public land use database from OSM, where we selected the public areas with one of the following types: `cemetery`, `forest`, `grass`, `heath`, `meadow`, `orchard`, `park`, `recreation_ground`. Again, we mapped those points to a cell in the discretized space given by $\Lambda$ and obtain a matrix $\boldsymbol{\pi}$ with the same shape as $\boldsymbol{\tau}$, where $\boldsymbol{\pi}_{x,y} = 1$ when the element corresponds to one of those areas and 0 otherwise. Fig. 3.14 shows both sidewalks and public areas in green, the roads dedicated to vehicles in yellow, and buildings in gray. We call $\Lambda_{\boldsymbol{\pi}}$ the set of points $(x, y, z)$ in which $\boldsymbol{\pi}_{x,y} \neq 0$, and we refer to it as the *walkable* area.

### 3.3.3 Experimental Setup and Metrics

We consider two different settings, one in which we optimize the coverage for the vehicular traffic ($\boldsymbol{\tau} = \boldsymbol{\gamma}$), and another one in which we optimize for the areas where pedestrians might be located ($\boldsymbol{\tau} = \boldsymbol{\pi}$). We apply Algorithm 3.2 to compute the optimal locations for the gNBs, increasing their number $k$. We consider a density $\lambda$ of gNBs per squared km going from 5 to 45 at steps of 5, and we set $k = \lambda S$.

We obtain two solutions for the coverage:

$$\Omega^{\star}_{\lambda,\boldsymbol{\gamma}} = \Gamma(\Omega, \lambda * S, \boldsymbol{\gamma}) \tag{3.27}$$

$$\Omega^{\star}_{\lambda,\boldsymbol{\pi}} = \Gamma(\Omega, \lambda * S, \boldsymbol{\pi}) \tag{3.28}$$

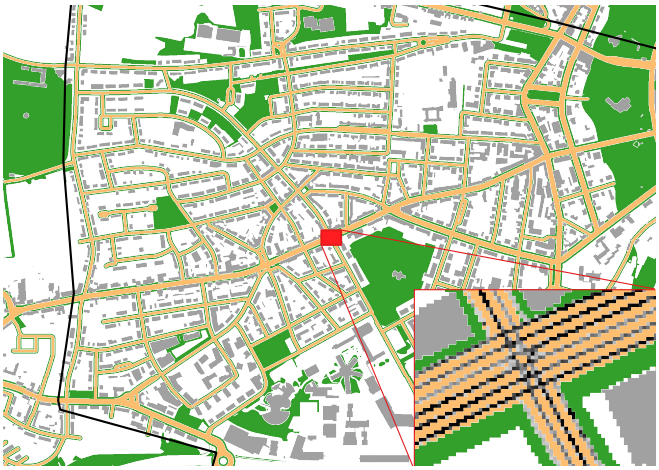


**Figure 3.15:** Detailed view of a portion of the area considered in the analysis. Yellow highlights roads, green indicates public areas for pedestrians (sidewalks and parks), grey indicates buildings, and white indicates private areas. The zoomed portion shows the vehicular traces with a grayscale representing the number of passages per point (darker means higher).

| Parameter | Value |
|---|---|
| Area size | 3.98 $km^2$ |
| Carrier frequency | 28 GHz |
| Bandwidth (B) | 400 MHz |
| Thermal Noise (T=300 K) | −87.8 dBm |
| Noise Figure ($N_t$) | 5 dB |
| 3GPP Channel Model | ETSI TR 38.901 Urban Micro |
| Reception gain | 3 dBi |
| MIMO layers ($\mu$) | 2 |
| Transmission power | 30 dBm |
| Transmission gain | 10 dBi |
| Maximum distance for LoS links | 300 m |

**Table 3.5:** Simulation Parameters

that we aggregate with the OR operator to obtain a full coverage matrix:

$$\Phi_{\lambda,\gamma} = \bigvee_{\sigma^j \in \Omega^\star_{\lambda,\gamma}} \sigma^j; \quad \Phi_{\lambda,\pi} = \bigvee_{\sigma^j \in \Omega^\star_{\lambda,\pi}} \sigma^j \tag{3.29}$$

In brief, the elements in $\Phi$ indicate whether a point (x, y) is covered by at least one gNB.

#### 3.3.3.1 Coverage Metrics

We use four metrics to compare the results in terms of coverage (again, $\odot$ is the Hadamard product and $|\cdot|$ the norm-1). The first two refer to the drivable area and are:

$$Dcov_{\text{ve}}(\lambda) = \frac{|\boldsymbol{\gamma} \odot \Phi_{\lambda,\gamma}|}{|\boldsymbol{\gamma}|}; \quad Dcov_{\text{pe}}(\lambda) = \frac{|\boldsymbol{\gamma} \odot \Phi_{\lambda,\pi}|}{|\boldsymbol{\gamma}|} \tag{3.30}$$

The metrics in Eq. (3.30) tell how good the coverage of drivable areas is when we optimize for vehicles ($Dcov_{\text{ve}}$) or when we optimize for pedestrians ($Dcov_{\text{pe}}$). $Dcov_{\text{pe}}$, in practical terms, tells us what happens if we try to optimize the coverage for pedestrians but we measure the results only on the points where the vehicles pass (with their multiplicity). Of course, we expect $Dcov_{\text{ve}}(\lambda)$ to be larger than $Dcov_{\text{pe}}(\lambda)$, yet we are interested in the difference.

We use two more metrics to evaluate the complementary set-up:

$$Wcov_{\text{pe}}(\lambda) = \frac{|\boldsymbol{\pi} \odot \Phi_{\lambda,\pi}|}{|\boldsymbol{\pi}|}; \quad Wcov_{\text{ve}}(\lambda) = \frac{|\boldsymbol{\pi} \odot \Phi_{\lambda,\gamma}|}{|\boldsymbol{\pi}|} \tag{3.31}$$

Both metrics express how good the coverage of walkable areas is, in the first case when we optimize for pedestrians ($Wcov_{\text{pe}}$) while, in the second, when we optimize for vehicles ($Wcov_{\text{ve}}$).

#### 3.3.3.2 Channel Capacity Model

Besides the evaluation on pure coverage, we also evaluate the quality of the coverage in terms of capacity, using the Shannon channel capacity formula. To obtain the received signal strength we consider a transmission power of 30 dBm at a frequency of 28 GHz. To compute the pathloss between a gNB located at $p_i$ and a point $(x, y, z)$ in the city, we employ the Urban Micro model defined in ETSI TR 38.901 [29], detailed in Sect. 2.2.1.3. The model provides the path loss for both LoS and NLoS conditions, and we can easily distinguish between the two cases thanks to the pre-computed viewsheds. In addition, we consider a gNB to be in LoS with a point only if the distance between them is smaller than 300 m. We set the noise floor for a bandwidth $B = 400$ MHz to be $N_t = -87.8$ dBm and we compute the Signal-to-Noise-Ratio (SNR) and the corresponding Shannon channel capacity, as explained in Sect. 2.2.5. Tab. 3.5 details all the parameters.

Fig. 3.16, shows two interesting aspects. First, by looking at the distance between the two curves, we can understand that a terminal continuously

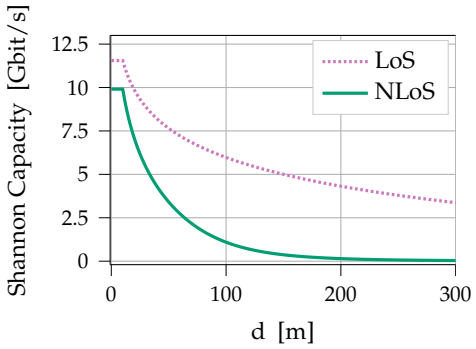



**Figure 3.16:** Shannon channel capacity as a function of the distance between two points $d$ and the LoS/NLoS conditions.

switching between LoS and NLoS would experience a constant change in network performance that would make it very hard to support any application, so a partial LoS coverage may be worse than a fully NLoS coverage. Second, by shifting the NLoS curve to the right, it can intersect the LoS one. This translates into the fact that in certain points it may be convenient to choose a gNB that is in NLoS but is physically closer to the closest one in LoS. This is more likely if the NLoS link is very short, and impossible if it is longer than roughly 50 m. It is interesting to evaluate if and how often this event happens.

#### 3.3.3.3 *Capacity Metrics*

Consider a point $(x, y, z) \in \Lambda_\gamma$ in the drivable area, the set $\Omega^\star_{\lambda,\gamma}$ of the quasi-optimal viewsheds at density $\lambda$ optimized for the vehicles, and the corresponding positions of gNBs. We call $LC_{x,y,\gamma}(\lambda)$ the highest capacity we can achieve from any gNB to $(x, y, z)$ using only LoS links[4]. If $LC_{x,y,\gamma}(\lambda) = 0$ then there is no LoS link between $(x, y, z)$ to any gNB. We call $NC_{x,y,\gamma}(\lambda)$ the highest capacity when using only NLoS links, which is always larger than zero (there are always NLoS links to any $(x, y, z)$). The same metrics are also defined for walkable areas using the $\pi$ pedix. We collect all the capacities in four sets, divided by LoS/NLoS:

$$Dcap^L_\lambda = \{LC_{x,y,\gamma}(\lambda) \; \forall \; (x, y, z) \in \Lambda_\gamma \mid LC_{x,y,\gamma}(\lambda) > NC_{x,y,\gamma}(\lambda)\} \tag{3.32}$$

$$Dcap^N_\lambda = \{NC_{x,y,\gamma}(\lambda) \; \forall \; (x, y, z) \in \Lambda_\gamma \mid NC_{x,y,\gamma}(\lambda) > LC_{x,y,\gamma}(\lambda)\} \tag{3.33}$$

$$Wcap^L_\lambda = \{LC_{x,y,\pi}(\lambda) \; \forall \; (x, y, z) \in \Lambda_\pi \mid LC_{x,y,\pi}(\lambda) > NC_{x,y,\pi}(\lambda)\} \tag{3.34}$$

$$Wcap^N_\lambda = \{NC_{x,y,\pi}(\lambda) \; \forall \; (x, y, z) \in \Lambda_\pi \mid NC_{x,y,\pi}(\lambda) > LC_{x,y,\pi}(\lambda)\} \tag{3.35}$$

[4] As previously said, we don't use $z$ as an index in the subscript since $(x, y)$ uniquely addresses one point.

There are two important things to note in these definitions. The first is that we estimate the capacity on the drivable (walkable) area only when we optimize on vehicles (pedestrians), so we do not consider the cross-metrics like in Sect. 3.3.3.1 and capacity metrics lose the *ve*/*pe* subscript. The second is that we consider NLoS links only when a LoS link is not available or it offers a lower capacity than a NLoS link. This last observation leads to the definition of a further metric: the fraction of points for which there exist some LoS links, but a NLoS link provides a higher capacity.

$$W^{diff}(\lambda) = \frac{|\{(x,y,z) \mid NC_{x,y,\gamma}(\lambda) > LC_{x,y,\gamma}(\lambda) > 0\}|}{|\Lambda_\gamma|} \tag{3.36}$$

$$D^{diff}(\lambda) = \frac{|\{(x,y,z) \mid NC_{x,y,\pi}(\lambda) > LC_{x,y,\pi}(\lambda) > 0\}|}{|\Lambda_\pi|} \tag{3.37}$$

We analyze the sets of capacity metrics in the following section in terms of averages, cumulative distribution functions, and coefficient of variation to provide a link quality estimation of the coverage strategies.

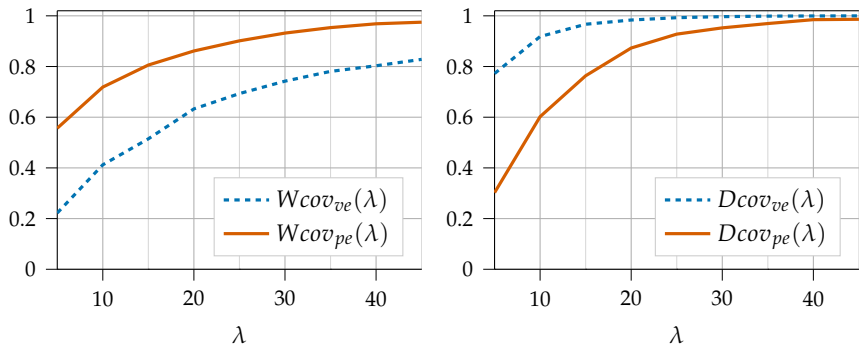
## 3.3.4 Results

### *3.3.4.1 Analysis of the coverage*

We start the analysis by looking at coverage metrics as observed by different areas in the city (walkable and drivable areas) for the two optimization strategies. Fig. 3.17a shows the coverage as perceived on walkable areas when optimizing for pedestrians ($Wcov_{pe}$) and for vehicles ($Wcov_{ve}$) as a function of the density of gNBs. Conversely, Fig. 3.17b shows the coverage as perceived on drivable areas when optimizing for pedestrians ($Dcov_{pe}$) and for vehicles ($Dcov_{ve}$).

Fig. 3.17b shows two very relevant conclusions. The first is that $Dcov_{ve}$ reaches 90 % coverage with $\lambda = 10$, 95 % coverage with $\lambda = 15$ and 99.9 % coverage with $\lambda = 25$ while $Dcov_{pe}$ needs 150 % and 100 % more gNB to cover 90 % and 95 % of the vehicles, respectively, and can not reach 99.9 % even with $\lambda = 45$. Considering that vehicles' coverage for autonomous driving requires high reliability, we see that there is a relevant difference when we specifically optimize for vehicles, rather than for pedestrians. The second conclusion is more generic: so far we did not have any concrete indication of how much we need to increase the density of gNBs to achieve vehicles coverage, and this result tells us that in urban areas, a reasonably low density can still be sufficient for a reliable service.

Fig. 3.17a instead tells a different message. There is a remarkable difference in the coverage of walkable areas when optimizing for vehicles or not. In particular, $Wcov_{ve}$ (the vehicles' optimization) allows us to cover only slightly more than 80 % of the ground. This result is worse than in our previous work [87] where we show that we could not cover more than 95 % of the ground because here we included larger public areas such as




**(a)** Walkable areas **(b)** Drivable areas

**Figure 3.17:** Coverage (fraction of points reached by at least one LoS link) for different optimization strategies. On the left, we measure the coverage for walkable areas ($Wcov_{pe}(\lambda)$ and $Wcov_{ve}(\lambda)$) while on the right we measure the coverage for drivable areas ($Dcov_{ve}(\lambda)$ and $Dcov_{pe}(\lambda)$).

parks and squares in the analysis. On the other side, while optimizing for pedestrians and measuring on walkable areas, $Wcov_{pe}$ reaches 90 % with $\lambda = 25$ and 95 % with $\lambda = 35$.

The takeaway for the operator that needs to start deploying gNBs for LoS communications is that the goals of covering vehicles or pedestrians are concurrent ones. Optimizing for vehicles would reduce significantly the required density of gNBs but would not allow to reliably cover pedestrian areas.

### 3.3.4.2 *Analysis of the capacity*

While coverage gives us a qualitative measure of the impact of different optimization strategies, it does not provide us with quantitative insights. In this section, we analyze the capacity resulting from different optimization strategies, but before that, we introduce Fig. 3.18, which shows the fraction of NLoS links chosen even if a LoS link was available ($W^{diff}(\lambda)$ and $D^{diff}(\lambda)$). Regardless of the density, less than 1 % of the links will obtain better connectivity from a NLoS link than from a LoS link. This means that in realistic coverage conditions, the best link is almost always the LoS one and that NLoS links are of prevalent importance only when the LoS coverage is low, that is, in the leftmost part of the curves in Fig. 3.17. The fact that the curves in Fig. 3.18 grow with $\lambda$ is due to the average reduction of the distance to any gNB with the growth of $\lambda$. As we already noticed this makes it more likely that the two curves in Fig. 3.16 intersect, as the average link length stays in the lower range.

**ESTIMATING THE AVERAGE CAPACITY** For each density value $\lambda$, we compute the average capacity for LoS and NLoS links, as well as the overall average. Fig. 3.19a shows the average capacity measured over the walkable areas when optimizing the locations of gNBs for pedestrians. The

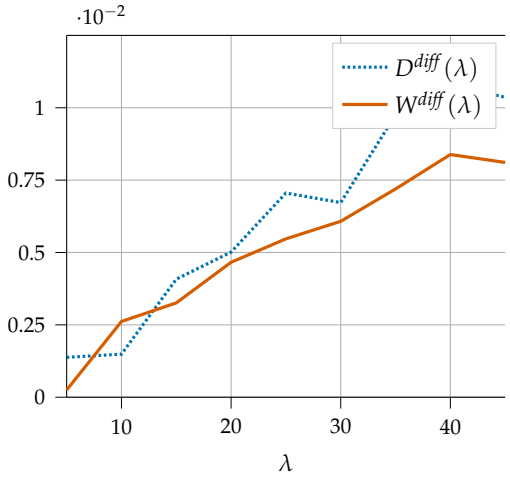



**Figure 3.18:** Fraction of points for which a NLoS gNB has been chosen even though a LoS gNB was available, for drivable areas ($D^{diff}(\lambda)$) and walkable areas ($W^{diff}(\lambda)$).

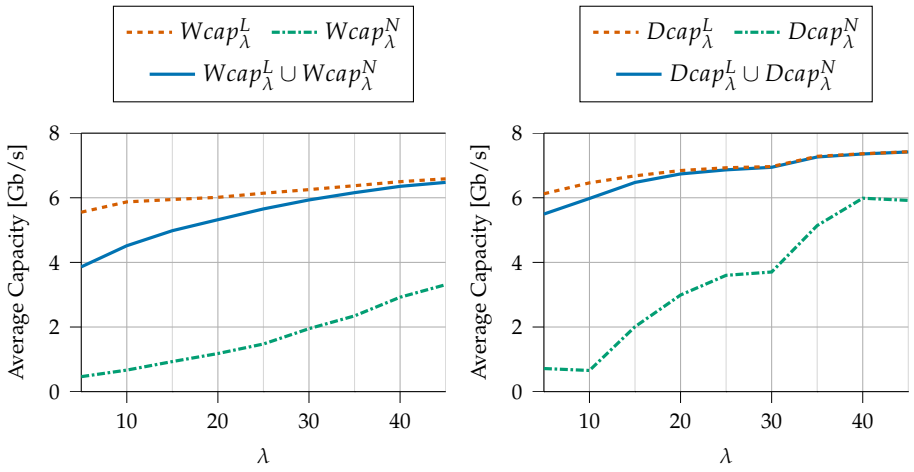
absolute values are to be considered as an upper bound (as in conditions of interference, the capacity can be far from the Shannon limit) however the comparison is interesting, with a minimum average of 5.5 Gbit/s achieved at the lowest gNB density for LoS links, versus less than 500 Mbit $s^{-1}$ for NLoS links. Especially at low gNB densities, this has a large impact on the overall (LoS and NLoS) average capacity: As the fraction of NLoS links for $\lambda = 5$ is 45 % (Fig. 3.17), the overall average is as low as 4 Gbit/s.

Fig. 3.19b instead shows the average capacity measured over drivable areas only when optimizing for vehicles. Again, there is a striking difference between LoS and NLoS links, however, since with the same density the drivable areas reach a better coverage, the overall average is closer to the average of LoS links. This also applies to the right end of the curve, in which a higher density reduces the average length of all links (including NLoS) and thus the green curve gets closer to the orange one.

**CAPACITY DISTRIBUTION** As averages hide information about the distribution of the capacities, Fig. 3.20b show the ECDF of the $Dcap_{ve}$ and $Wcap_{pe}$ sets. Each graph shows the distribution for $\lambda = 5$, the density to cover 95 % of the area ($\lambda = 35$ for pedestrians, $\lambda = 15$ for vehicles), and for $\lambda = 45$, for both LoS and NLoS links.

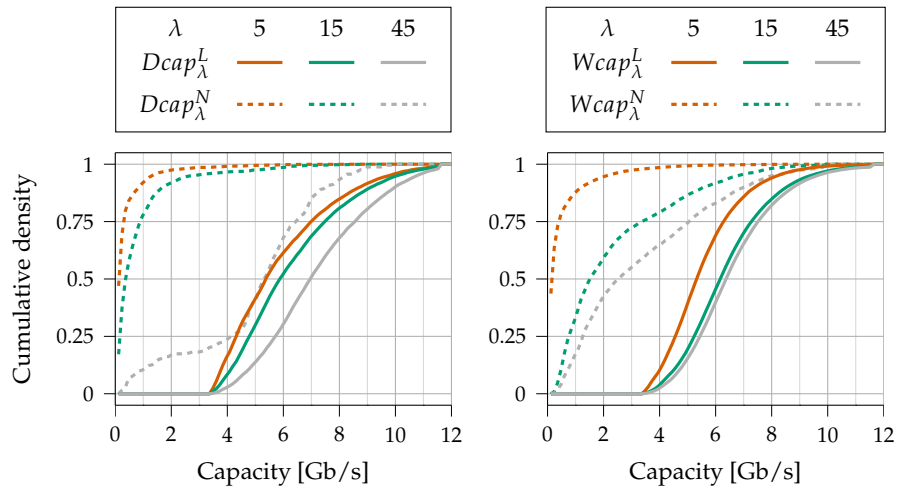
By looking at LoS capacities, regardless of the optimization strategy and the density, for the smallest possible density, we can see that all distributions have a reasonable shape. Most importantly we can see that no capacity is smaller than 3 Gbit/s. This means that, regardless of the objective, deploying gNBs focusing on LoS will result in very good performance.

With respect to NLoS, the quality of links is instead highly dependent on density. At low densities, a vast majority of links experience capacities that are very close to zero. For the densities providing 95 % coverage, instead, the distribution depends on the optimization strategy. When optimizing




**(a)** Walkable areas **(b)** b) Drivable areas

**Figure 3.19:** Average Capacity for different areas and LoS conditions. On the left, we measure the average capacity for walkable areas, divided LoS links ($Wcap^L_\lambda$), NLoS links ($Wcap^N_\lambda$) , and both of them ($Wcap^L_\lambda \cup Wcap^N_\lambda$)). On the right, we measure the same metrics for drivable areas.




**(a)** Drivable areas **(b)** b) Walkable areas

**Figure 3.20:** ECDF of the capacity for different areas, LoS conditions and gNB densities. At the top, the metrics are displayed for drivable areas, while at the bottom for walkable areas. The gNB densities displayed, corresponds to the minimum considered ($\lambda = 5$), the maximum ($\lambda = 45$) and the density that guarantees a LoS coverage of 95 % of the areas ($\lambda = 15$ for drivable areas and $\lambda = 35$ for walkable areas).

for vehicles (Fig. 3.20a), roughly 20 % of the links have close to zero capacity, and 75 % have a capacity lower than 1 Gbit/s. This means that, in high density zones, vehicles will experience very good network conditions, but in less dense areas communication might be at risk and some roads might be left completely without coverage or with poor communication. For the pedestrian strategy, there are fewer links with this problem, as 50 % of them have an available capacity of at least 1.5 Gbit/s, but this comes at the cost of more than doubling the density of gNBs. At high density, the vehicular strategy results in very good performance, with just 20 % of links experiencing a capacity smaller than 3 Gbit/s and 75 % of them experiencing at least 4 Gbit/s. On the contrary, the ECDF for the pedestrian strategy grows at a much slower rate, with almost 50 % of the links having a capacity smaller than 2 Gbit/s.

Finally, Fig. 3.21 shows the coefficient of variation (i.e., the ratio between standard deviation and the average) of the capacity for different values of $\lambda$. Regardless of the optimization strategy, the coefficient of variation decreases with the density of gNBs. This indicates that, as we increase the density of gNBs the change in the experienced capacity between different locations reduces, indicating a more even distribution of resources among users, but also in the trajectory of one single user. Increasing the coverage not only improves the average performance but makes it more stable.

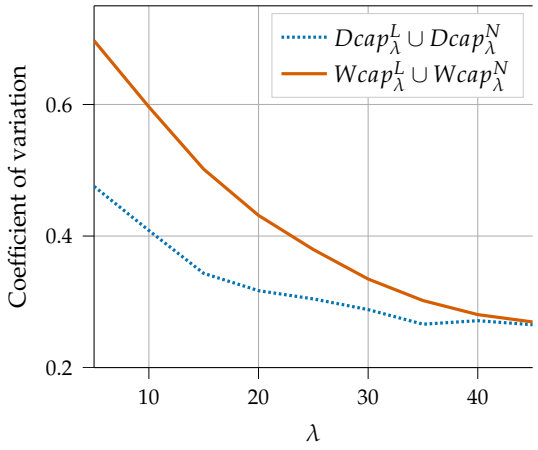



**Figure 3.21:** Coefficient of variation (ratio between standard deviation and average) of the capacity for drivable areas ($Dcap^L_\lambda \cup Dcap^N_\lambda$) and walkable areas ($Wcap^L_\lambda \cup Wcap^N_\lambda$).

## 3.4 FEASIBLE TOPOLOGIES FOR BACKHAUL MESH NETWORKS

In this section, we used similar visibility analysis techniques as in the previous works with a different objective. Here, instead of focusing the analysis on the ground points in LoS with the gNBs, we tried to analyze what kind of network could be made by interconnecting rooftop-placed gNBs using LoS links in different environments: Urban, Suburban, and Rural.

We built visibility graphs for 9 Italian municipalities divided in Urban, Suburban, Rural environments, and we compared them to a full-mesh graph realized on the same set of nodes by focusing on robustness metrics. The motivation came from recent works that dealt with mesh backhaul assuming a full mesh between the network nodes [71], [88] and we wanted to test how this assumption influences the results in a more generic setting.

Results showed how important it is to include in topological models of mesh networks details deriving from topographic characteristics. Once this need is assessed, a plethora of possibilities (and needs) open up for the research community: finding synthetic models that include these characteristics so that realistic studies can be carried out without the need of using specific or particular data; define strategies to select the best positions for mesh nodes given an area; study the impact of the mesh topologies on vertical applications or paradigms like Mobile Edge Computing (MEC) and many others

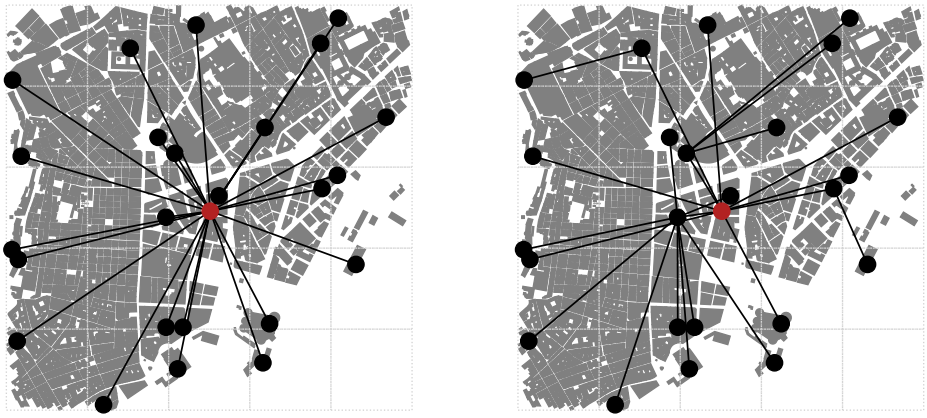
### 3.4.1 Methodology

Given a geographical area, we choose a building in the area which serves as a gateway $n_g$. This building is selected according to considerations that define a good place for a gateway, and we initialize $\mathcal{N} = \{n_g\}$. We select a $1400\,m \times 1400\,m$ area centered on the gateway and we divide it in a $5 \times 5$ regular grid of squares with side 280 m. In every square we choose the tallest building and we place there a node of the graph, completing a set $\mathcal{N}$ with 26 nodes. In some cases there are no buildings in some squares, or the tallest building is not in line-of-sight with any other building, so not all the graphs are made of exactly 26 nodes.

Given $\mathcal{N}$ we compute two different graphs:

- i) the visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ or $G^v$ for brevity, as per Sect. 2.3.5.4;
- ii) the full mesh graph $G^f$ obtained connecting every pair of nodes in $\mathcal{N}$.

Fig. 3.22 reports the spanning tree rooted in the gateway computed with Dijkstra algorithm when the cost of links is proportional to their length for an area in Urban 3 scenario (around 'Piazza del Municipio'); the thin dotted lines define the regular grid and one of the squares is in the gulf of Napoli and has no building. The left hand side refers to $G^f$, and



(a) FullMesh (b) `TrueNets`

**Figure 3.22:** $G^f$ vs. $G^v$: minimum spanning tree to reach $n_g$ in a Urban 3 area with full visibility or `TrueNets` model. © 2021 IEEE

it is obviously a star. The right hand side refers to $G^v$ and the outcome is clearly very different. The goal of the figure is qualitative: to show how different the topology is when it is built considering effective topographic data from the one built on a simple visibility model, heuristically intuitive given the nodes are on the tallest buildings of each area.

The goal of this initial work is to study how $G^v$ influences the feasible network topology given a set of nodes, and specifically, how different robustness metrics of networks built on $G^v$ and $G^f$ are.

### 3.4.2 Robustness and Performance Metrics

#### *3.4.2.1 Effective Graph Resistance*

The first metric we consider is related to the overall potential capacity of the network and its robustness, because it takes into account the presence of parallel (possibly disjoints) paths. The metric is called *Effective Graph Resistance* [89], and has been used and defined in the previous section. As in Eq. (3.43), let it be defined as $\xi$.

We use this metric to give a high level estimate of the quality of a mesh build on a specific visibility graph $G^v$. $\xi$ strictly decreases when an edge is added to the network, so the smaller $\xi$ the more robust is the mesh, but also the overall capacity can be larger, as it can exploit more disjoint paths between nodes. To guarantee an easy interpretation of results independently from the number of nodes of the graph, we use the ratio between the effective resistance of $G^f$ and that of $G^v$: $\xi_R = \frac{\xi_{G^f}}{\xi_{G^v}}$. Since the effective resistance of a full mesh is the minimal possible one, $\xi_R \leq 1$ for any visibility graph $G^v$, and the smaller it is the worst are the properties of meshes built on $G^v$.

#### 3.4.2.2 k-edge Connectivity

Next, we consider $r[k]$, the fraction of nodes that are part of a k-edge connected graph embedded on a graph $G$ and including $n_g$. Algorithm 3.3 describes how we compute $r[k]$.

**Algorithm 3.3** The spanning-tree robustness metric

```
Require: G(N_v, E_v), n_g, k_conn_max
procedure (r (Number of nodes in the k-edges-connected subgraphs))
    k = 1
    r =dict()                                   ▷ a mapping int–>object
    for iin1 . . . k_conn_max do
        r[i] = []                               ▷ initialize to an empty list
    end for
    newG = G
    while k <= k_conn_max do
        S(N_s, E_s) = compute_min_spanning_tree(newG)
        for e in E_s do
            newG.remove_edge(e)
        end for
        newG = connected_component(newG, n_g)
        sizeG = |newG|                          ▷ // number of nodes in newG
        if sizeG > 1 then
            r[k].append(sizeG)
        end if
    end while
    k += 1
    return r
end procedure
```

The algorithm starts from complete graph $G$ and creates a minimum spanning tree $St$, then removes from $G$ all the edges of $St$ and checks the size $sizeG$ of the largest connected component including $n_g$. It iterates this process as long as $n_g$ does not get disconnected from the rest of the graph ($sizeG = 1$). At each $k$-th iteration we save $r[k] = sizeG$. A simple interpretation of $r[k]$ is that when a node $n$ is k-edge-connected with the gateway, then there are at least $k$ independent spanning trees that connect $n$ to $n_g$. This means that $n$ can survive any pattern of $k - 1$ failures, as in the worst case scenario one edge fails on a different $St$, but still there is at least one $St$ on which $n$ can communicate with $n_g$. The $r[k]$ metric tells what is the number of nodes of the graph that are k-edge-connected. Note that the spanning tree are generated with the classical Kruskal algorithm, and not centered on $n_g$ as in Fig. 3.22, otherwise only one spanning tree would be possible around the gateway in a full mesh, as all its outgoing edges are removed after the first loop of Algorithm 3.3.

Given a municipality, we select 5 gateways and areas $A$ around them with different sets of nodes $S_A = \{\mathcal{N}_0 \dots \mathcal{N}_4\}$. Then we compute 5 graphs $G_i^v$ as a visibility graph, and for each one we obtain $r_i[k]$.

#### 3.4.2.3 *Weather Disruption-Tolerant Networks*

The third metric is a state of the art algorithm for topology design [88]. This work tackles the problem of designing a robust backhaul topology using a realistic model for the probability of link availability under heavy rain or snow. It proposes two optimization algorithms. The first one, which assumes uncorrelated link failures, is formulated as an integer linear program and it is more scalable and easier to reproduce, the second introduces correlation among link failures and is formulated as a quadratic linear program. We use the first one, as it is more scalable and useful to highlight our findings. This model, called in the original work TD_IF, takes as input all the paths from any $n \in \mathcal{N}$ to $n_g$ and selects the least number of disjoint paths so that the probability that $n$ is disconnected is below a certain threshold $\epsilon$. As shown in algorithm 3.4, TD_IF returns a set of optimal paths for every $n$, which are joined into a single topology. We introduce two modifications on the original algorithm to make it run on our graphs, first, we limit the set of paths to the ones with less hops than $diam(G) + 2$, where $diam(G)$ is the diameter of the graph G; second we relax the robustness requirement otherwise some of the `TrueNets` graphs do not support a solution: reliability $\epsilon$ is achieved for at least 90% of the nodes in the graph. The metric is the cost of the graph, which is proportional to the sum of the length of the links in all the paths. This is, like $\xi_R$, a synthetic number whose absolute value has no straightforward interpretation, but that can be used to compare two strategies.

**Algorithm 3.4** Modified TD_IF algorithm.

**Require:** $G(\mathcal{N}_v, \mathcal{E}_v)$, $n_g$

**procedure** DO SOMETHING($G_t$)
  $\xi_g = \emptyset$ ▷ empty set of edges
  **for** $n$ in $\mathcal{N}$ **do**
    $P_{all}$ = calc_simple_paths($G$, $n$, $n_g$, $diam(G)$ + 2)
    $P_{best}$ = TD_IF($P_{all}$, $\epsilon$)
    **for** $e$ in $P_{best}$ **do**
      $\xi_g = \xi_g \cup e$
    **end for**
  **end forreturn** $G_t(\mathcal{N}, \xi_g)$
**end procedure**

### 3.4.3 Results

We can now analyze the nine different municipalities in Italy, verify the impact of the `TrueNets` modeling and check if there are regularities for Urban, Suburban, and Rural areas.

Fig. 3.23 reports the ratio $\xi_R$ between the effective resistance of $G^f$ and $G^v$ for all the municipalities under analysis. It is clear that $G^v$ differ significantly from a full mesh in all cases, even in densely populated areas as the Urban municipalities in Italy. There is no specific "trend" related to the density of building, while what seems to have more impact is the landscape itself (Suburban 1 is mostly flat, while Rural 1 is in the mountains) and possibly the presence of buildings that are indeed much higher than others.

Fig. 3.24 shows the value of the robustness $r[k]$ defined in Sect. 3.4.2.2. Recall that $r[k] = |\mathcal{N}|$ means that there are $k$ completely independent spanning trees that cover the entire network, while decreasing values means that some nodes are not k-edge-connected. It's clear that building a mesh on a graph that allows a large number of independent spanning trees makes it very robust and resilient, while if removing the links of the first spanning tree leaves some node disconnected the mesh is clearly fragile. Orange lines refer to FullMesh, while blue ones to `TrueNets` and we report results for the 5 different networks we generated in each area; shaded areas are the envelope of the five curves. The difference is evident with $G^v$ that rarely allows the presence of more than two independent spanning trees without leaving some node disconnected; in some cases only one exists.

Fig. 3.25 finally describes the cost of the topologies generated with the TD_IF algorithm, again on 5 graphs per strategy. The $x$ axis is the probability $\epsilon$ of letting a node disconnected, the $y$ axis reports the cost of the corresponding graph generated using the TD_IF algorithm. We report only 6 out of 9 municipalities. In one area (Suburban 1) `TrueNets` generates topologies that are close to a FullMesh, and thus, the cost of the

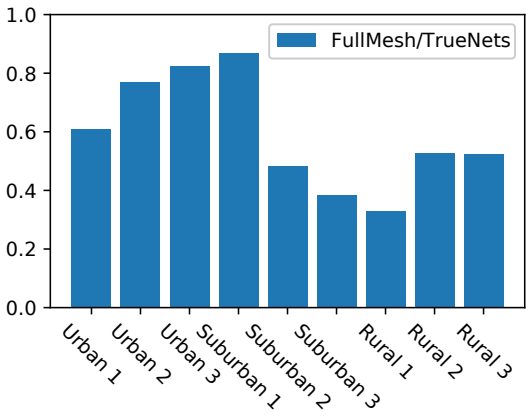



**Figure 3.23:** The ratio $\xi_R$ between the effective resistance of $G^f$ and $G^v$. © 2021 IEEE

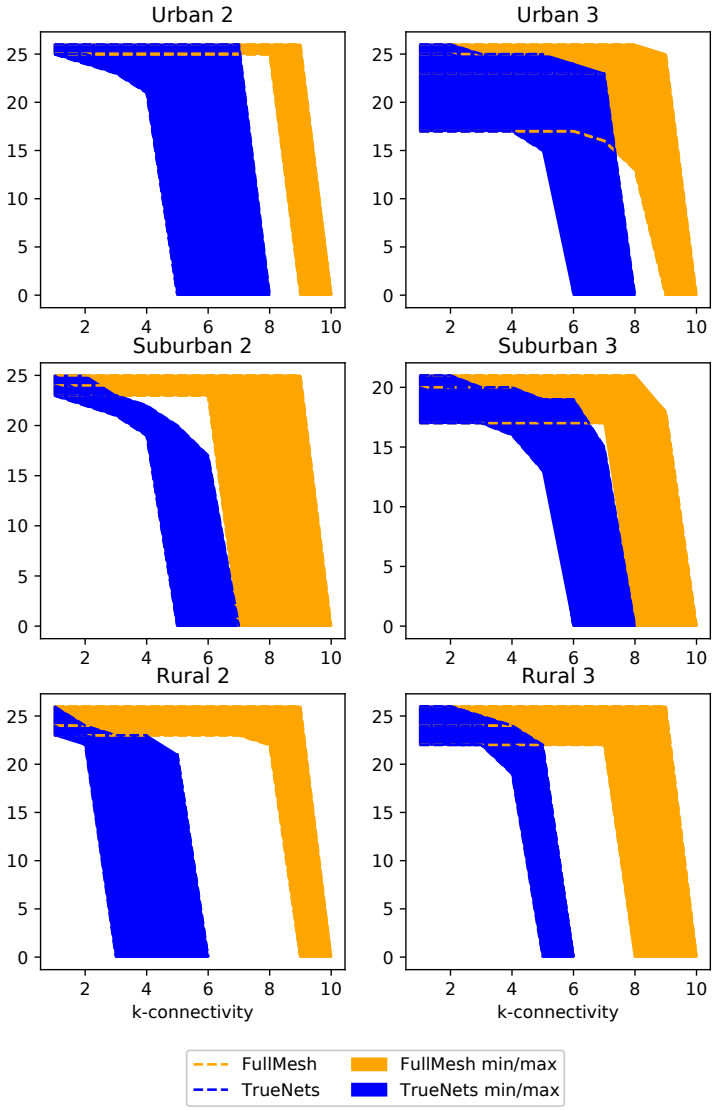



**Figure 3.24:** The value of $r[k]$ computed on 5 instances of $G^v$ and $G^f$ in 6 out of 9 areas. © 2021 IEEE

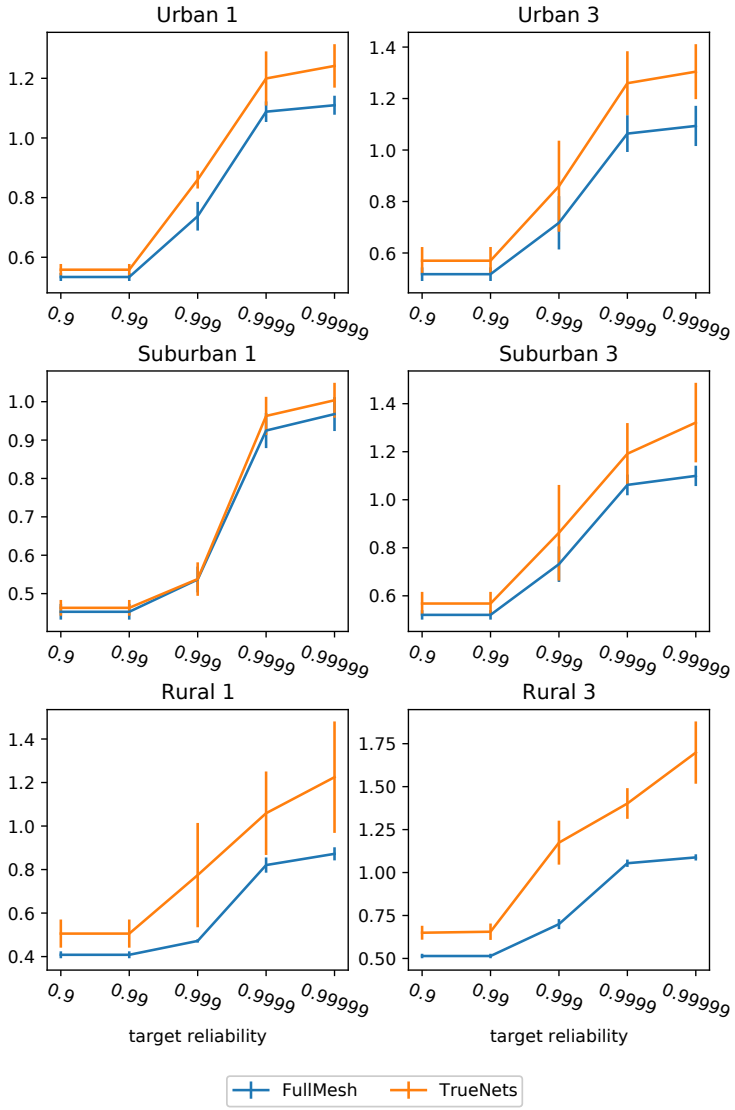



**Figure 3.25:** Cost of the network produced by the TD_IF algorithm for on 5 instances of $G^v$ and $G^f$ in 6 out of 9 areas. © 2021 IEEE

network is similar. This is consistent with Fig. 3.23 in which Suburban 1 has the closest score to the FullMesh. In all the other areas the cost is strongly affected by the underlying topology, especially with higher levels of robustness, with an more visible effect in rural areas.

## 3.5 WIRELESS BACKHAULS FOR URBAN MESH NETWORKS

The last contribution of this chapter extended the work-in-progress research, presented in the previous section, by evaluating the visibility graphs and their applications on larger areas, up to 100km$^2$ and 50000 buildings.

We did so by first developing an heuristic to find the most suited location to deploy a gNB on a rooftop and then by building a visibility graph between all the couples of buildings in that area. Using the same set of locations, we then built a more simplistic version of visibility graphs by leveraging on widely used LoS probability models.

We compared the different graphs in terms of their topological properties and then we used two state-of-the-art topology design algorithms to embed an IAB topology into those graphs.

The analysis was carried out using Open-data coming from 9 different Italian municipalities and led us to understand that the use of synthetic models, such as the ones proposed by ITU, lead to wireless backhauls whose property and behavior is far from the realistic one and thus, from reality.

### 3.5.1 Visibility analysis: generating the visibility graph

We first define some fundamental concepts of visibility analysis, which we then apply to generate $G(\mathcal{N}_v, \mathcal{E}_v)$.

Given a DEM $\mathcal{D}$ and it's associated matrix $\mathbf{E}$, let $\Upsilon(\mathbf{E}, p_i, p_j)$ be a function that returns 1 if there is direct LoS between $p_i$ and $p_j$ and 0 otherwise, as defined in Eq. (2.24).

Evaluating $\Upsilon(\mathbf{E}, p_i, p_j)$ is a computationally intensive task, but recent advances in GPU optimized algorithms [90] allow performing this process in areas that contain tens of thousands of buildings, with billions of potential links to be tested. The design of this process, together with the software provided to the community, is one of the contributions of this paper.

We start from the DEM of 9 real-world areas, corresponding to 9 administrative municipalities in Italy. The areas belong to urban, suburban, and rural areas[5], whose properties are reported in Tab. 3.6. To process the data we use the Numba libraries that exploit the CUDA architecture for NVIDIA GPUs[6].

Our goal is to study the visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ where each vertex of the graph corresponds to one point on the roof of an existing building, so we need the shapes of the buildings in the specified areas. This information

[5]The urban/suburban/rural label was obtained using the Eurostat definitions, see https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Territorial_typologies_manual_-_degree_of_urbanisation

[6]See the Numba documentation at https://numba.pydata.org/numba-doc/latest/cuda/index.html.

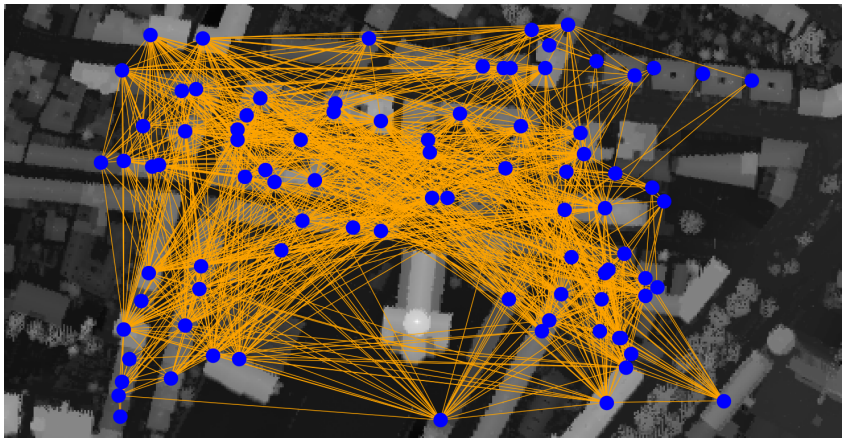


**Figure 3.26:** Visibility graph of the sample area of Fig. 3.27 (for clarity, only a subset of the points are reported). © 2023 IEEE

| | Name | Extension ($km^2$) | Pop. dens. (ab./$km^2$) | Buildings | Build. dens. (buil./$km^2$) |
|---|---|---|---|---|---|
| U1 | Trento | 187.88 | 749.24 | 16490 | 104.45 |
| U2 | Firenze | 102.32 | 3702.54 | 44598 | 435.87 |
| U3 | Napoli | 119.02 | 8058.82 | 50879 | 423.90 |
| S1 | Mezzolombardo | 13.88 | 517.10 | 1285 | 92.01 |
| S2 | Barberino di Mugello | 133.29 | 82.05 | 2338 | 17.54 |
| S3 | Sorrento | 9.95 | 1639.25 | 3856 | 387.32 |
| R1 | Predaia | 80.05 | 83.39 | 2333 | 29.14 |
| R2 | Pontremoli | 182.48 | 39.36 | 4079 | 22.35 |
| R3 | Visciano | 10.90 | 401.60 | 1266 | 116.11 |

**Table 3.6:** Main urbanization parameters for the 9 areas

is obtained using two different sources: OSM and the technical regional charts data made available by the single municipalities. More details on this are available in Sect. 2.3.2.

The latter source is more accurate but not necessarily updated, while the first source is more frequently updated especially in urban areas, but may miss some buildings. For each area, we use the source that provides the highest density of buildings per km$^2$ without attempting to merge the data, a process that would be too error-prone. Therefore, given a closed polygon $s_k$ that represents building $b_k$ we implement a function that provides a binary matrix $\mathbf{S}^k$ of the same dimension of $\mathbf{E}$ so that:

$$\mathbf{S}^k_{x,y} = 1 \iff (x,y) \in A(s_k) \tag{3.38}$$

where $A()$ returns the area limited by $s_k$.

In the following, we detail the steps needed to create the annotated visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ once we have $\mathbf{S}^k$ for all the buildings in a certain area.

#### 3.5.1.1 Roof Node Placement

The first step for the computation of $G(\mathcal{N}_v, \mathcal{E}_v)$ is to determine $\mathcal{N}_v$, which requires to attribute one point $p_i$ to each building $b_k$, so $|\mathcal{N}_v|$ equals the number of buildings in the area. The precise position is important because it influences the chances of having LoS with the other chosen points. In the real world this is done inspecting the roof and visually searching for the place with the best visibility towards other buildings. In an automated procedure, we need an algorithm that tries to maximize the probability of LoS towards other buildings.

For every $s_k$ we are looking for a point $p_i$ with coordinates $(x,y) \in A(s_k)$ so that the resulting graph $G(\mathcal{N}_v, \mathcal{E}_v)$ has the highest number of edges. An exact algorithm needs to explore all the possible combinations of all the $(x,y) \in s_k$ for every $k \in [1, \ldots |\mathcal{N}_v|]$, whose number grows exponentially with $|\mathcal{N}_v|$. By modeling the problem as a colored graph, where each point $k$ on the roof is colored with the same color, we can reduce it as the search for the connected rainbow subgraph, which has been proven to be NP-Hard [91]. For this reason we need a heuristic.

A reasonable assumption is that points that have a good visibility toward the centroids of buildings will also have good visibility toward the good visibility points on the buildings, so that, starting from centroids, we can select these points with an algorithm that scales quadratically with the number of buildings. However, centroids may actually fall outside the building roof (e.g., internal courts), so we use the point $c_k$, which is the point of coordinates $(x_k, y_k, z_k)$ where $(x_k, y_k)$ are the coordinates of the a suitable point that ensures $(x_k, y_k) \in A(s_k)$ as defined by the C++ function `GEOSPointOnSurface` from the `GEOS` library. For the sake of readability we will call these points pseudocentroids.

We define the pseudocentroid $c_k = (x_k, y_k, z_k{=}\mathbf{E}_{x,y}{+}2)$ adding 2 m to $z_k$ since we assume gNB antennas are elevated on the roof with a pole[7] and we compute all the viewsheds $\sigma^k$ from any point $c_k$ to any other point also elevated by 2 m. Since Y is a symmetric function, given a point $(x, y, \mathbf{E}_{x,y} + 2)$, each element $\sigma^k_{x,y}$ represents the availability of LoS from $(x, y, \mathbf{E}_{x,y} + 2)$ to $c_k$. Summing all $\sigma^k$ we obtain $\mathbf{V} = \sum_{k=1}^{|\mathcal{N}_v|} \sigma^k$, which is a cumulative visibility matrix whose values range from 0 to $|\mathcal{N}_v|$ indicating how many centroids $c_k$ are in LoS from any elevated point on every building. The second step is searching, for every building $b_k$, the point $\bar{p}_i$ that has the highest visibility of centroids. This is obtained masking $\mathbf{V}$ with $\mathbf{S}^k$: $\mathbf{V} * \mathbf{S}^k$ (where $*$ is the element-wise multiplication) selecting all the values of $\mathbf{V}$ belonging to points inside $b_k$, and returning the coordinates of the maximum:

$$(\bar{x}, \bar{y}) = arg\,max_{(x,y)}(\mathbf{V} * \mathbf{S}^k) \tag{3.39}$$

then $\bar{p}_i = (\bar{x}, \bar{y}, \bar{z}_k = \mathbf{E}_{\bar{x},\bar{y}} + 2)$ is going to be the position of the gNB antennas for building $b_k$. We repeat this process for every building and we obtain $\mathcal{N}_v$ as the collection of all the $\bar{p}_i$. From now on, when we mention a generic point $p_i$, we always refer to points chosen with this procedure.[8]

Fig. 3.27 shows a sample of the visibility matrix $\mathbf{V}$ as a greyscale image in which every pixel is a point in $\mathbf{V}$, with lighter color meaning higher visibility index; the chosen $\bar{p}_i \in \mathcal{N}_v$ are reported in yellow.

#### 3.5.1.2 Building the visibility graph

Once the set $\mathcal{N}_v$ has been determined we need to build the set $\mathcal{E}_v$ computing $\mathrm{Y}(p_i, p_j)$ for each couple of points $p_i, p_j \in \mathcal{N}_v$, tracing a ray between $p_i$ and $p_j$ and checking whether any obstacle intersects the ray.

As in the previous task, implementing this algorithm in a GPU allows parallelizing the task among the large number of cores available. As described previously in Sect. 2.3.5.2, each core is assigned to one point $p_i \in \mathcal{N}_v$ and calculates the LoS towards any other point $p_j \neq p_i \in \mathcal{N}_v$ using well known visibility algorithms [47]. Our specific implementation is detailed in Algorithm 2.2.

Fig. 3.26 shows the visibility graph of the sample area. The properties of the resulting graphs are reported in Tab. 3.7 and commented in Sect. 3.5.3.

#### 3.5.1.3 Complexity analysis

The worst-case time complexity for building $G(\mathcal{N}_v, \mathcal{E}_v)$ is $O(|\mathcal{N}_v|^2) \cdot C(\mathrm{Y})$, where the first part represents the iteration on every pair of nodes and the second part represents the complexity of the LoS computation between two given points. $C(\mathrm{Y})$ depends on the longest LoS link for which the

[7] This is an arbitrary choice that can be modified according to the design constraints and requirements.

[8] Note that this introduces a little abuse of notation. So far we used $i$ as the index of point $p_i$ in the ordering $o(x, y)$ for the coordinates in the DEM, and $k$ as an index for buildings, so the two indexes are uncorrelated. From now on we will simply refer to $p_i$ as the point chosen for building $i$.

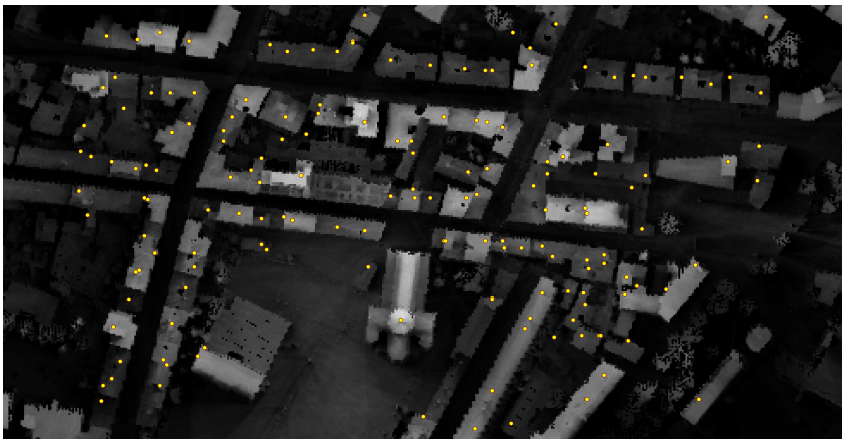


**Figure 3.27:** Outcome of the selection of the best visibility point –in yellow– for each building in a small portion of an urban area, the shade of grey indicates the visibility index with light grey indicating higher visibility. © 2023 IEEE

algorithm needs to evaluate every space element intersected by the ray. Assuming that **E** is a square matrix, the longest link would be its diagonal which is $\sqrt{2|E|}$, which leads to $C(Y) = O(\sqrt{|E|})$, thus a general complexity equals to $O(|\mathcal{N}_v|^2 \cdot \sqrt{|E|})$. In order to express the complexity solely on the size of the area we can take advantage of the fact that the number of buildings grows linearly with the area. This leads to an overall complexity of $O(|E| \cdot \sqrt{|E|})$

On the other hand, the SoA algorithm for viewshed analysis [47] has a complexity of $O(|E|^2)$ to calculate a single viewshed. In order to calculate the visibility graph the complexity would be $O(|\mathcal{N}_v| \cdot |E|^2)$, which again expressed only in terms of the area is $O(|E|^2 \cdot \sqrt{|E|})$.

Computing such an algorithm on datasets like U3 (more than 50 k nodes and 100 km$^2$) would not be feasible using a normal CPU, however modern GPUs, with their high number of cores and large RAM allow to speed up the process and compute the whole visibility graph in a reasonable time. We used an NVIDIA Tesla P100 GPU which has 3584 cores and 16 GB of memory. This allows the computation of the whole process for the largest city in the data-set in roughly one hour at a speed of 40 M-links per second.

### 3.5.2 Generation of the topology

Given $G(\mathcal{N}_v, \mathcal{E}_v)$ and the position of a set $\mathcal{N}_\varphi \subset \mathcal{N}_v$ of gNBs, $G_\varphi$ is embedded into $G(\mathcal{N}_v, \mathcal{E}_v)$ choosing a set $\mathcal{E}_\varphi \subset \mathcal{E}_v$ that will create the wireless backhaul. Our goal is to show that the realistic data-set of $G(\mathcal{N}_v, \mathcal{E}_v)$ topologies that we publish is key to obtain realistic results in the design of a next generation wireless backhaul $G_\varphi$. Contrarily to classical large-scale, low density mesh networks that span across multiple municipalities [92] a Next Generation Wireless Backhaul (NGWB) is expected to extend a wired backhaul in localized regions, so in this section we describe three

processes: i) how we create realistic localized $G(\mathcal{N}_v, \mathcal{E}_v)$ and $\mathcal{N}_\varphi$; ii) how we create simplistic localized $G(\mathcal{N}_v, \mathcal{E}_v)$ and $\mathcal{N}_\varphi$ for comparison; iii) two strategies from the state of the art to choose $\mathcal{E}_\varphi$, whose performance will strongly vary when applied to realistic or simplistic data.

#### 3.5.2.1 A Realistic Localized visibility graph

In each of the 9 areas we select 5 sub-areas of approximately $1 \times 1$ km, and in each sub-area we call $\Gamma$ the border of the convex hull that includes all the buildings fully contained in the sub-area, with $A(\Gamma) \leq 1\,\text{km}^2$ its area. We choose $\mathcal{N}_v$ assigning a point $p_i$ to each building in the area with the procedure described in Sect. 3.5.1, and finally we produce the localized realistic $G(\mathcal{N}_v, \mathcal{E}_v)$ applying Algorithm 2.2 to these points. Given a desired density of gNB per km$^2$ ($\rho \in \{30, 60\}$), we set the size $|\mathcal{N}_\varphi| = \rho A(\Gamma)$ and we pick a random set $\mathcal{N}_\varphi \subset \mathcal{N}_v$. In the figures we simply refer to data generated with this process with the TrueNets label.

#### 3.5.2.2 Simplistic Localized visibility graph

Given the same 9 areas and $|\mathcal{N}_\varphi|$ as defined above, we generate two versions of a simplistic $G(\mathcal{N}_v, \mathcal{E}_v)$. As a baseline we pick $\mathcal{N}_v = \mathcal{N}_\varphi$ using a Homogeneous Poisson Point Process (which we refer to as HPPP) in which locations are chosen with a random uniform choice in $A(\Gamma)$, without any relation to the building maps. In order to create $\mathcal{E}_v$, for each couple of nodes in $\mathcal{N}_v$, we add an edge with a probability given by the ETSI model using Eq. (2.8). This simple strategy is the one used in most of the papers that propose approaches for the creation of wireless backhauls, such as the ones from Polese et al. [75], [76] mentioned later on.

In a second, slightly more realistic approach, we pick $\mathcal{N}_\varphi$ as in the realistic case of Sect. 3.5.2.1, but we use the ETSI model to generate the edges $\mathcal{E}_v$. This is an intermediate model in which the 2D distribution of the points is not completely uncorrelated from the city map, but is similar to the 2D distribution of the buildings. Yet, without 3D information, the edges are chosen with a synthetic model and not with the Y function. As we use the OpenStreetMap data, we refer to this process as OSM.

#### 3.5.2.3 Choosing the edges

Both the described processes produce a visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ and a set $\mathcal{N}_\varphi$ of gNBs. In the IAB terminology, some of the nodes are *donors*, i.e., they are connected to the wired backbone, and the other nodes need to build a multi-hop path to some donor. We randomly choose $\lceil 0.1\rho A(\Gamma) \rceil$ donors (at least 10% of the gNBs), as in [75]. For each donor a Directed Acyclic Graph (DAG) is created that interconnects the reachable nodes, and the union of all the DAGs provides $G_\varphi$. We mentioned in Sect. 2.1.1.3 that there are several proposals to choose the DAGs, and thus create the backhaul graphs, among which we pick two. The first is one of the heuristic proposed by Polese et al. [75] that assumes no centralized coordination.

We report the results for the algorithm named DPS_WF, in which each node tries to connect with a multi-hop path to the physically closest donor. The algorithm is distributed and greedy, thus not optimal. The second one is an optimal centralized strategy for the creation of the backhaul with the smallest distance (in hops) from each node to its donor. The strategy chooses $\mathcal{E}_\varphi$ as the union of all the edges that are in the shortest path from any node to the closest donor, computed with classical Dijkstra's algorithm. It is optimal in the sense that it minimizes the distance between each node and its donor.

Both algorithms are taken from literature and we use them to test the impact of $G(\mathcal{N}_v, \mathcal{E}_v)$ on the properties of $G_\varphi$.

### 3.5.3 Analysis of the visibility graph

This section presents the features of the visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ generated considering the whole areas and the 5 sub-areas of 1 km$^2$. Due to space constraints in the rest of the paper we include and comment only a small set of the figures we generated, that are enough to robustly support our conclusions. The rest of the figures can be found in the supplementary material, together with the links to the data sets and the source code.

#### *3.5.3.1 Size of the giant component*

Before we provide the results on the analysis of $G(\mathcal{N}_v, \mathcal{E}_v)$, we want to highlight the importance of the algorithm chosen to select $\bar{p}_i$. We compute $G(\mathcal{N}_v, \mathcal{E}_v)$ with three different strategies: *i)* using the heuristic described in Sect. 3.5.1.1; *ii)* using the pseudocentroid $c_k$; and *iii)* using the highest point on the roof. Fig. 3.28 shows the relative difference in terms of number of edges between our heuristic and the other two. For instance, in case of comparison between the heuristic (h) and the pseudocentroid (c) placement the metric is $\frac{|\mathcal{E}_v^h| - |\mathcal{E}_v^c|}{|\mathcal{E}_v^h|}$.

It is clear from the figure that the improvement with respect to the pseudocentroids is substantial. In fact in areas such as U3 the number of edges doubles. The comparison with the highest point of the roof still shows a relevant gain, up to 10% in U3, with one exception: R1, which is a mountain rural area composed of isolated hamlets at very different heights, where overall visibility is more influenced by the position of the hamlets than the characteristics of the buildings. Note that, in absolute terms in U3 we gain 7,696,210 edges while in R1 we lose 13,187 edges, so overall the advantage is considerable. Moreover, the highest point in the roof could be hardly accessible (e.g., a chimney). This confirms the importance of a solid and repeatable methodology to produce the visibility graphs like the one we provide. It also suggests that in the real world small differences in node positions reflect in large differences in the graph properties, i.e., the network density in the real world is extremely sensitive to small differences in nodes placement. This makes it an interesting challenge to define generic synthetic models able to capture

| | Name | Giant Component | Links | Avg. Degree |
|---|---|---|---|---|
| U1 | Trento | 16348 (99.1%) | 21531378 | 1317 |
| U2 | Firenze | 44221 (99.2%) | 36981097 | 836 |
| U3 | Napoli | 50552 (99.3%) | 64141794 | 1268 |
| S1 | Mezzo-lombardo | 1276 (99.2%) | 327579 | 256 |
| S2 | Barberino di Mugello | 2288 (97.9%) | 474618 | 207 |
| S3 | Sorrento | 3828 (99.2%) | 2041968 | 533 |
| R1 | Predaia | 1954 (83.7%) | 834965 | 427 |
| R2 | Pontremoli | 3869 (94.8%) | 621069 | 160 |
| R3 | Visciano | 1161 (91.7%) | 169083 | 145 |

**Table 3.7:** $G_\varphi$ Network metrics for the 9 areas.

that variability. From now on we only consider the points $p_i$ selected with the heuristic.

Tab. 3.7 reports the number of nodes that can not be connected to the giant network component. It is always less than 2.1% for the urban and suburban areas and below 10% in the rural areas, with the exception of Rural-1 where, due to the morphology of the area, roughly 17% of the nodes are not in the giant component. However, in the process of network construction that we followed we did not devote any effort to have full connectivity while in a real setting several ad-hoc solutions can be introduced, e.g., higher trellises on roofs, repeater nodes in strategic locations even in the absence of a building – recall that Rural-1 is in a mountain area. So the first key finding is that a NGWB mesh network covering almost entire cities or vast suburban/rural areas is possible in 8 out of 9 settings, without any specific attempt to maximize coverage. This is itself an extremely interesting result, as it confirms that the concept of IAB is feasible in practice and encourages several applications that rely on

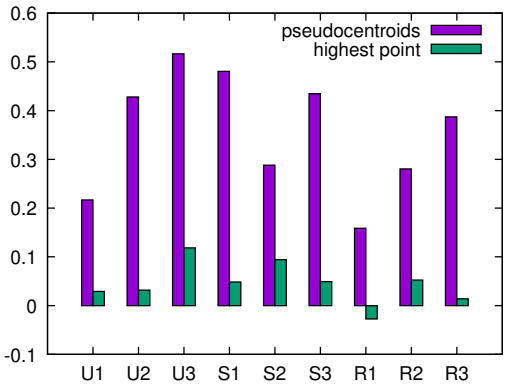



**Figure 3.28:** Relative gain on the number of edges between our heuristic and other point selection strategies. © 2023 IEEE

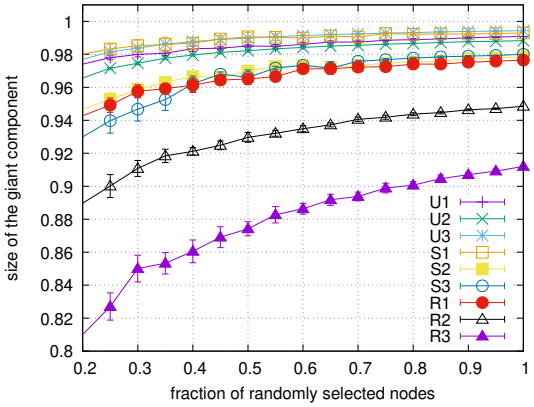



**Figure 3.29:** Size of the giant component with a randomly sampled subset of the nodes. © 2023 IEEE

LoS links, such as backup networks, community networks, the extension of wired access where it can not be provided, or even dedicated networks made of Free Space Optical links for quantum key distribution [93]. An additional result that provides further insights is shown in Fig. 3.29, where the size of the giant component is evaluated in the case when only a subset of the buildings, randomly selected, are used. The figure shows that even with a small percentage of randomly selected buildings it is still possible to build a backhaul connecting most of the nodes. In fact with just 30% of the buildings it is possible to connect more than 95% of the nodes in 7 out of 9 areas, and in all cases more than 80% of the selected nodes are connected. Note that this positive result should be considered as a lower bound, as operators can easily improve it choosing high buildings or those in strategic positions, instead of choosing at random.

#### 3.5.3.2 Coverage, Degree Distribution, Link Length, LoS Probability

The topological properties of $G(\mathcal{N}_v, \mathcal{E}_v)$ do not show clear regularities among different areas or even inside the same area. Fig. 3.30 reports the degree distribution of all rural and urban areas, and shows that while the distributions for Urban-2 and Urban-3 seem to suggest a power law trend, the same trend is less distinguishable for Urban-1 (also due to a more compact distribution). Rural areas have a noisy behavior indicating irregular degrees, also due to the smaller number of nodes in the area, and Rural-3 has a second mode close to the maximum degree. The average degree is high, as it ranges from 145 to 1317. In the network with the lowest (highest) edge density, the average degree corresponds to 1.8% (20%) of the number of nodes.

The physical length of the links of $G(\mathcal{N}_v, \mathcal{E}_v)$ displays large differences even among areas of the same kind. Fig. 3.31 reports the cumulative density function of the link length for all the areas, and it shows that the curves for the three areas of the same kind are always distinct and they start to diverge very early in the graph.

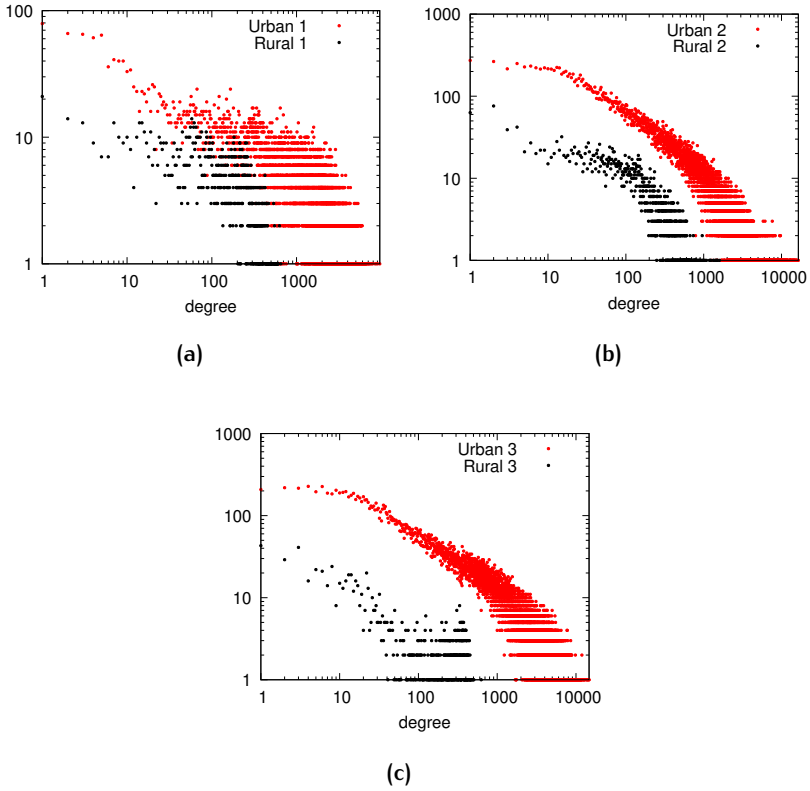
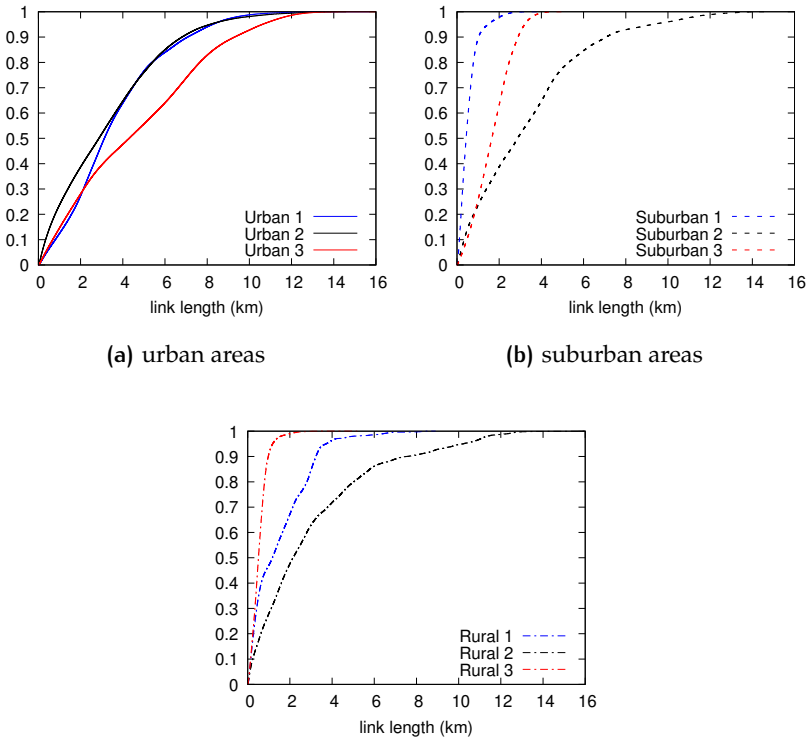



**Figure 3.30**: Degree distribution for every location $p_i$ in Urban and Rural areas.




**(a)** urban areas

**(b)** suburban areas

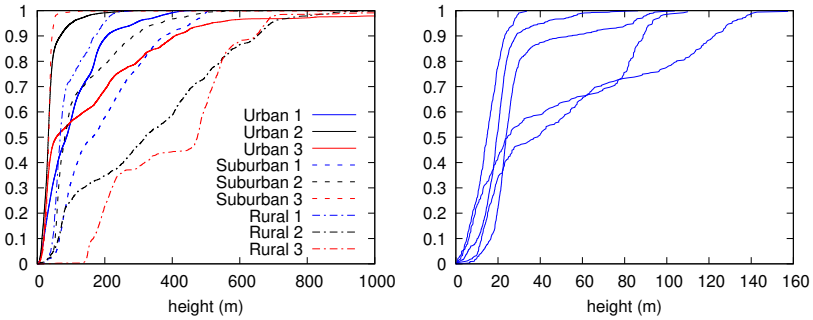
**(c)** rural areas

**Figure 3.31:** EPDF of links length for all the 9 areas. © 2023 IEEE

These observations have two key direct implications for the performance of NGWBs, the first is that a high average degree implies a huge number of $G_\varphi$ networks that can be embedded on $G(\mathcal{N}_v, \mathcal{E}_v)$. This is very important because it allows to divide the physical network into multiple virtual backbones to support different applications. This can be an enabler of the *network slicing* features of 5G: A high density of links provides many possible physical IABs to map slices on, each one with different performance in terms of delay, robustness, etc. The second implication is that the diversity in the link length strongly impacts the performance of the network, and the choice of the technology to build links on the selected edges, as the propagation of signals changes significantly with the technology selected. This calls for techniques to build $G_\varphi$ that are tailored for the specific target area, discouraging a one-size-fits-all approach, and justifies the need for the real-world data sets we publish.

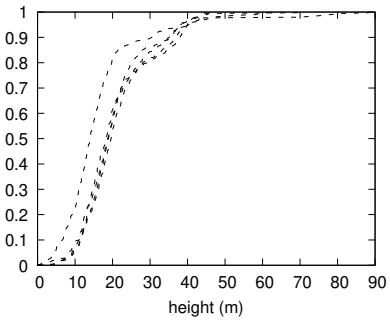
#### 3.5.3.3 *Antenna elevation*

The differences in $G(\mathcal{N}_v, \mathcal{E}_v)$ shown in Sect. 3.5.3.2 are due to two concurrent factors, a different distribution of building elevation, mostly due to terrain factors, and a different distribution of buildings in the 2D map of the area. Here we focus on the first one.




**(a)** All areas

**(b)** Urban 1, five 1 km$^2$ sub-areas




**(c)** Suburban 2, five 1 km$^2$ sub-areas

**Figure 3.32:** ECDF of the $z$ value (height) of all $p_i$ in the 9 areas a) and for 5 subareas b) and c). © 2023 IEEE

Fig. 3.32 reports the ECDF of the $z$ value for all the points $p_i$ (accounting for the ground elevation, building height and 2 m pole) in all areas and 2 sets of sub-areas. The $z$ values are referred to the lowest point of the area.

Fig. 3.32 shows clear differences between areas, without a recognizable pattern even among areas of the same kind. Rural-3 shows a bimodal behavior that is due to the earth altitude, rather than the buildings' height, while Rural-2 shows a smooth trend that is quite different from the other curves. If we zoom on a single area, we find again different situations. Urban-1 is surrounded by mountains, so that a different choice of the sub-area yield even more variability (Fig. 3.32b). The same can not be said for SubUrban-2, in which the five sub-areas show a very similar behavior (Fig. 3.32c). Again, we observe that the variability of the data do not allow a single model for all the areas, and not even for sub-areas inside the same one. This has a strong impact on the accuracy of simulations that use a single model to describe every scenario, as we discuss in Sect. 3.5.4.

#### 3.5.3.4 *Comparison with SoA LoS models*

We now compare the probability of LoS estimated on the graphs generated with TrueNets with the two LoS models introduced in Sect. 2.2.2: The ITU Model in and the Al-Hourani model. As explained, these two models were not derived for rooftop backhauls, thus the goal of this analysis is understanding if these model are somehow adequate for the design of rooftop backhauls or not.

Consider a certain area on which we compute $G(\mathcal{N}_v, \mathcal{E}_v)$, let us call $\mathcal{E}_p$ the set of all the *potential edges*, i.e. all the couples $(p_i, p_j)$ of the points in $\mathcal{N}_v$, with $i \neq j$. An edge $e = (p_i, p_j)$ is present in $\mathcal{E}_p$ even if $\Upsilon(p_i, p_j) = 0$. We call $d(e)$ the length of $e$. For computational reasons, we extract a random fraction $r$ ($r = 1\%$ in urban areas and 10% in sub-urban and rural areas) of edges from $\mathcal{E}_p$ called $\overline{\mathcal{E}}_p$, and we bin the edges in $\overline{\mathcal{E}}_p$ based on the edge length, with interval $\Delta = 200$ m:

$$B_l = \{e \in \overline{\mathcal{E}}_p \mid l\Delta < d(e) < (l+1)\Delta\} \tag{3.40}$$

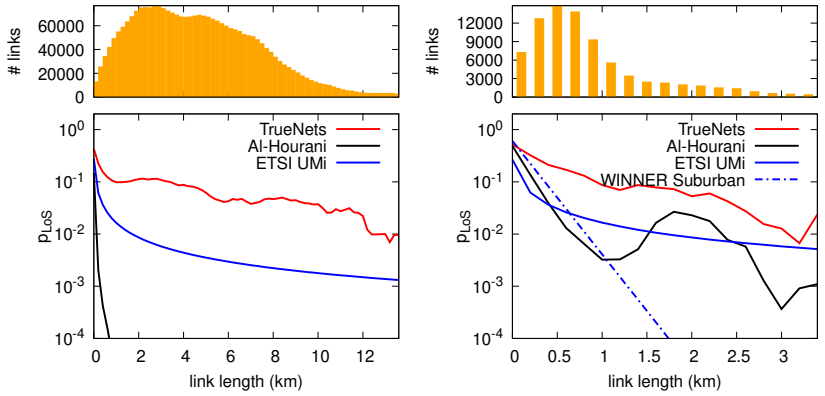
Then for each $e \in B_l$ we compute the LoS probability $P_{LoS}^{UrbanMicro(UMi)}(d(e))$ using the ETSI model as in Eq. (2.8), and we calculate the average on the whole bin:

$$M_{\text{LoS}}^{\text{ETSI}}(l) = \frac{1}{|B_l|} \sum_{e \in B_l} P_{LoS}^{UMi}(d(e)) \tag{3.41}$$

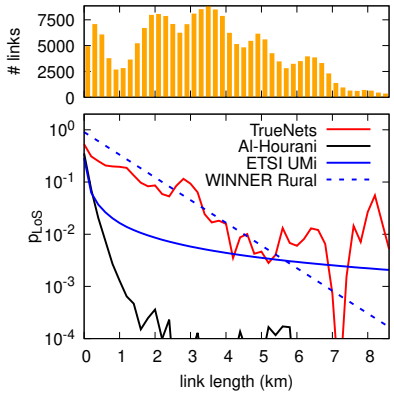
We repeat the same procedure to obtain $M_{\text{LoS}}^{\text{WINNER}}(l)$ (using the WINNER model for rural areas) and $M_{\text{LoS}}^{\text{AL-H}}(l)$ (for the Al-Hourani model). In the latter case we compute $P_{\text{LoS}}^{\text{AL-H}}(d(e), z, z')$ using Eq. (2.11), with $e = (p_i, p_j)$ and $z, z'$ the elevation values of $p_i$ and $p_j$. We replace $P_{\text{LoS}}^{\text{ETSI}}$ with $P_{\text{LoS}}^{\text{AL-H}}$ in Eq. (3.41) and we obtain $M_{\text{LoS}}^{\text{AL-H}}(l)$. Finally, we use TrueNets to compute $\Upsilon(\mathbf{E}, p_i, p_j)$ and we have a fourth value:

$$M_{\text{LoS}}(l) = \frac{1}{|B_l|} \sum_{(p_i, p_j) \in B_l} \Upsilon(\mathbf{E}, p_i, p_j) \tag{3.42}$$

Fig. 3.33 contains different curves for three different areas (the supplementary material contains all the figures for all the areas), the $x$ axis of each graph is cut to avoid the noise introduced by bins with less than 0.1% of the sampled edges. The upper graphs report the numerosity of the bins, which shows that in the urban areas the density of buildings smooths the distribution, with some small fluctuations. We observed this behavior also in the other two urban areas. The sub-urban area maintains some regularity, while the rural area shows a completely different behavior. In this case, the area seems to be partitioned in small clusters that generate the multi-modal shape of Fig. 3.33c, due to the settlement structure done by small, dense hamlets scattered in a mountain environment: two hamlets facing each other across a valley at distance $x$ give a very large contribution of LoS links around this value.




**(a)** Urban 1 **(b)** Suburban 1




**(c)** Rural 1

**Figure 3.33:** Comparison between synthetic LoS probability models (ETSI and Al-Hourani) and measures with TrueNets. © 2023 IEEE

The curves in the bottom part of the figure report the four values of $M_{\text{LoS}}$. We observe that the Al-Hourani model seems to differ largely from TrueNets values, which is due to two effects. The first is that Al-Hourani was designed for short links (below 250 m) so its application to longer links extends beyond its initial purpose. In the supplementary material

we report the curves for the range 0-500 m that show (for some cases) a trend that is closer to $M_{\text{LoS}}$. Second, Al-Hourani models the LoS of drones moving in a 3D space, so the position of the drones is in the empty area among buildings, while we put our nodes on top of the buildings. We already observed the large difference caused by a change in the position of $p_i$ inside the same building area (Tab. 3.7), so it is not surprising that the Al-Hourani model does not fit a real-world LoS probability between building roofs.

The ETSI model for Urban-Micro has a trend that is reasonably similar to $M_{\text{LoS}}$ but on a totally different scale, which suggests that it could be adapted to fit the real data at least in the initial part of the curve. One of the most interesting observation is that the WINNER models for the area types we consider (suburban and rural) yield completely different results: Not only they are a pure exponential models, but they also follow a decay that is often completely different from the measured values.

#### 3.5.3.5 *Robustness*

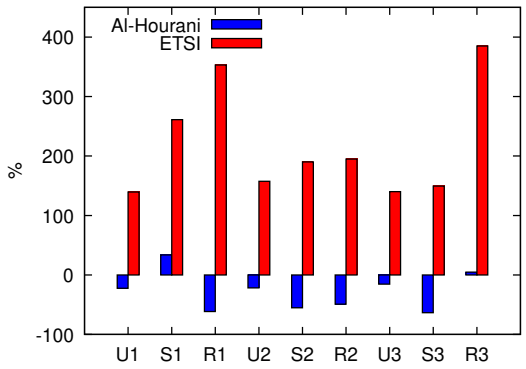



**Figure 3.34:** Effective Graph Resistance $\xi$ of ETSI and Al-Hourani models as a percentage difference from $\xi$ measured with TrueNets with 30 gNB/km$^2$. © 2023 IEEE

One intrinsic limit of both the ETSI and Al-Hourani model is that they use a radial symmetry, so given $p_i$ they assume the probability of LoS with $p_j$ is independent of the angle of the segment between $p_i$ and $p_j$, which is of course not true in a real setting. To quantify the effect of the path loss model we need to evaluate the properties of visibility graphs generated with different LoS probability models. We already studied the properties of $G(\mathcal{N}_v, \mathcal{E}_v)$ on entire areas, so here we focus on $G(\mathcal{N}_v, \mathcal{E}_v)$ built among a subset of buildings in sub-areas that are most likely to be interesting for designing real communication graphs.

We use the methodology described in Sect. 3.5.2.1 to select a random set of buildings in one sub-area, and then we build the visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$ among all the $p_i$ on the selected buildings. We repeat the process ten times with different random seeds ($\rho$ = 30 gNB/km$^2$) in all the 45 sub-areas obtaining a total of 45 $\times$ 10 random topologies.

Since availability and fault tolerance is a key requirement for nextG, we focus on a robustness metric: the *Effective Graph Resistance* $\xi$. $\xi$ takes into account the presence of parallel (possibly disjoint) paths, and is computed as the average of the resistance between any two nodes $s, d$ in the network, computed as if the graph were an electrical resistive circuit where the links have unitary resistances (the interested reader can refer to Ellens et al. for details[89]). $\xi$ is defined as:

$$\xi = |\mathcal{N}_v| \sum_{i=2}^{|\mathcal{N}_v|} \frac{1}{\lambda_i} \tag{3.43}$$

where $\lambda_i$ is the value of the $i$-th eigenvalue (ordered by their value) of the Laplacian matrix of graph $G(\mathcal{N}_v, \mathcal{E}_v)$. $\xi$ strictly decreases when an edge is added to the network, so the smaller $\xi$ the more robust is in general the mesh, but also the larger is the overall capacity of a mesh built on top of it, as it can exploit more disjoint paths between nodes. For each generated graph we compute the relative percent difference as follows:

$$100 \frac{\xi_{\text{ETSI}} - \xi_{\text{TrueNets}}}{\xi_{\text{TrueNets}}} \tag{3.44}$$

for the case of the ETSI model, and similarly for the Al-Hourani model.

Fig. 3.34 reports the values averaged on all graphs. It is evident that even using the ETSI model, that was somehow closer to the TrueNets data in Fig. 3.33, the robustness metric is completely different, and so the properties of the potential embedded $G_\varphi$ will be. The Al-Hourani model maintains a higher similarity, but robustness still differs in a range between $+33.6\%$ and $-63.4\%$. Again, this has direct implications on the design of reliable networks and in the support of integrated network slices.

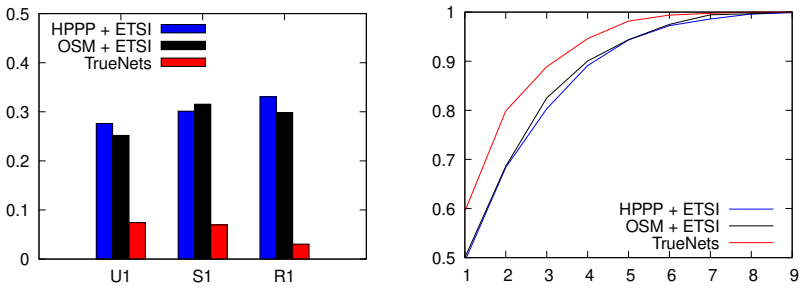
### 3.5.4 Analysis of the topology

To underline the impact of realistic data in the evaluation of scientific contributions we selected two state-of-the-art algorithms for the generation of $G_\varphi$, and we compare its performance when applied to synthetic $G(\mathcal{N}_v, \mathcal{E}_v)$ models based on HPPPs (Homogeneous Poisson Point Processes) to place nodes and synthetic models of LoS probability, or to realistic $G(\mathcal{N}_v, \mathcal{E}_v)$ estimated with TrueNets. The first one is a greedy distributed algorithm from Polese et al. [75], which was proposed as an algorithm to deploy IAB networks [76]. The second one is the classical centralized shortest path algoritm from Dijkstra used to compute a multi-source spanning tree. As documented in Sect. 3.5.2 we dissect the comparison in three parts, showing the impact of both modifications incrementally and finally together. The baseline, which we call HPPP+ETSI, is the synthetic model used by Polese [75]: $\mathcal{N}_\varphi$ is selected with an HPPP and $\mathcal{E}_\varphi$ is based on the ETSI UMi model. The intermediate one we call OSM+ETSI uses the realistic positions of $\mathcal{N}_\varphi$ from real-world data and retains the ETSI model

for the selection of $\mathcal{E}_\varphi$. Finally, TrueNets uses both realistic node location and realistic edge LoS measured from our data set.

For each of the 45 sub-areas of 1 km$^2$ and for each strategy we generate 10 different possible networks, each one with a different random choice of $\mathcal{N}_\varphi$ and donor nodes and a different resulting $\mathcal{E}_\varphi$ (the process is the same explained in Sect. 3.5.2.1) and we average the results. As usual, we report only the minimal set of results to support our conclusions, more results which corroborate the same conclusions are available in the supplementary material.

### 3.5.4.1 *Distributed Algorithm*




**(a)** $\rho = 30$ gNB$/km^2$

**(b)** $\rho = 30$ gNB$/km^2$

**Figure 3.35:** a) Ratio of unconnected nodes with DPS_WF strategy. b) ECDF of the distance of IAB nodes from the closest gNB donor in Urban-3. © 2023 IEEE

The first result, reported in Fig. 3.35 a), shows the ratio of IAB-nodes that can not be connected with a multi-hop path to any IAB-donor. These nodes are isolated and can not be part of the backhaul network. This result indicates that the main difference is due to the LoS probability model. In fact, regardless of the nodes selection strategy for $\mathcal{N}_\varphi$, the fraction of disconnected nodes remains very similar. In all three areas the difference among the baseline and OSM+ETSI (blue and black bars) is below 30%. On the other hand, the usage of a realistic visibility graph dramatically impacts the ratio of unconnected nodes, with differences up to 700% due to the strong difference in the LoS probability as reported in Fig. 3.33. Increasing the density of gNBs, the fraction of disconnected nodes becomes marginal and also differences between strategies decrease as one may expect.

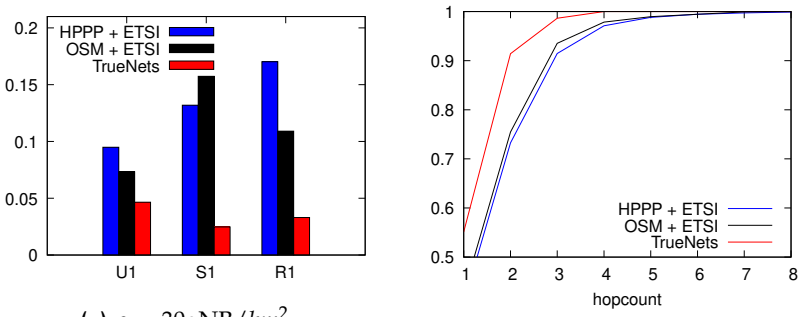
The second result, reported in plot b) of Fig. 3.35, is the metric used by the authors in the original research: The distance between each IAB node and its donor gNB in terms of hops, which is a key parameter to estimate the latency in an NGWB, but also to estimate the effective capacity given a technology to set up links on edges. The plots report the ECDFof the hopcount and show that not only the $G_\varphi$ generated using synthetic data

has between 20% and 30% fewer nodes (Fig. 3.35 a)), but also the length of the path to the donor is much longer.

### 3.5.4.2 Centralized Algorithm

The results found for the distributed algorithm generally holds true also for the centralized (and optimal) algorithm. The first result, reported in plot a) of Fig. 3.36 shows that the usage of a realistic visibility graph still provides some gains in the connectivity of nodes. In fact, albeit the values are more compressed, the usage of a realistic visibility graph diminish the number of unconnected nodes.

The second result, reported in plot b) of Fig. 3.36, also confirms the results found for the distributed algorithm. By using a realistic visibility graph, the paths are generally shorter due the fact that longer edges are less likely to be present in synthetic graphs than in reality. Additionally, we note that when employing a centralized algorithm the average number of hops diminishes significantly, suggesting that the algorithm proposed in [75] was not conceived to minimize the number of hops. Without further findings, a centralized algorithm may be preferable for delay-sensitive applications.




**(a)** $\rho = 30\text{gNB}/km^2$

**(b)** $\rho = 30\text{gNB}/km^2$

**Figure 3.36:** a) Ratio of unconnected nodes with Dijkstra strategy. b) ECDF of the distance of IAB nodes from the closest gNB donor in Urban-3. © 2023 IEEE

We can conclude that, in this case, the use of synthetic data produces results that are extremely pessimistic, primarily in terms of admission to the network, but also in terms of performance, so that an evaluation based on these models may lead to hamper the development of a technology that is instead very promising. These results do not change if we use synthetic models to estimate the LoS and realistic node positions, but they dramatically change when we use realistic data for both node positions and LoS availability, confirming the importance of data-based approaches and models for the generation of $G(\mathcal{N}_v, \mathcal{E}_v)$.

# 4 IAB TOPOLOGY OPTIMIZATION USING OPEN RAN

The advent of the 5th generation (5G) of cellular networks brought with it the promise of unprecedented speeds and connectivity. Critical to realizing this potential is the efficient deployment and management of networks in dense urban and diverse rural settings. millimeter wave (mmWave) communication stands as a promising solution for such deployments, given its capacity for multi-gigabit mobile throughput. However, challenges like the notorious short range and high susceptibility to interference associated with mmWave communications require a denser network, which in turn escalates deployment costs.

Integrated Access and Backhaul (IAB) emerges as an elegant solution to this problem. By integrating backhaul and access transmissions within the same radio technology, IAB promotes the use of a Wireless Backhaul Network (WBN) composed of several Next Generation Node Bases (gNBs) which successfully reduces the deployment costs and thus enables Mobile Network Operator (MNO) to reach such levels of densification.

This architecture, however, introduces a new dimension of complexity in the network design and management that cannot be addressed using traditional techniques. The Open Radio Access Network (O-RAN) paradigm, with its concept of RAN Intelligent Controller (RIC) and closed-loop control previously explained in Sect. 2.1.1.4, is the ideal solution to effectively optimize different aspects of IAB networks, such as energy efficiency, capacity, delay, etc.

This chapter encompasses three research works that delve into the complexity of the IAB architecture in the context of the O-RAN paradigm.

The first research, presented at the 21st Mediterranean Communication and Computer Networking Conference [6], details the challenges faced extending the O-RAN architecture to IAB, proposing a tentative architecture to integrate the two different systems that was missing. Then it validates such framework on an experimental platform by leveraging on OpenAirInterface (OAI), an open-source implementation of the 5G stack.

The second research, submitted at IEEE International Conference on Computer Communications [1], focuses on the optimization of IAB topologies, by proposing algorithms to optimize IAB networks under different constraints, such as a minimum capacity or a reliability level. Then the algorithms are framed in the O-RAN context to automatically tune the network during its lifecycle.

The third research, presented at the 2023 IEEE Global Communications Conference [5], tackles the problem of energy optimization. By showing a novel model for joint routing and energy optimization, it offers insights into leveraging the capabilities of O-RAN to maintain network efficiency while simultaneously minimizing energy costs.

It is worth mentioning that all works of this chapter build on top of the data and techniques detailed in the previous chapter. Here, in fact, we leverage optimal placements and visibility graphs to assess the performances of the experimental frameworks and optimization techniques we propose.

## 4.1 RELATED WORKS

The foundation of these studies lies in diverse yet interconnected fields of research: the first field is the newly born field of O-RAN enabled optimization of the Radio Access Network (RAN), and the second one is related to specifics optimization for IAB networks.

### 4.1.1 Closed-loop control using Open RAN

The capabilities introduced by Open RAN, enable different kinds of optimization of the RAN that were not possible before. For this reason, in recent years we have seen an increase in research works leveraging O-RAN with a variety of use cases.

The most flourishing area is, perhaps, the one involving resource allocation; with a large number of articles optimizing scheduling, slicing, and in general, the provision of network functions, many of which leverage Machine Learning based optimization. Bonati et Al. [94] provide the first demonstration of closed-loop control of softwarized cellular networks optimizing the scheduling policies towards different targets. Johnson et Al [95] similarly take advantage of an O-RAN xApp to optimize the slicing of a mobile network running on an experimental over-the-air testbed. Polese et Al. [96] propose ColO-RAN, a pipeline to design, train, and test Deep Reinforcement Learning based control in O-RAN, which is showcased by controlling slicing and scheduling policing through an xApp.

Another promising use case for closed-loop control is resource sharing. This technique allows different MNOs to share parts of, or all, the network infrastructure. By doing so both the Capital Expenditure (CapEx) and the Operational Expenditure (OpEx) needed to deploy and operate a mobile network decrease. Bonati et Al. [23] proposes NeutRAN, a framework based on the O-RAN architecture to optimize and balance the use of resources among different MNO on a single infrastructure. Baldesi et Al. [97], with ChARM, devise an O-RAN compliant framework that, leveraging on Neural Networks, senses the spectrum in real-time and, upon sensing an incumbent transmission on the current channel, re-configures the ran on a different frequency. Polese et Al. [98], similarly leverages external information about the satellite orbits to perform automatic and dynamic spectrum sharing to avoid interfering with passive incumbents.

### 4.1.2 IAB topology optimization

The optimization of wireless mesh networks is an extensively studied area of research, from planning [99] to scheduling [100], and reliability [101]. As the concept of IAB emerged more recently, there are fewer studies explicitly dealing with its optimization. Some of the most general research on mesh networks could be adapted, however, the adoption of higher frequencies and the introduction of beam-steering created new challenges that are being currently investigated. Zhang et Al. [102], for instance, leverage on Multi-Agent Reinforcement Learning techniques to optimize resource allocation in a mmWave IAB network. Pagin et Al. [103], instead, adopt a centralized approach to optimize scheduling on an IAB topology. Polese et Al. [104], devise and evaluate several distributed algorithms to select the best hop for an IAB network.

There is a lack of research that specifically leverage the Open RAN framework to optimize IAB networks, due to the novelty of both systems. Gahtan et Al. [105] leverages Deep Reinforcement Learning to optimize the scheduling of an IAB network and discuss its integration into an O-RAN framework. Munldamuri et Al. [106] showcase an IAB-like implementation using OpenAirInterface in a O-RAN network, showing how a flying IAB node can improve the throughput of an User Equipment (UE) which is in Non-LoS (NLoS) with the gNB.

In the remainder of this chapter, we will describe our efforts, both from an experimental perspective and optimization perspective, on how IAB networks can leverage on O-RAN.

## 4.2 A FRAMEWORK TO EXPERIMENT WITH IAB AND OPENRAN ON COLOSSEUM

In this effort, we lay off the basis to allow the dynamic control of IAB network through O-RAN, leveraging on OpenAirInterface and publicly available testbeds to validate our contribution.

We started by considering 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 5G IAB networks. We then proposed an experimental framework, built on top of IABEST [107], where researchers can easily deploy an end-to-end O-RAN-enabled IAB network with Over-The-Air (OTA) and hardware-in-the-loop emulation capabilities. Furthermore, to further facilitate experimental research activities, we packaged and integrated the entire framework into OpenRAN Gym, a publicly available research platform for data-driven O-RAN experimentation at scale [108]. Finally, we validated the proposed framework on Colosseum, the world's largest wireless network emulator [109], where a synthetic scenario and a realistic scenario, based on the data and techniques developed in the previous chapter, have been used to test its performance.

### 4.2.1 Integrating IAB in Open RAN

The existing 3rd Generation Partnership Project (3GPP) standards offer adaptation and control mechanisms through the IAB Backhaul Adaptation Protocol (BAP) layer, the F1 interface, and the Radio Resource Control (RRC) layer across the IAB-Donor Central Unit (CU) and the IAB-Node Distributed Unit (DU). However, the specifics of control and adaptation within these configurations are often dependent on the individual vendor's implementation.

In this context, integrating IAB networks within the O-RAN architecture becomes significant. This integration allows for the functionalities of IAB-Donor and IAB-Node to be made accessible to the RICs. These RICs have the advantage of a centralized overview of the RAN and access to a vast array of analytics and information, which are typically not available at the level of individual IAB-Donors and Nodes. For IAB networks, this

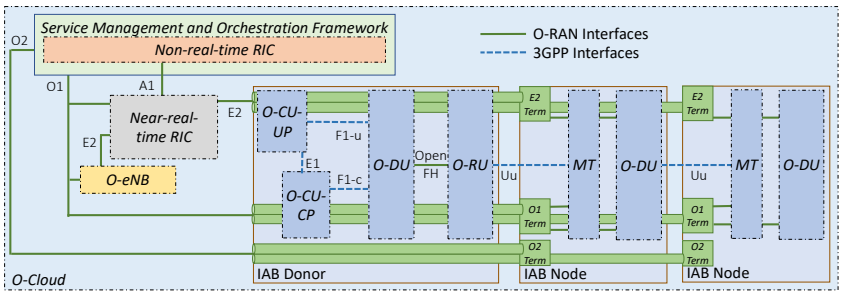



**Figure 4.1:** IAB and O-RAN integrated architectures. © 2023 IEEE

could mean improved coordination across multiple IAB-Donors, leading to reduced interference, more adaptable network topologies, and dynamic resource allocation. This approach also facilitates proactive congestion management across both access and backhaul links through data-driven strategies.

#### 4.2.1.1 Extensions to Open RAN

Adapting the O-RAN architecture to accommodate IAB deployments involves several design and architectural challenges. A key decision is whether to support O-RAN interfaces directly in the IAB-Nodes, which could involve either terminating these interfaces at the IAB-Donor or transmitting the data over the wireless backhaul. The former is simpler and does not require changes in the architecture but limits control to the capabilities of the IAB-Donor, lacking visibility into the IAB-Nodes. The latter offers more detailed control but adds complexity and necessitates data tunneling over the wireless backhaul.

3GPP anticipates the execution of Self-Organizing Network (SON)-like operations via the wireless backhaul interface. Therefore, our proposed architecture, as depicted in our framework, opts for the latter approach, which aligns closely with integrating O-RAN and IAB deployments. This approach involves encapsulating O-RAN interface data in dedicated bearers, necessitating coordination between the network's control plane and user plane transport. This could, for instance, be managed through a dedicated Packet Data Unit (PDU) session between local User Plane Functions (UPFs) in both the IAB-Donor and IAB-Node Mobile Termination (MT). Subsequently, local interface termination would be established in the IAB-Node, similar to a traditional, fiber-equipped RAN node. However, it's crucial to prioritize O-RAN traffic correctly on the wireless backhaul to maintain control objectives without negatively impacting user performance or causing congestion.

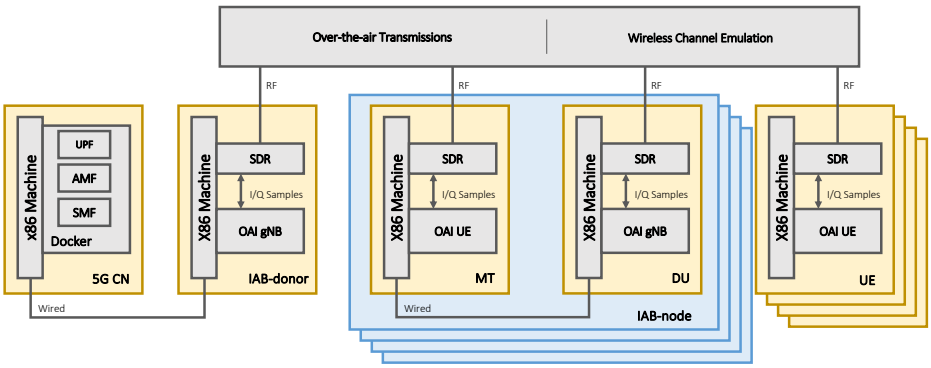



Figure 4.2: Overview of the RAN architecture deployed over white-box hardware. © 2023 IEEE

**O1 EXTENSION** The O1 interface could connect the Service Management and Orchestration (SMO) framework to the IAB-Node for tasks like

maintenance and updates of the components (MT and DU) of the IAB-Node. Since the control loops via O1 operate at a longer timescale (around 1 second or more), its traffic can be assigned a lower priority compared to E2 traffic. This scenario advocates for the use of dedicated bearers and tunnels for each O-RAN interface over the backhaul.

**E2 EXTENSION** Enhancing the E2 interface for IAB usage may require new, dedicated E2 Service Models (E2SMs). These models define the semantics of the E2 interface interactions with xApps in the near-real-time RIC. For IAB, an extension of the E2SM Key Performance Measurement (KPM) [110] can be utilized to reveal performance metrics related to the MT, in addition to the DU. Other potential targets for near-real-time control over E2 could include the allocation of resources between backhaul and access traffic or dynamic Time Division Duplexing (TDD) slot configurations to adjust to fluctuating traffic patterns.

**O2 EXTENSION** This interface could integrate IAB-Nodes as part of the O-Cloud resources. In contrast to traditional virtualization infrastructure, the IAB-Nodes in the O-Cloud are accessible and reachable over O2 only when a session is established from an IAB-Donor to the IAB-Node.

### 4.2.2 Developing an Experimental Framework for IAB and O-RAN

In this section, we present a comprehensive experimental framework that encapsulates the entire software suite necessary for operating an O-RAN-enabled IAB network, as outlined in Section 4.2.1. This framework is designed with a multi-layer architecture and is compatible with Commercial Off-the-Shelf (COTS) hardware, including generic x86 machines and USRP Software-defined Radios (SDRs).

We have customized the OpenAirInterface (OAI), an open-source 5G RAN framework [37], to support the functionality of IAB-Donors, IAB-Nodes, and IAB-capable core functions. Furthermore, our framework integrates agents for E2 and O1 interfaces into both IAB-Donor and IAB-Node, facilitating the architectural integration proposed in Section 4.2.1. These interfaces enable the non-real-time and real-time Radio Intelligent Controllers (RICs) included in our framework to manage the IAB network components effectively.

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

#### *4.2.2.1 RAN and Core Network Components*

Figure 4.2 represents an overview of the radio access functional components that enable end-to-end communication in our framework. In particular, we provide the following: a minimal yet functional deploy-

ment of 5G Core Network (CN) functions, software-defined IAB-Nodes and IAB-Donors and software-defined UEs.

**IAB-NODES AND IAB-DONORS** According to 3GPP specifications [76], an IAB-Donor consist of a CU and multiple DUs. Likewise, an IAB-Node is split into a DU and an MT. The DU is responsible for enabling the downstream connectivity, while the MT is responsible for the upstream connectivity. At the time of first developing this framework, OAI's implementation of the CU/DU functional split does not support multiple DUs connected to a single CU [111][1]. This limitation is not compliant with the IAB architecture. Thus, we employ a full OAI gNB in place of both CU and DU. As for the MT, this component is functionally equivalent to a UE, as it connects to upstream nodes using the same resources and protocols. Consequently, we have selected OAI's software-defined UE to act as MTs in this framework. The resulting system is thus made up of two concurrently running instances: an OAI gNB, acting as a DU, and an OAI UE, acting as a MT. These two components can either be deployed over two separate machines, interconnected with high-speed networking, as shown in Fig. 4.2, or on a single machine, provided that sufficient computing power is available.

**CORE NETWORK** A minimal set of 5G CN functions have been included in our framework: Network Repository Function (NRF), Access and Mobility Management Function (AMF), Slicing Magangement Framework (SMF) and User Plane Function (UPF), all based on the OAI 5G core implementation. All these functions run as Docker containers on a single x86 machine, as shown in Fig. 4.2, which hosts all the network core functions. Due to the selected IAB system design, the UPF required modifications to enable IAB operations. As previously mentioned, UEs acts as MTs in IAB-Nodes, connecting to upstream nodes. The established GPRS Tunneling Protocol (GTP) tunnels are then used to provide direct connectivity between the DU component of the IAB-node and the CN functions. However, OAI's UPF implementation lacks support for the required forwarding capability, as any packet whose destination is not a UE is dropped. Therefore, we have modified the UPF implementation to allow forwaring packets to an UE that are not intended to be received by it, so that they can be forwarded to the DU.

**USER EQUIPMENT** From the perspective of the UE, an IAB network deployed using the components described above is entirely standard-compliant. As such, both software-defined UEs (as shown in Fig. 4.2) and COTS UEs can be used in the proposed framework.

[1] At the time of writing this thesis, there exists a branch that supports multiple DU per CU, in the next future this framework will be updated to support it

#### 4.2.2.2 O-RAN Components

As mentioned earlier in Sect. 4.2.1, O-RAN defines a set of standardized and open interfaces with which the RAN exposes data collection and control primitives to the RICs. In the proposed framework, we have enabled IAB-Nodes and IAB-Donors to be O-RAN-compatible by integrating software agents for the E2 and O1 interfaces into the codebase of OAI. Furthermore, our framework comprises a Near Real-time RAN Intelligent Controller (Near-RT RIC) and a Non-Real-Time Ran Intelligent Controller (Non-RT RIC).

**O1 INTERFACE** In order to properly manage all the different aspects of networked elements, the O1 interface defines various Management Services (MnS), which can be used either from the managed entities (the gNBs) to report information back to the RIC or from the managing entity (the SMO and the rApps running on it) to deploy configurations changes, transfer files or update the software on the managed entities [19], [112]. Among all the different MnS, we have focused our contribution on implementing the Heartbeat MnS, which periodically transmits heartbeats; the Fault Supervision MnS, which reports errors and events; and the Performance Assurance MnS, which streams performance data. Those MnS have been integrated into the OAI codebase by implementing a scheduler that, running on a dedicated thread, periodically sends Virtual Network Function (VNF) Event Stream (VNF) notifications in JSON format over HTTP. This format and communication protocol has been chosen among the different options defined in the standard, as it is widely known and easily extendable by other researchers. As of now, our implementation reports performance metrics, such as the throughput and information on the channel quality between IAB-Nodes, and failure events, such as RRC or Uplink Shared Channel (UL-SCH) failures, which can be used in rApps to monitor and optimize the backhaul network. Provisioning MnS, which can be used by the rApps to deploy configuration changes (e.g., topology optimizations), have not been implemented by following the O1 specifications, as it would have needed major reworks in the OAI codebase. Instead, we have taken advantage of *IAB-Manager*, a software component we developed to orchestrate IAB experiments, as discussed next.

**E2 INTERFACE** The E2 interface is functionally split into two protocols: E2AP, tasked with establishing a connection with the Near-RT RIC, and E2SM, which implements specific monitoring and control functionalities, namely Service Models (SMs), as discussed in Section 4.2.1. In our software implementation, E2AP has been adopted from O-RAN Alliance Software Community reference implementation and, as such, it is entirely compliant with O-RAN. The E2SMs however, are defined using ASN.1: a powerful abstract description language which is, however, cumbersome and challenging to use in the fast-paced research and development envi-

ronments targeted by our framework. For this reason, we employ custom SM that are defined through Protocol Buffers (protobuf), an abstract definition language developed by Google, that is easier to handle and allows for fast prototyping and testing, facilitating the development of IAB-aware control solutions. Since the E2 interface is such that the E2SM messages are encoded and decoded only in the RAN and xApp, the custom SM definitions are transparent to the RIC, allowing our proposed solution to retain generic O-RAN compliance. At the time of this writing, we have implemented a set of protobuf messages that can be used to reproduce both the KPM and RAN Control (RC) SMs [19]. These can be used to develop data collection and control xApps, respectively.

**IAB-MANAGER** IAB networks are expected to include several IAB-Nodes, and the framework we propose can scale to such numbers. Managing experiments with tens or more RAN components is however a challenging task. Each component can be hosted by a different machine, thus setting up an IAB deployment requires the users to activate and configure each machine according to a sequence that starts from the CN functions and ends with the terminal IAB-Nodes. To facilitate experimenting at such a large scale, we have developed *IAB-Manager* [107]: an orchestrator that automates the IAB network deployment and experimentation through an Application Programming Interface (API) and a command line interface. Specifically, *IAB-Manager* is the entry point for controlling the entire experiment. It handles all the different network components, the setup of the emulated radio environment, and the management and reconfiguration of the IAB topology. Furthermore it also automates the tests and collects the results upon completion. From a functional point of view, IAB-manager instantiate a Secure Shell (SSH) connection to the machines involved in the experimentation and runs some commands to configure them according to the assigned roles. Additionally, as previously mentioned, *IAB-Manager* is used in place of the O-RAN O1 interface to execute the network configuration changes mandated by the rApps.

**RAN INTELLIGENT CONTROLLERS** The proposed framework includes standard compliant RICs, specifically the Near-RT RIC and Non-RT RIC implemented by the O-RAN Software Community.

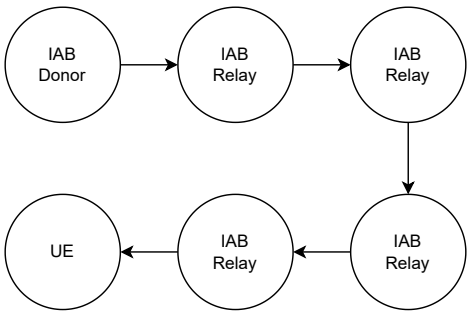
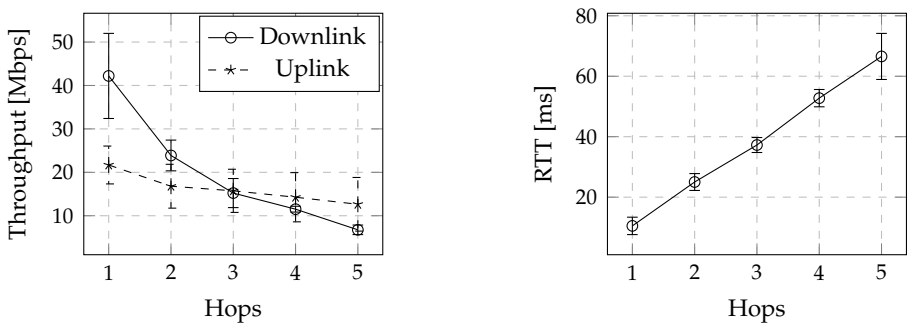



**Figure 4.3:** Linear IAB topology. © 2023 IEEE




**(a)** Throughput measurements for the linear topology.

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

**Figure 4.4:** Results for the linear chain. © 2023 IEEE

### 4.2.3 Validation and Results

This section focuses on the experimental validation of our framework. Specifically, we characterize some fundamental Key Performance Indicators (KPIs) of the deployments allowed by our IAB framework, and we validate its correct functioning.

We chose to base our validation campaign on Colosseum [109], but we stress that the openness and flexibility of the software components are such that the framework can run on generic x86 hardware. Colosseum is a publicly available large-scale testing platform with hardware-in-the-loop capabilities. It comprises 128 Standard Radio Nodes (SRNs), each composed of a powerful x86 computing node and a USRP X310 SDR. All the components of our framework can be trivially deployed on Colosseum's SRNs. Every SRN radio is interconnected by an FPGA mesh, called MCHEM, that emulates arbitrary radio channels defined through tapered delay models. Specifically it can emulate 65536 distinct channels, each one with a 4 TAPs characterization, allowing to faithfully model multipath

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

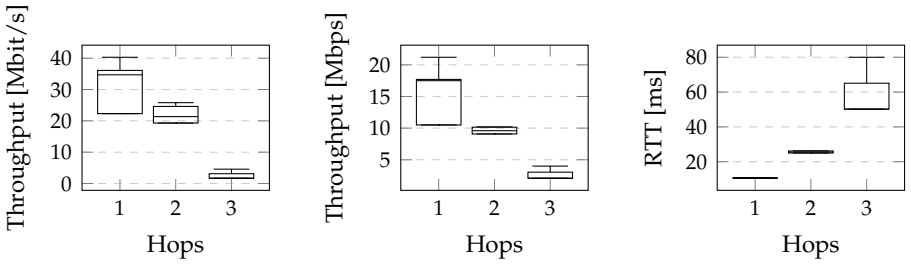
**Table 4.1:** Table of System Settings

effects. With the capability of emulating complex scenarios made of tens of nodes, Colosseum makes it possible to perform system-level evaluation of large IAB networks over complex propagation scenarios thus representing an ideal validation platform for our framework. Furthermore, Colosseum is open to the research community, and the validation tools are made available, allowing interested parties to start experimenting with a minimal initial effort.

#### 4.2.3.1 Experiments with a linear chain

We start by evaluating the performance of an IAB network deployed in a tightly controlled scenario. To this end, we consider a 5-hop linear topology, as shown in Figure 4.3. As detailed in Section 4.2.1, each IAB-Node comprises an MT and a DU, bringing this experiment's overall radio node count to 10. In order to characterize the upper-bound performance of the proposed framework, we employ an ideal propagation scenario. Through properly manipulating Colosseum's channel emulator, a 0 dB pathloss model is selected for nodes connected in the linear topology, and an infinite pathloss is set for all the other channels, effectively suppressing any possible interference. In other words, this radio scenario is equivalent to connecting the SDRs with coaxial cables.[2] Transmissions occur on band n78 with 106 Physical Resource Blocks (PRBs) available, for a total of 40MHz bandwidth.

Figure 4.4a shows the downlink and uplink TCP throughput against the number of hops, as measured between the core network and the specific MT/UE. The first-hop values of 47 Mbps in DL and 21 Mbps in UL represent the maximum throughput attainable in the testing settings. This upper bound is far from the theoretical maximum allowed by the available bandwidth. It is limited by several factors that depend on the experimental platform, OAI software implementation, and system design. Most notably, the strongest detractor to the final throughput performance is given by the OAI implementation of the software-defined UE, which is employed to build the MT. In particular, the OAI UE is inefficient in reception and transmission, thus becoming a bottleneck for the entire communication




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

**Figure 4.5:** Measurements for the realistic scenario. © 2023 IEEE

[2]

chain. Efforts are ongoing to improve the performance and stability of this software framework. Furthermore, the frameworks' system design is such that each IP packet is encapsulated into as many GTP packets as the number of hops. This increased overhead can cause packet fragmentation with a further negative impact on the overall performance. Additionally, even if the emulated channel is set to a 0 dB pathloss, Colosseum's architecture includes an unavoidable base loss of 50 dB [113] due to characteristics of the hardware architecture. This, together with the aforementioned inefficiencies, make such that packet drops and subsequent retransmissions happen also in this ideal scenario.

As the number of hops increase, the downlink throughput experiences a sharp decrease before stabilizing on a per-hop loss of around 6 Mbps. The notable throughput loss experienced at the second hop can be explained by observing the standard deviation of the throughput, represented by the whiskers in Figure 4.4a. This value is at its maximum for the first hop, suggesting that the first radio link is unstable due to the RX pipeline of the MT being overwhelmed. This substantial variability is caused by packet loss and retransmissions and internal buffer overflow, which negatively affect the performance of the second hop, as it is noticeable in the numerical results. At the same time, the second hop's throughput standard deviation is lower, as the decreased traffic volume causes less drops in the involved MTs. This stabilizing effect propagates down the topology, as both the decreasing standard deviation and the linear per-hop loss testify. On the other hand, the uplink throughput is relatively stable and close to the upper bound, even at the fourth hop. This is because the limited OAI UE performance and BS scheduling process limits the uplink traffic volume, and the gNBs are far from being overwhelmed. On the other hand, since the uplink throughput does not significantly decrease from the maximum, the UE's congestion level remains relatively stable and high, as proven by the constant standard deviation values.

RTT is measured when the network is unloaded, that is when there is no traffic flowing through the IAB network. As shown in Figure 4.4b, the first hop latency is around 11 ms. This value represents the base processing delay plus a small fixed propagation delay that is, however, the same for each hop. As the number of hops increases, the RTT experiences a linear increase comparable with the first hop latency, as expected. This shows how the system does not introduce any spurious latency when the network is unloaded. Finally, the relatively higher RTT standard deviation of the last hop (as represented by the whiskers in Figure 4.4b) suggests that multiple packet retransmissions are required.

#### 4.2.3.2 Validation over realistic RF scenarios

After having validated the system performance in a controlled environment, we move to more realistic urban scenarios, representing the typical deployment environment of an IAB network. We take advantage of the results, released as open-data, of the research on gNB placement detailed in Sect. 3.2. Specifically, we take the optimal placement for different gNB


**Figure 4.4:** Realistic deployment scenario in Florence, Italy. Donors are represented in red, while IAB-Nodes are represented in yellow. © 2023 IEEE

density values and we build the visibility graph among their locations. Then, we characterize the propagation between each gNB according to the ITU Urban Micro Pathloss Model (see Sect. 2.2.1.3), and we produce a tapered-delay representation of the communication channels, which Colosseum can then emulate. This process has been carried out for several European cities and four different scenarios are made available.[3]

Motivated by the fact that IAB is unanimously considered as a key enabler of mmWave RAN [114], we are interested in providing an experimental solution that enables testing in such conditions. While Colosseum is not directly capable of operating at frequencies higher than 6 GHz, we can approximate these radio scenarios by reproducing the most relevant propagation characteristics of mmWaves, namely the extremely directive transmissions through beamforming and the increased pathloss [115]. In particular, the pathloss between nodes that are not directly connected in the provided topologies has been set to infinite. The resulting suppression of inter-node interference might appear too ideal at first. However, this is compatible with the highly directive transmissions typical of mmWave, where interference in static conditions (i.e., as in a backhaul IAB topology) can be practically neglected [116]. A more refined mmWave channel emulation will be subject of future extensions. In addition, since Colosseum's channel emulation happens in base-band, we can apply arbitrary pathloss independently of the radio frequency employed during the experiments. Thanks to this flexibility, we could compute pathloss for a

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

carrier frequency of 28 GHz and apply them to Line of Sight (LoS) links. Nonetheless, the scenarios made available to the Colosseum community are available for both 3.6 GHz and 28 GHz, both with and without inter-node interference suppression.

For the experimental evaluation presented in this work, we have selected a scenario based on the city of Florence, Italy. Figure 4.4 shows both the urban layout and the IAB deployment, which is extended over 0.7 $km^2$ and comprises 21 nodes (3 of which are IAB-Donors). To determine which nodes are going to become IAB-Donors, we have applied the group closeness centrality metric [117] to the visibility graph. This centrality metric selects $k$ nodes such that their distance to all the other nodes is minimized. Then, we have determined the IAB topology as a Shortest-Path Forest computed over the visibility graph of the area with the well-known Dijktra's Algorithm. Similar to what has been done for the previous analysis, we characterize the throughput and latency at each hop in the network. In this case, however, the different link lengths cause performance variations in the per-hop throughput and latency. As such, we employ box plots to synthetically describe the network performance statistics in Figure 4.5. In particular, consider Figure 4.5a. Here the bottom and top edges of each box represent the first and third quartile of the downlink throughput measurements taken at all the different hops in the scenario. Similarly, the central marks indicate the median, and the whiskers represent the extreme data points. The plotted values indicate how the realistic pathloss introduced in the study scenario causes lower performance than the ideal case previously analyzed, independently of the considered hop. The same can be noted for the uplink throughput, as shown in Figure 4.5b. In both cases, the decreasing per-hop throughput trend is conserved. However, the throughput variability is the same for the two transmission directions. This is because, as opposed to the ideal scenario, the link length now represents the main performance-determining factor. This is testified by the significant distance between the first and third quartile of the first hop in both downlink and uplink throughput, which is consistent with the high variance of the first hop length in the topology of study. As for the second and third hop, the relatively closer quartiles are motivated by lower link length variations for these hops in the considered topology. Finally, the upper whiskers represent the performance of the shortest links, giving a further characterization of the system performance in this realistic scenario.

Figure 4.5c shows the RTT statistic through the same plotting technique. Differently from the throughput, the latency is not affected by the link length variations in the considered scenario for the first two hops. Additionally, the RTT increase at hops 1 and 2 is consistent with the one experienced in the controlled scenario. On the other hand, the high RTT variance of the third and last hop suggests a high probability of requiring retransmissions along the IAB path.

## 4.3 OPTIMAL TOPOLOGY DESIGN FOR IAB

In this research, we focused on pre-deployment and post-deployment approaches for the optimal identification and management of topologies across complex IAB networks, with the goal of (i) providing a minimum guaranteed area capacity, in line with the International Telecommunication Union (ITU) recommendations for next-generation cellular networks [27]; and (ii) minimizing the downtime of the IAB tree in case of failures of links between parent and child IAB nodes.

First, we introduced optimization problems that mix topological constraints (i.e., maximum degree and maximum diameter of the network graph), robustness constraints (resilience to link failure), and flow constraints (characterization of the demand and conservation of the flow on the path to a donor). The problems belong to the family of mixed Integer-Linear Problems (ILPs), which we tested on synthetic graphs to show that with accessible hardware we are able to find a very close to optimal solution for fault-tolerant topologies with flow guarantees and up to 45 nodes with a number of fiber-connected gNBs (i.e., the IAB-donors) between 66% and 23% of the total number of gNBs.

Second, we further evaluated our approach on realistic IAB topologies by using open 3D surfaces representing urban areas. In these scenarios, we were able to apply the optimization with networks of up to 60 nodes, in these large networks we need as little as 19% of the nodes to be IAB-donors.

Finally, we embedded and implemented the optimization routine using the O-RAN architecture [6], [19] in an rApp (i.e., a piece of custom control logic running in the O-RAN Non-real-time (Non-RT) RIC). The rApp dynamically recreated the IAB topology in case of link failures in the backhaul topology, based on the guidance from the optimization problem. We used Colosseum [109], the world's largest wireless network emulator with hardware-in-the-loop, to confirm the viability of our approach using the experimental framework presented in Sect. 4.2.

### 4.3.1 Problem Statement

This section introduces the proposed system model for the design of optimal, reliable, and reconfigurable IAB topologies, based on the main blocks illustrated in Fig. 4.5. First, we leverage realistic representations of the areas where the IAB network operates. Second, we develop optimization techniques that deploy optimal connectivity graphs over such topologies, including redundant paths to enforce network resiliency. Third, we leverage the O-RAN architecture, and, specifically, the O1 interface between the RAN and the Non-RT RIC, to perform reconfigurations of the network in case of failures.

IAB supports a backhaul network made of multiple trees rooted in IAB-donors. In our optimization we start from the positions of gNBs and we study how to create an optimal IAB backhaul. The number of IAB-

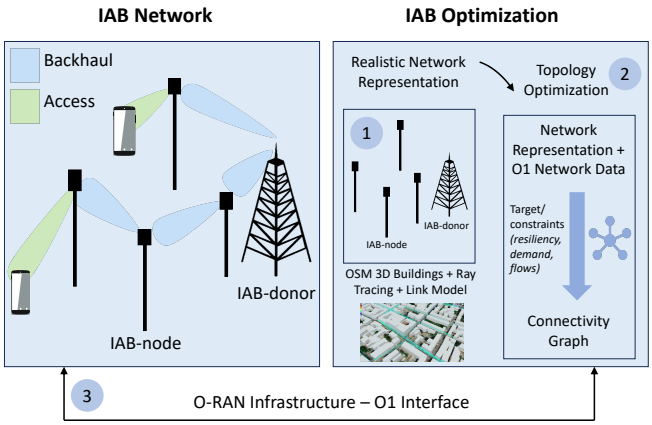



**Figure 4.5:** System model of the IAB closed-loop optimization using the (i) realistic topologies representations; (ii) graph-based optimization; and (iii) the O-RAN infrastructure.

donors needs to be minimized as they are fiber connected to the core network, and this impacts the overall capital expenditure. Based on the gNBs positions, the proposed optimization will jointly decide their roles (IAB-donor or IAB-node), and the backhaul topology, with different levels of redundancy. As such, the optimization should be performed before deploying the network, as it determines the nodes with wired access and those using a wireless backhaul, so we do not target real-time network optimization. However, we support and generate redundant paths that the controller can configure at run-time to repair link failures. Note that the focus of our work is on the backhaul network, the access network can be realized with a separate radio or using portions of the bandwidth we don't use for backhauling.

#### 4.3.1.1 *Realistic Deployment Area Representation*

The starting point of our analysis is the reconstruction of a 3D surface that represents with high fidelity a certain area that we need to cover with a next-generation network. This approach has been recently adopted by a number of papers for various goals, such as, studying propagation models [118], localization [119], network planning [10], and LoS estimation [30]. We use open data from public administrations and the OSM buildings project [120]. In this paper we use the methodology and heuristic proposed in [10] to deploy a network of gNBs in an outdoor urban area, and we then evaluate LoS and link capacity between the gNB similarly to [30]. The result is that given a certain urban scenario, and a desired density of gNB per square kilometer, we can identify the positions of the gNBs that provide an optimal outdoor coverage.

We also leverage a realistic traffic demand profile, based on the knowledge of the area covered by each gNB and requirements for next-generation wireless systems, as we detail in Sect. 4.3.2.1, and a link model based on parameters for 3GPP systems and the OAI reference implementation, as discussed in Sect. 4.3.2.2. In the remainder of the paper, we focus on

a IAB network deployed at mmWaves, to evaluate the impact of large bandwidth availability to the performance of the system.

#### 4.3.1.2 Topology Optimization

For the optimization of the network topology and access/backhaul link distribution, we consider the visibility graph $G(\mathcal{N}_v, \mathcal{E}_v)$, representing all the feasibile connection between the gNBs in a specific area. The first goal of the optimization is to find the smallest number of IAB-donors, i.e., the smallest subset $\mathcal{N}^* \subseteq \mathcal{N}_v$ that must be connected to the core network with a wired connection. The second goal is to choose the subset $\mathcal{E}^* \subseteq \mathcal{E}_v$ of edges that create a multi-hop path from every gNBs in $\mathcal{N}_v$ to some IAB-donor. The resulting topology must respect some performance and reliability requirements. All together, the final goal is to define $\mathcal{N}^*$ and a new graph $G^*(\mathcal{N}_v, \mathcal{E}^*)$ made of all the gNBs and the set $\mathcal{E}^*$ of edges resulting from the union of all the edges of all the trees. We will impose two classes of constraints: topological constraints that impose reliability features and flow conservation based on the estimated link capacity and traffic demand of every gNB.

#### 4.3.1.3 Open RAN for Optimized IAB Deployments

This generated graph is used to optimally deploy the IAB network in the first instance, but also in the context of network programmability, driven by primitives and capabilities provided by the O-RAN architecture depicted in Fig. 4.5. O-RAN introduced the Non-RT RIC, which performs optimization and service provisioning with a closed-loop control with a granularity higher than 1 s, and connects to the RAN through the O1 interface. The control logic is defined by custom applications called rApps [121]. In our previous research (see Sect. 4.2) we also extended the O-RAN architecture to support IAB operations, with the interfaces to the RICs (e.g., O1 and E2) implemented as tunnels over the backhaul network.

The IAB rApp is then used in two phases. During the network setup, it receives the optimized topology and configures each network component to enforce it. When the network is running, it continuously receives performance metrics and failure events through the O1 interface, and uses them to detect relevant changes in the network topology (e.g., the degradation of a radio link). The rApp eventually reacts by reconfiguring the IAB topology based on the optimization output. We test the rApp on Colosseum, a wireless network emulator that includes 256 SDR together with a digital channel emulator to run large-scale experiments involving real hardware and software [109].

### 4.3.2 Estimating Demand and Link Capacity

As in the previous research, the starting point of this analysis is the visibility graph computed between all gNBs of an area, whose location have been determined using the optimal placement algorithms from Sect. 3.2.

We then need to estimate the traffic demand for every gNB, and the link capacity that every point-to-point link between gNBs can offer. This will create the annotated graph $G(\mathcal{N}_v, \mathcal{E}_v)$over which we run the optimization. Both estimations are data-driven and are part of our original methodology.

#### 4.3.2.1 Estimating Demand

The ITU provides precise parameters to simulate gNB deployments when offering "extended mobile broadband" services [27], which correspond to 10 UE per gNB, each one with a demand of at least 100 Mb/s, so each gNB should be able to serve a load $\lambda = 1000$ Mb/s. This work focuses on the coverage of outdoor public areas (streets, public gardens, etc.), so assigning a fixed load per gNB is realistic only if the goal is to minimize coverage overlap in a context where gNBs can be deployed without constraints. However, in scenarios based on actual topographies, as those used in this paper, the buildings make the area of interest non homogeneous, with some gNBs that are placed in positions that cover specific areas which would otherwise be significantly shadowed [122]. As a result, coverage areas are often partially overlapping and the UEs may be shared among gNBs. This calls for a strategy to model shared demand among gNBs.

We consider a sampling of the area under analysis with one point per square meter. From the initial placement algorithm, we obtain $\Sigma = \{\sigma^0 \dots \sigma^n\}$ that is a family of sets. $\sigma^i$ includes all the discrete $(x, y)$ coordinates on the ground that are covered by gNB $i \in \mathcal{N}_v$. In this paper we consider a mmWave deployment thus we restrict $\sigma^i$ to the points on the ground that are in LoS with the gNB $i$, but the same approach can be used with a different definition (i.e., all the points with a minimum estimated capacity). The area occupied by $\sigma^i$ is then simply $|\sigma^i|$ (the number of points in the sampled 2D space). One point can belong to more than one set, so we define the multiplicity $m_{x,y} = |\{\sigma^i \in \Sigma \mid (x, y) \in \sigma^i\}|$. We define the demand $d_i$ of a gNB as:

$$d_i = \frac{|\sigma^i|\lambda}{\sum_{(x,y)\in\sigma^i} m_{x,y}} \tag{4.1}$$

Note that if $\sigma^i$ is not overlapping with any other $\sigma^j$, then $m_{x,y} = 1 \; \forall \; (x, y) \in \sigma^i$ and $d_i = \lambda$, while if gNB $i$ covers an area that overlaps with some other gNB (as it happens for the majority of the real-world scenarios) then $d_i < \lambda$.

#### 4.3.2.2 Link Capacity Estimation

To estimate the capacity of each link between a pair of nodes $(s, d)$, we model the propagation through a ray tracing analysis using the MATLAB suite, and combine a link abstraction model based on the physical layer implementation of OAI. First, we load the 3D model of the buildings obtained from OSM Buildings, then for each pair of gNBs we perform ray tracing using the shooting and bouncing method [31]. We consider up to

a maximum of 4 reflections and we ignore the effect of diffraction that is negligible at mmWave frequencies [33].

Then, we apply the Multiple Input, Multiple Output (MIMO) channel model, described in Sect. 2.2.3.1, onto each pair of nodes $(s, d)$ and finally we use the 5G capacity model, described in Sect. 2.2.5.3, to obtain a realistic estimation of the capacity $C_{s,d}$ of each link. The detailed parameters are reported in Tab. 4.2

### 4.3.3 Optimization Models

Let us consider the edges in $\mathcal{E}^*$ as directed, representing the downstream links from the donors down to the gNBs. We focus on downstream for simplicity but the problem can be trivially extended to both downstream and upstream. Given an IAB-node $i$, let the out-degree $i$ be its number of direct children and the distance of $i$ be the number of hops that separates $i$ from the IAB-donor.

Let the binary variable $u_{i,l}$ be equal to 1 if and only if node $i$ is at distance $l$ from the IAB-donor. If $u_{i,0} = 1$, then node $i$ is an IAB-donor. Let $P_{i,j}$ be a binary variable that is set to 1 if and only if the potential edge from $i$ to $j$ is part of $\mathcal{E}^*$. The optimization algorithm will assign both sets of variables and thus define the network topology. The parameters of the optimizations are:

- $D$: the maximum distance from a gNB to an IAB-donor;
- $\delta$: out-degree of a gNB;
- $e_{i,j}$: a set of parameters so that $e_{i,j} = 1$ if and only if the edge from $i$ to $j$ is present in $\mathcal{E}_v$, i.e., the SNR is sufficient to negotiate a link.

The objective of the optimization is to minimize the number of donors:

$$\text{objective:} \quad \min \sum_{i \in \mathcal{N}_v} u_{i,0}. \tag{4.2}$$

We first introduce some topological conditions:

$$\sum_{l=0}^{D} u_{i,l} = 1 \quad \forall i \tag{4.3}$$

$$\sum_{i \in \mathcal{N}_v} P_{i,j} = 1 - u_{j,0} \quad \forall j \tag{4.4}$$

$$\sum_{j \in \mathcal{N}_v} P_{i,j} < \delta \quad \forall i \tag{4.5}$$

$$P_{i,j} \leq 1 - u_{j,l} + u_{i,l-1} \quad \forall i, j \quad \forall l \in \{1 \dots D\} \tag{4.6}$$

$$P_{i,j} \leq e_{i,j} \quad \forall i, j \tag{4.7}$$

$$P_{i,j} + P_{j,i} \leq 1. \tag{4.8}$$

Briefly, the equations can be explained as follows: Eq. (4.3) imposes that every IAB-node is assigned to one single tree rooted in one IAB-donor;

| Parameter | Value |
|---|---|
| Carrier frequency ($f_c$) | 27 GHz |
| Bandwidth ($B$) | 400 MHz |
| Resource Blocks ($RB$) | 132 |
| Numerology ($\mu$) | 3 |
| Uplink to downlink slot ratio ($R_{slot}$) | 0.7 |
| Control channel overhead ($Oh$) | 0.18 |
| Noise density ($N_0$) | $-174$ dBm/Hz |
| Noise Figure ($N_f$) | 7 dB |
| Antenna Elements | 8x8 |
| Maximum number of reflection | 4 |
| MIMO Layers ($\mu$) | [1,2] |
| Transmission power ($\Pi$) | 33 dB m |
| Maximum BLER | 0.1 |

**Table 4.2:** Simulation parameters. Most parameters are adapted from [27].

Eq. (4.4) imposes that IAB-donors have no incoming edges, and other nodes have one (thus it is a tree topology); Eq. (4.5) imposes the maximum out-degree; Eq. (4.6) imposes consistency on paths, hop by hop, that is, if a link connecting IAB-node $i$ to IAB-node $j$ is active, then the distance of IAB-node $i$ equals the distance of IAB-node $j$ minus one; Eq. (4.7) imposes that only existing links can be used; Eq. (4.8) imposes that links are unidirectional.

The problem is a merge between two known problems, a multi-commodity flow problem, and the shortest path tree problem, both of which can be modeled with an ILP approach. There are however two key differences. The first is that the sources of the commodity are not decided a-priori, but must be decided by the optimization. The second is that we give to the network designer the freedom to choose a maximum out degree of and a maximum distance from the IAB-donors. The second parameter in particular will affect the delay of the traffic, so it is important that the operator can define it based on the specific applications that must be supported.

This model has a number of variables that scales as $|\mathcal{N}_v|^2$ as the number of variables is dominated by the dimension of $P_{i,j}$.

#### 4.3.3.1 Link-Failure-Resilient Model

The previous formulation can be extended to provide edge-failure resiliency, which improves IAB operations at mmWaves, where links between gNBs are subject to blockage causing an abrupt degradation of the performance. We can divide the set of edges in $R$ disjoint sets, and impose that every tree rooted in an IAB-donor must use edges coming from only one out of $R$ sets, while every IAB-node must belong to all trees. If $R = 2$, then there are two separate trees that serve each IAB-node $i$, so that if one edge fails in one tree, the second tree can be used to serve IAB-node

$i$. This requires the reconfiguration of the backhaul topology, which we implement in the rApp. The model can be formulated as follows: let $u_{i,l,k}$ be the binary variable that is set to 1 if IAB-node $i$ is at distance $l$ from the IAB-donor that is the root of a tree whose edges belong to the sub-set $k \in [1, \dots R]$. $P_{i,j,k}$ is the binary variable that is set to one if edge from $i$ to $j$ is active in a tree that belongs to sub-set $k$.

The objective of the optimization is still the minimization of the number of IAB-donors:

$$\text{objective:} \quad \min \sum_{i \in \mathcal{N}_v} \sum_{k=1}^{R} u_{i,0,k}, \tag{4.9}$$

under the following constraints:

$$\sum_{l=0}^{D} u_{i,l,k} = 1 - \sum_{r \neq k} u_{i,0,r} \quad \forall i, \forall k \tag{4.10}$$

$$\sum_{k} u_{i,0,k} \leq 1 \quad \forall i \tag{4.11}$$

$$\sum_{i \in \mathcal{N}_v} P_{i,j,k} = 1 - \sum_{r} u_{j,0,r} \quad \forall j, k \tag{4.12}$$

$$\sum_{j \in \mathcal{N}_v} P_{i,j,k} < \delta \quad \forall i, k \tag{4.13}$$

$$P_{i,j,k} \leq 1 - u_{j,l,k} + u_{i,l-1,k} \quad \forall i, j, k \quad \forall l \in \{1 \dots D\} \tag{4.14}$$

$$P_{i,j,k} \leq e_{i,j} \quad \forall i, j, k \tag{4.15}$$

$$\sum_{k} P_{i,j,k} + \sum_{k} P_{j,i,k} \leq 1 \quad \forall i, j. \tag{4.16}$$

These equations extend the previous formulation in a multi-tree version, thus we explain only the ones that differ significantly from their single-tree formulation: Eq. (4.10) imposes that every gNB must belong to at least one tree per edge-set $k$, unless it is an IAB-donor, which must belong to exactly one tree, the one that it is the root of; Eq. (4.11) imposes that each IAB-donor must be the root of a single tree; Eq. (4.16) imposes that every edge is used only in one direction and only in one tree. This last equation produces trees that are edge-disjoint. As a result we have $R$ disjoint sets of edges, however we may have more trees, because the topological constraints and the availability of edges may make it impossible to have only $R$ trees. Each tree will have disjoint edges from each other tree, and each IAB-node will have two parents, belonging to two different trees rooted in two different IAB-donors.

Fig. 4.6 shows an example of an optimized topology when $D = 3$ and $R = 2$. We can see that the optimization creates two separate edge-set, represented using two colors. The optimization yields 3 IAB-donors out of 18 total gNBs (thus 15 IAB-nodes), and each IAB-donor is the root of a tree using edges coming only from one set. Every IAB-node has two parents connected with links using two distinct colors. As a consequence, if one edge fails and an IAB-node can not reach the IAB-donor at the root of the tree using the blue edge-set, it can still reach a IAB-donor using the tree made of orange edges.

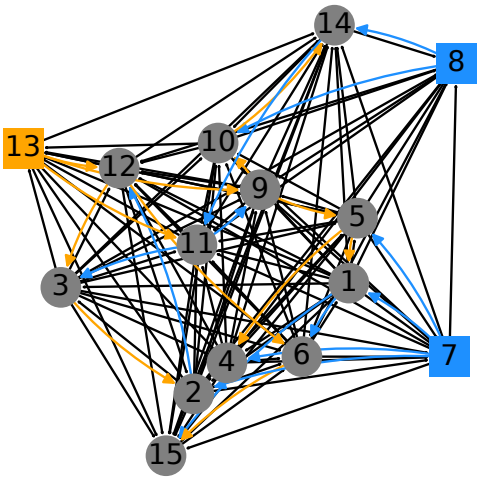



**Figure 4.6:** An example realization of the robust backhaul graph with $R = 2$. Circles are IAB-nodes, squares are IAB-donors, orange/blue edges belong to two disjoint edge-set, grey edges are all the potential edges. Every grey node has two incoming edges.

The cost associated to using this model is the added complexity, as it comes with a number of variables that scales as s $R|\mathcal{N}_v|^2$ as it is dominated by the dimension of $P_{i,j,k}$.

#### 4.3.3.2 *Flow-based Model*

Finally, we introduce the flow-based model, which can be implemented as a set of new equations on top of the topology-based ones, thus producing a topology and flow-based model. The flow-based model introduces three new sets of parameters, that are $d_i$ (the flow demand for IAB-node $i$), $C$ (the total maximum outgoing flow for an IAB-donor), and $L_{i,j}$ (the link capacity of $e_{i,j}$). It also introduces two new sets of variables: $f_{i,j,h}$ that is the flow destined to IAB-node $h$ that passes through edge $e_{i,j}$, and $a_{i,j}$ that is a real value in $[0, 1]$ that represents the fraction of time link $e_{i,j}$ is used. As each node $j$ has a single radio for access and backhaul, it becomes necessary to impose a constraint on the total number of active links for $j$, modeled through variable $a_{i,j}$. If $j$ is connected to, for example, node 1 in upstream and node 2 in downstream, then $a_{1,j} + a_{2,j} \leq 1$. In this multi-commodity flow problem, IAB-nodes are not marked a priori as a

generators or consumers of the commodity, and the decision is part of the optimization. The constraints are set as follows:

$$f_{i,j,i} = 0 \quad \forall\, i,j \tag{4.17}$$

$$\sum_h \sum_j f_{i,j,h} - \sum_{h \neq i} \sum_j f_{j,i,h} <= C u_{i,0} \quad \forall\, i \tag{4.18}$$

$$\sum_j f_{j,i,i} \geq (1 - u_{i,0}) d_i \quad \forall\, i \tag{4.19}$$

$$\sum_h f_{i,j,h} \leq L_{i,j} a_{i,j} \quad \forall\, i,j \tag{4.20}$$

$$a_{i,j} \leq P_{i,j} \quad \forall\, i,j \tag{4.21}$$

$$\sum_i \sum_h f_{i,j,h} \leq \sum_i L_{i,j} a_{i,j} \forall j \tag{4.22}$$

$$\sum_i (a_{i,j} + a_{j,i}) \leq 1 \;\forall\, j \tag{4.23}$$

$$\sum_h f_{i,j,h} \leq a_{i,j} L_{i,j} \quad \forall i,j. \tag{4.24}$$

Here, Eq. (4.18) imposes flow conservation for all IAB-nodes, except for donors that can provide at most $C$. Note that the real flow outgoing node $i$ is capped in every link by Eq. (4.20), so $C$ is just an arbitrary non zero upper bound that makes the equation formally correct, and it can model the bit-rate available on the wired link. Eq. (4.19) imposes that every IAB-node receives at least the amount of flow corresponding to its demand; Eq. (4.20) imposes that flow on a link does not exceed the capacity of the link itself, limited to the fraction of time it was allocated; Eq. (4.21) imposes that the usage on a link must be zero if the link is not chosen by the topology optimization; Eq. (4.22) imposes that an IAB-node can not receive in input more flow that what all its incoming link allow; Eq. (4.23) imposes that the fractional time allocated for all links does not exceed 1 for every IAB-node; and, finally, Eq. (4.24) imposes that flow on a link does not exceed its allocated capacity. The objective function is still Eq. (4.9) as these conditions just impose more constraints on the choice of the IAB-donors. There is always one trivial acceptable solution in which all gNBs are IAB-donors.

Again, the cost associated with the model is the added complexity. This model has a number of variables that scales as $|\mathcal{N}_v|^3$ as the number of variables is dominated by the dimension of $f_{i,j,h}$. However, since every IAB-node has only one parent if the capacity of one incoming link to IAB-node $j$ is lower than the demand of IAB-node $j$ ($L_{i,j} < d_j$), we can set $e_{i,j} = 0$ as that link can not be used for serving the demand of $d_j$. This allows us to prune some edges in $\mathcal{E}_v$ before running the optimization.

Further, if the underlying topology model is the link-failure-resilient one, some equations need to be slightly modified. First of all, the flow variables need to be independent for each tree so they gain a new index: $f_{i,j,h,k}$ now represents the flow destined to IAB-node $h$ passing through edge $e_{i,j}$ belonging to tree $k$. This is required because even if edges can not belong to more than one tree, the flow constraints need to be guaranteed in each edge and in every tree. Similarly, $u_{i,j}$ is replaced with $u_{i,j,k}$, which

is true if the distance of IAB-node $i$ from IAB-donor in tree $k$ is $j$. The equations used for the single-tree flow-based model need to be adjusted with the new variables, together with a new constraint that imposes that if an edge is not assigned to tree $k$ the corresponding flow in tree $k$ is zero, i.e.,

$$f_{i,j,h,k} \leq L_{i,j} P_{i,j,k} \quad \forall i, j, h, k \tag{4.25}$$

The total number of variables becomes $R|\mathcal{N}_v|^3$.

Finally, while our model is designed for a greenfield deployment, in which the operator is planning the network from scratch, it can also be used in a brownfield deployment in which some IAB-donors are already connected to the core with a fiber connection. The only required change in the model is the need to force some of the $u_{i,0,k}$ variables to be constant set to 1. This is important because we can use the optimization for both new networks, or to upgrade or dynamically control existing ones, as shown in Sect. 4.3.5.

### 4.3.4 Numerical Results

In this section we report on the tests realized to verify the feasibility and the effectiveness of the proposed optimization, leveraging the resilient and non-resilient versions of the flow optimization problem. We use as a metric the fraction $\rho$ of IAB-donors in the network, which is the ratio between the value of the objective function and the total size of the network $|\mathcal{N}_v|$:

$$\rho = \frac{\sum_{i \in \mathcal{N}_v} \sum_{k=1}^{R} u_{i,0,k}}{|\mathcal{N}_v|}.$$

First, we consider results obtained on synthetic random graphs, in order to verify the feasibility of our approach in a controlled scenario, then we use realistic graphs generated from open data of four European cities.

#### *4.3.4.1 Synthetic Graphs*

The synthetic graphs are generated by placing nodes in random positions in a 2D area, with an average density of 45 gNBs per square km, that is the density reported to be sufficient to achieve 95% coverage of the outdoor urban areas [10]. We use a constant gNB density, and vary the area to deploy 15, 30, and 45 gNBs. For each gNB, the demand is estimated with Eq. (4.1), assuming a coverage radius of 100 m. We use the 3GPP TR 38.901 technical report for modeling both the probability of LoS between two gNBs and the path loss (using two different decay values for LoS and NLoS links) [26]. We thus do not explicitly model obstacles, and obtain networks with an edge density substantially higher than in real settings. Given the path loss on every link, we compute the link capacity as described in Sect. 4.3.2.2, using the parameters in Tab. 4.2 and varying the number $\mu$ of MIMO layers from 1 to 2. Execution times are evaluated on a 16 cores server (Intel Xeon Gold 6342 CPU @ 2.80GHz), with 64GB

of RAM using the Gurobi solver with 30 randomly generated graphs for each number of gNBs.

Fig. 4.7 shows the ratio $\rho$ between IAB-donors and total number of gNBs in the simulation. The top row reports the results for the failure resilient model ($R = 2$), while the bottom row reports the results with only one tree ($R = 1$). In the left column, $\mu = 1$, while in the right column, $\mu = 2$. As a general trend, higher $\mu$ and lower $R$ decreases $\rho$. This is expected since with a higher capacity per link the backhaul network can transport more traffic to a smaller number of donors, while redundancy requires more independent trees and thus more donors. Increasing $|\mathcal{N}_v|$ enlarges the space of possible results and can lead to a further reduction of the number of required IAB-donors. For the most challenging configuration ($\mu = 1, R = 2$), the median value of $\rho$ ranges from 0.66 to 0.6, which means that we can save up to 40% of the donors. Savings increase to up to 60% with $\mu = 2, R = 2$ and 45 gNB/km$^2$, which means that 27 gNBs out or 45 do not need to be connected with a fiber cable.

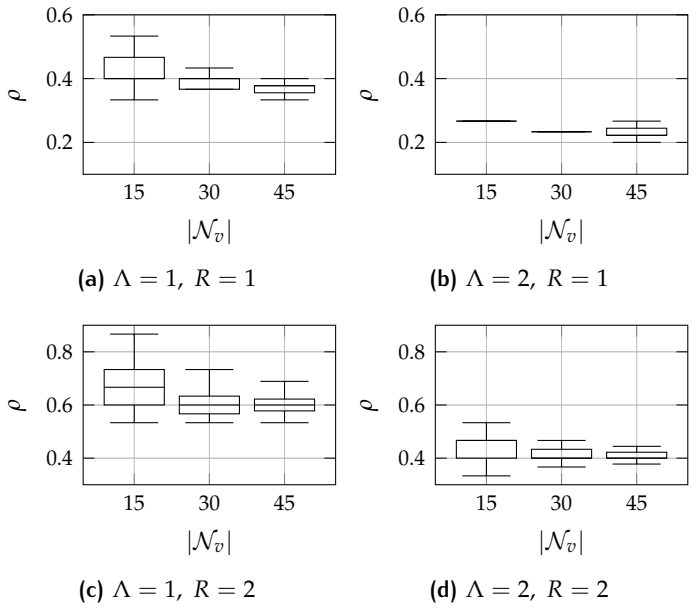
The failure resilient topology imposes a very strong condition, i.e., that all flow is conserved and there is no performance degradation upon the failure of a link. If we relax this condition, and set $R = 1$, we can further reduce the number of IAB-donors down to 37% (45 gNBs, $\mu = 1$) and 23% (45 gNBs, $\mu = 2$), introducing a non zero probability of network outage. This does not necessarily imply that the failure of one link disconnects some UE, but it could produce a certain performance degradation. Specifically tailored models could provide a trade-off between these two approaches, which obtain $\rho$ close to the ones generated with $R = 1$ with a predictable performance penalty in case of failure.

Finally, we ran the optimization for at most 48 hours. In 67% of the cases this produces the guaranteed optimal solution, in the rest of the cases, the upper bound of the distance of $\rho$ from the global optimum is on average 6.3%. This is perfectly compatible with the use-case we target, that is, the planning of a network that is yet to be realized.

#### 4.3.4.2 Realistic Graphs

Here we report the results with realistic graphs, which were generated for 4 cities (Florence and Milan in Italy, Barcelona in Spain, and Luxembourg city). For each city we draw on a map an area of roughly one square km, as in Fig. 4.8. As we select areas following the boundaries of streets, the area size will not be exactly 1 square km and will be different from city to city so the exact number of gNBs is not constant. For each scenario, we use densities of 30, 45, and 60 gNBs/km$^2$, we run the placement heuristic to position the gNBs and we then compute the capacity of each link using ray-tracing as described in Sect. 4.3.2.2. Some gNBs are placed in locations that do not allow them to have any neighbor (the Signal-to-Noise-Ratio (SNR) is too low to negotiate a link) and in those cases, we remove them from the graph, as they can not be part of the backhaul network.

Fig. 4.9 shows $\rho$ in the realistic scenario, that confirm results obtained in the synthetic scenarios, i.e., as the density increases (moving left to




(a) $\Lambda = 1,\ R = 1$ (b) $\Lambda = 2,\ R = 1$

(c) $\Lambda = 1,\ R = 2$ (d) $\Lambda = 2,\ R = 2$

**Figure 4.7:** Box-plots of the fraction of donors $\rho$ for the synthetic topologies, with different MIMO configuration ($\mu = 1$, $\mu = 2$), the single tree model ($R = 1$) and the failure resistant model ($R = 2$). The box plot show the median, 25% and 75% quartile and 1.5*IQR (inter quantile range) whiskers.

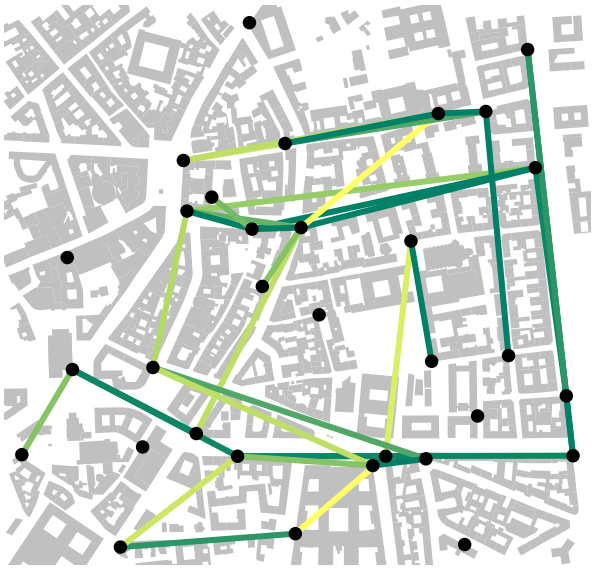


**Figure 4.8:** Map showing one of the four realistic networks in the city of Milan, with density 45 gNBs/km$^2$. IAB-nodes are depicted in black, and feasible backhaul links are depicted with a color gradient corresponding to their capacity (yellow for low capacity and green for high capacity).

right in the figure), fewer IAB-donors are required. Remarkably, since the presence of a realistic distribution of obstacles makes topologies less dense, we can complete the optimization even with 60 gNBs/km$^2$, and in some configurations $\rho$ is lower than 20% (Fig. 4.9(b)). Only in 4 cases over 48 runs the optimization did not reach the optimal value, and in those 4 cases the upper bound of the distance of $\rho$ from the global optimum is on average 6%.

Overall our results confirm that in both synthetic and realistic scenarios our optimization scheme is fully usable and produces robust topologies that can save a large number of IAB-donors. This turns into a massive cut of capital expenditure costs given by the use of wireless backhaul links instead of wired ones.

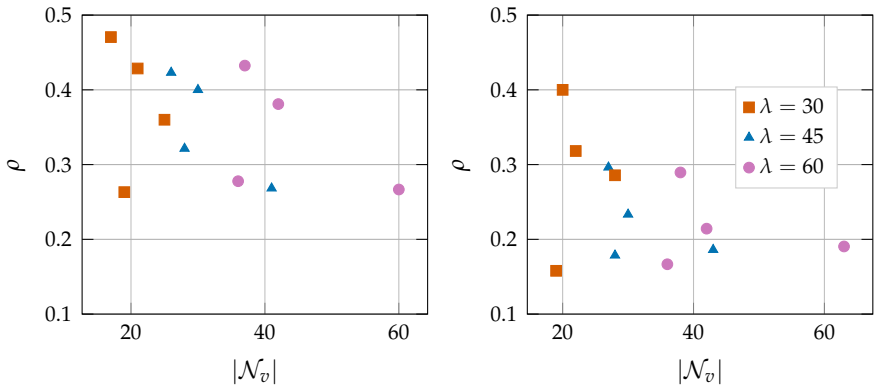
### 4.3.5 Implementation and validation of the rApp

The goal of this section is to show that we can integrate our methodology within the O-RAN framework, and so our approach provides all the necessary tools to set-up and operate a real network. We focus on implementing a prototype rApp running in the Non-RT RIC that takes in input a network made of 2 trees (a primary and a back-up one) and gracefully handles the failure of a link by reconfiguring all the gNBs to switch from their primary tree to the backup one. We validate the rApp using OpenAirInterface on Colosseum.

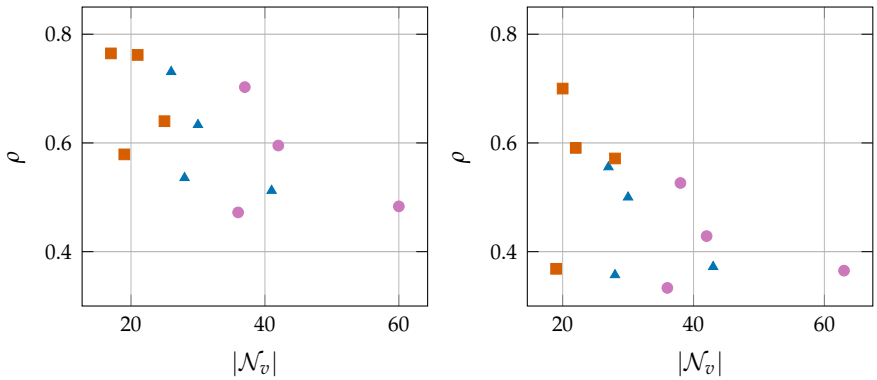
The application gets initialized by the user with the multitree topology obtained in the network planning phase. Once the network is set-up and running, the rApp maintains an updated representation of the IAB network using some specific messages (HeartBeats, Performance Reports, and Fault Events) arriving from the RAN through the O-RAN O1 Interface (see Fig. 4.5). Upon the failure of a radio link between IAB-nodes, the upstream node, still connected to the core, detects the RRC failure and sends a Failure Event message to the RIC through the O1 interface. The message gets forwarded to the rApp, which reconfigures the network by restarting the IAB-node which has lost connectivity, so that it connects to the other IAB-node. It also restarts all the IAB-nodes downstream since they need to reconnect to the network core using a new path. Finally, it reconfigures the core of the network so that the routing table to reach the downstream IAB-nodes reflects the new topology.

#### *4.3.5.1 Validation on Colosseum*

The current implementation of OAI does not support fast handover between the network components, so it introduces fluctuations and long delays. For this reason, our goal is not to provide an accurate performance analysis, but to show that our approach can be implemented in a real network. We thus have set up a representative IAB-network topology, shown in Fig. 4.10, comprising two IAB-donors, two IAB-nodes and twenty UEs. Each IAB-node is connected to a different IAB-donor, and a feasible (but




**(a)** $\Lambda = 1,\ R = 1$ **(b)** $\Lambda = 2,\ R = 1$




**(c)** $\Lambda = 1,\ R = 2$ **(d)** $\Lambda = 2,\ R = 2$

**Figure 4.9:** Fraction of donors $\rho$ for the realistic topologies, with different MIMO configuration ($\mu = 1$, $\mu = 2$), the single tree model ($R = 1$) and the failure resilient model ($R = 2$).

unused) link is available between the two IAB-nodes. Note that we assume gNBs have a secondary communication channel for their management through the O1 interface, possibly using the sub-GHz bands in order to be less subject to path loss due to obstacles [123].

After the rApp initializes the whole network, all UEs and all IAB-nodes start sending Internet Control Message Protocol (ICMP) packets to a server in the core network, routed through the parent IAB-node or donor. Then, by altering the radio environment we simulate the failure of the link between IAB-node 2 and its donor. This triggers the transmission of a Fault Event message to the rApp, that in turn triggers the reconfiguration of IAB-node 2, which will end up passing through IAB-node 1 to reach the core. Since the current implementation of IAB for Colosseum and OAI creates end-to-end tunnels from UEs to the core network [6], the UEs attached to IAB-node 2 must also be reconfigured. Moreover, the rApp reconfigures the 5G core, so that the backwards path to the UEs is also restored.

Figure 4.11 shows how the RTT of the transmitted packets is affected during and after the reconfiguration of the network. Since the RTT on the experimental testbed is noisy, we compute a rolling average with a window of 10 s, and we average the RTT of all the UEs connected to the same IAB-node. At $t = 0$ both IAB-nodes report a RTT of roughly 10 ms and their UEs measure an RTT of roughly 24 ms, which accounts for one more wireless hop and some switching time. At time $t = 19$ s, we induce the failure of the link between IAB-node 2 and IAB-donor 2, which interrupts the successful transmission of the ICMP packets and triggers the reconfiguration of the network from the rApp. OAI does not yet support a fast handover so the whole stack needs to be rebooted after the reconfiguration, which takes roughly 15 s. Around second $t = 35$ s, IAB-node 2 is able to ping the core again, with an average RTT of 22 ms, due to the additional hop. At that point, also the software-defined UEs stack restarts, adding 25 more seconds before the UEs successfully ping the core, with their RTT increased to an average of 47 ms, corresponding to 3 hops of distance to the core and two switching delays.

Albeit the current state of OAI affects the performance of the reconfiguration, this experiment fully confirms the practical viability of our approach in a realistic network scenario, based on the O-RAN specifications.

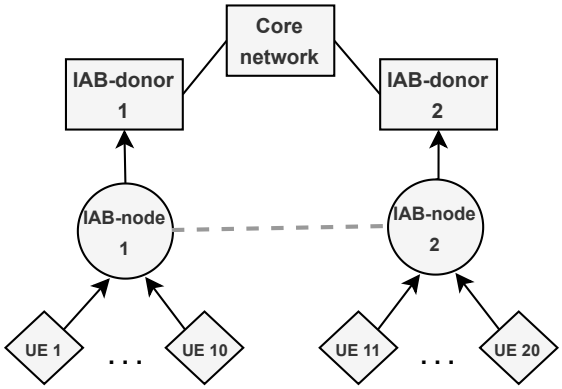



**Figure 4.10:** IAB topology including the Core network, two IAB-donors, two IAB-nodes and 20 UEs. Black arrows indicate the normal topology and the grey dotted line indicates the backup link between the two IAB-nodes.

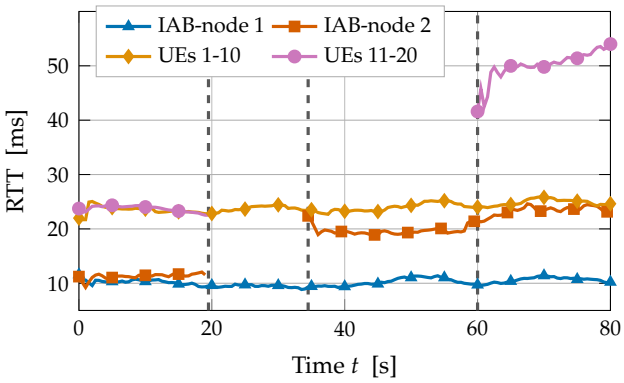



**Figure 4.11:** RTT from the IAB-nodes and the UEs before, during, and after the link failure. The vertical lines indicate the time of fault and the time of the two recoveries.

## 4.4 ENERGY-EFFICIENT IAB TOPOLOGY CONTROL

In this last research, we showcased how IAB networks can be optimized to minimize energy consumption, by leveraging on closed-loop control logic deployed in the RIC.

Specifically, we introduced the concept of *measurements graph*, a representation of all the potential wireless links of an IAB network, that can be obtained by collecting metrics and messages from the RAN through the O-RAN interfaces. On this representation, we run an optimization algorithm, developed ad-hoc for this contribution, that selectively deactivates certain IAB nodes that are not essential to guarantee a certain level of service to the UEs.

To validate our optimization model, we devised a time-varying demand model for UEs from openly available datasets. This model, which describes an ideal week in the city center of Milan, Italy, estimates the density of UEs per hour of the day. We then used this information to generate 168 instances of the IAB measurements graph (one per each hour of the week).

Finally, we validated our optimization model on all the different measurements graph, showing that the proposed optimization model reduces the energy consumption of the RAN by 47%, while guaranteeing a minimum downlink capacity per each UE equal to $80 Mb\ s$.

### 4.4.1 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. It is conceptually similar to a visibility graph, but instead of being constructed by applying visibility analysis algorithms on open-data, it is constructed dynamically by the network, by collecting measurements from the RAN.

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. Fig. 4.12a and Fig. 4.12b 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.

#### 4.4.1.1 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 [124], 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 [125]—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 multicommodity 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

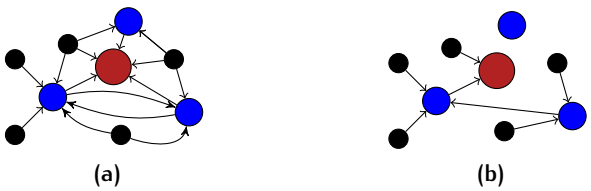



Figure 4.12: Example of a measurements graph $\mathcal{G}$ (a) and a possible IAB Tree $\mathcal{T}$ (b). IAB-donors are depicted in red, IAB-nodes in blue, UEs in black, and deactivated IAB-nodes in light blue. © 2024 IEEE

bandwidth to reserve on the path from $s_k$ to $t_k$. These commodities are decided by the 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:

$$\min \quad \sum_{v \in \mathcal{V}} a(v) \tag{4.26}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} f_k(u,v) \cdot d_k \leq c(u,v) \times \frac{1}{\sum_{w \in \mathcal{N}_{in}(v)} f(w,v)} \quad \forall (u,v) \in \mathcal{E}, \forall k \in \mathcal{K} \tag{4.27}$$

$$\sum_{v \in \mathcal{V}} f_k(u,v) - \sum_{v \in \mathcal{V}} f_k(v,u) = 0 \qquad \forall u \in \mathcal{V}, \forall k \in \mathcal{K} \tag{4.28}$$

$$\sum_{v \in \mathcal{V}} f_k(s_k,v) - \sum_{v \in \mathcal{V}} f_k(v,s_k) = 1 \qquad \forall k \in \mathcal{K} \tag{4.29}$$

$$\sum_{v \in \mathcal{V}} f_k(v,t_k) - \sum_{v \in \mathcal{V}} f_k(t_k,v) = -1 \qquad \forall k \in \mathcal{K} \tag{4.30}$$

$$a(v) \geq \frac{1}{deg(v)} \left[ \sum_{\forall u \in \mathcal{N}_{in}(v)} f(u,v) + \sum_{\forall u \in \mathcal{N}_{out}(v)} f(v,u) \right] \qquad \forall (v) \in \mathcal{E} \tag{4.31}$$

$$f(u,v) \geq f_k(u,v) \qquad \forall (u,v) \in \mathcal{E}, k \in \mathcal{K} \tag{4.32}$$

$$\sum_{v \in \mathcal{N}_{out}(u)} f(u,v) \leq 1 \qquad \forall u \in \mathcal{V} \tag{4.33}$$

$$a(v), f(u,v), f_i(u,v) \in \{0,1\} \tag{4.34}$$

Our objective in Eq. (4.26) is to minimize the number of nodes that are turned on, *e.g.*, the energy consumption of the network. In Eq. (4.31), the value of variable $a(v)$ is enforced to be 1 if any flow uses node $v$. Eq. (4.27) to (4.30) are multi-commodity flow constraints, where Eq. (4.28) to (4.30) enforce the equilibrium of the flow and Eq. (4.27) ensures the capacity constraints are respected. This constraint is different from the classic multicommodity flow problem, in which it would be

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

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. (4.32) ensures that an edge is activated if any commodity uses it and Eq. (4.33) 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. (4.27). 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. (4.27) and to replace it with a weighted sum of those binary variables.

$$\min \quad \sum_{v\in\mathcal{V}} \sum_{i=1}^{in(v)} x_i(v) \tag{4.35}$$

$$\text{s.t.} \quad f(u,v)\cdot d_k \leq c(u,v)\cdot \left(x_1(v) - \sum_{i=2}^{in(v)} \frac{x_i(v)}{(i-1)i}\right) \quad \forall (u,v)\in\mathcal{E} \tag{4.36}$$

$$x_i(v) \geq \left(\sum_{u\in\mathcal{N}_{in}(v)} f(u,v) - (i-1)\right) / in(v) \quad \forall v\in\mathcal{V}, \forall 1\leq i\leq d(v) \tag{4.37}$$

$$x_i(v)\in\{0,1\} \quad \forall v\in\mathcal{V}, \forall 1\leq i\leq d(v) \tag{4.38}$$

(4.28), (4.29), (4.30), (4.32), (4.33), (4.34)

**Theorem 1.** *The BNLP (4.26) - (4.34) has the same optimal solution as the BLP (4.28) - (4.38).*

*Proof.* Let us first observe that $\left(\sum_{u\in\mathcal{N}_{in}(v)} f(u,v) - (i-1)\right)$ 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. (4.37) 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 $(4.28) - (4.38)$, $x_i(v)$ 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. (4.36), 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 (4.27). 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. (4.35), is equivalent to the objective in Eq. (4.26).

□

### 4.4.2 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.

#### 4.4.2.1 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 $\lambda_{\text{gNB}}$. The exact position is computed by relying on the placement optimization techniques described in Sect. 3.2. UEs are then randomly distributed both in public areas, such as streets, and inside buildings. Specifically, given a density of $\lambda_{\text{UE}}$, 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 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 3GPP technical report [126]. In short, we consider that in our simulations 80% of the UEs are placed indoors. Fig. 4.13 shows a deployment with $\lambda_{\text{gNB}} = 45$ and $\lambda_{\text{UE}} = 900$ UE/km$^2$.

#### 4.4.2.2 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 ITU Urban-Micro (UMi) channel model,described in Sect. 2.2.1.3. However, instead of using the stochastic LoS probability model provided by the same UMi model, we deterministically evaluate the LoS by employing visibility analysis on the same 3D models used to find the optimal locations, obtaining a more accurate estimation. For indoor UEs, we always consider them to be 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 [26].

Finally, we compute the 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. 4.3 details all the values used in our simulations, which are aligned with typical literature and 3GPP studies on this topic.

#### 4.4.2.3 Time-varying UE density model

Most studies dealing with topology optimization focus their analysis on a single, or a handful, value of $\lambda_{\text{UE}}$. 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

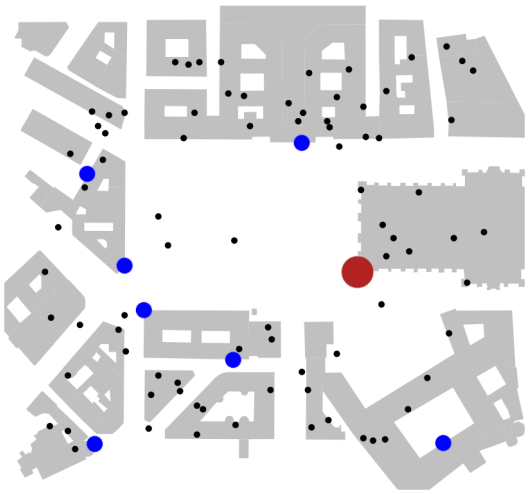


**Figure 4.13:** Sample deployment of a network in the center of Milan, with 1 IAB-donor (in red), 7 IAB-nodes (in blue), and 83 UEs (in black). It corresponds to $\lambda_{\text{UE}}(9) = 900$ UE/km$^2$ (Mon 9am). © 2024 IEEE

of values of UE density, ideally following a realistic trend. Therefore, we employ a technique used in similar research [127] 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 [128]. We then normalize it in the range $(0, 1]$, and we model the UE density as a function 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 [126]. Finally, we generate a set of 168 graphs spanning an average week with a one-hour granularity.

Fig. 4.14 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.

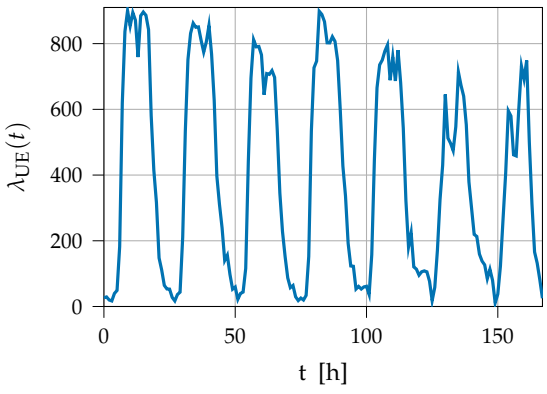



**Figure 4.14:** Weekly profile for the UE density in central Milan. © 2024 IEEE

| 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 4.3:** Simulation Parameters

### 4.4.3 Results

We evaluate our model on a 0,092$km^2$ area in the center of the city of Milan (Italy), for which we computed the UE density trend $\lambda_{UE}$ 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.
- **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.

#### *4.4.3.1 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 4.15 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**.

### *4.4.3.2 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 Fig. 4.16 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

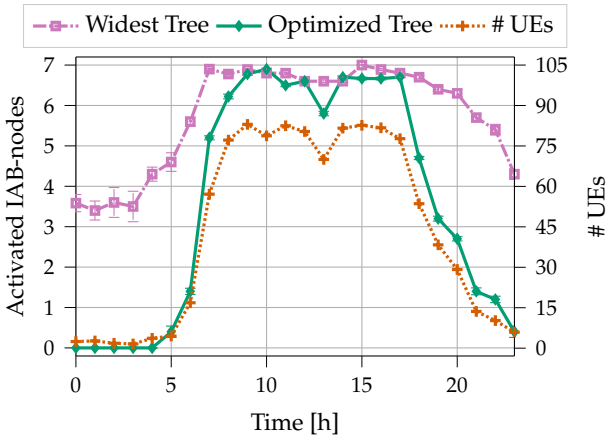



**Figure 4.15:** Number of IAB-nodes activated in the first 24 of the week (Mon).

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. (4.39) 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.

$$\hat{c}_I = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} c(i,j) \tag{4.39}$$

Figure 4.16a 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.

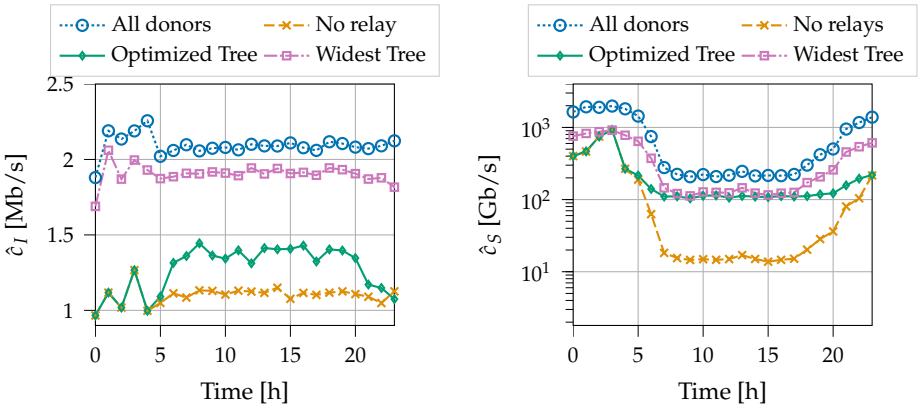
The second metric, called Average Saturation Capacity and detailed in Eq. (4.40), is formulated in a very similar way as Eq. (4.39). However, here we assume that all the UEs try to access the network at the same time, thus we divide the capacity of each edge $c_{s,t}$ by the number of inner neighbors of the node $t$, since those edges share the same resources through the scheduler.

$$\hat{c}_S = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} \frac{c(i,j)}{N_{in}(j)} \tag{4.40}$$

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.

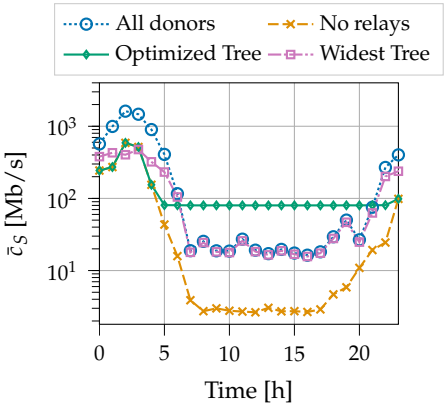
The third metric, called Minimum Saturation Capacity and detailed in Eq. (4.41), measures the capacity delivered to worst UE while the network is under saturation by all the UEs, *e.g.* it defines the minimum level of Quality-of-Service provided by the topology. It is defined similarly to the previous one, but instead of averaging over the UEs we take the worst value.

$$\bar{c}_S = \min_{u \in \mathcal{U}} \min_{(i,j) \in p(u,t)} \frac{c(i,j)}{N_{in}(j)} \tag{4.41}$$




**(a)** Average idle capacity.

**(b)** Average saturation capacity.




**(c)** Minimum saturation capacity.

**Figure 4.16:** Capacity Metrics for the first 24h of the week (Mon).© 2024 IEEE

Fig. 4.16b 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

# 5 WIRELESS BACKHAUL NETWORKS IN RURAL AREAS

The pandemic showed us that a working Internet connection is not *just* an enabler of other human rights, as we already knew, but it is a necessity for many people to study, work, and access basic services. However, 2.9 billion people, almost 40% of the world population did not have access to the Internet in 2021[1], a rate that grows to 61% in rural areas, worldwide. One key reason for this situation is that telecommunication companies have a business model that works best in densely inhabited areas, have large capital expenditures to deploy the infrastructure, and rely on the high number of potential customers to return on the investment. It is an *all-or-nothing* model that fails in regions where the population density is very low, houses are clustered in groups several km apart, and the cost of the *middle-mile* connecting them is very high.

One of the successful instruments that we have to reduce the digital divide consists in Wireless Backhaul Networks (WBNs), a network architecture usually adopted either by Wireless Internet Service Provider (WISP) or by Wireless Community Networks (WCNs) [2]. With this architecture, the customers, or participants, of the network host part of the infrastructure on their premises and leverage wireless technology to cut capital costs related to the deployment of fiber optic, the deployment of trellises where to place antennas, etc.

In this chapter, we focus on analyzing the feasibility and economic sustainability of these models in rural areas [3] through two main pieces of research.

The first research, published in IEEE Network Magazine [3], analyzes the phenomena of WCN from a historical and evolutional perspective. The evolution of WCN is divided into three main generations of networks using different technologies and solving different problems. More focus is given to the 3rd generation of WCN, still in an emerging phase, that we foresee will be using millimeter wave (mmWave) frequencies and massive Multiple Input, Multiple Output (MIMO) antennas to increase its capacity and thus provide a viable alternative in providing multi-gigabit broadband access. Finally, we evaluate the possible coverage of a 3rd generation WCN by adopting open-data methodologies as done in the studies presented in Chapter 3

[1] See ITU Facts and Figures 2021 https://www.itu.int/itu-d/reports/statistics/facts-figures-2021/

[2] Note that even though most WCNs started using only Wireless technology, now many of them have moved to hybrid wireless/wired architecture and are generally referred simply as Community Networks

[3] The definition of rural areas used in this chapter is the same adopted in Chapter 3, which is the definition made by Eurostat.

The second research, presented at the 18th International Conference on Network and Service Management [9] and then extended and submitted to IEEE Transaction on Network and Service Management [4], proposes an open-data-based economical model for WBN in rural areas, estimates the demand for this kind of service and then evaluate both the level of service and the economical sustainability of such model. Additionally, the reliability of the infrastructure is assessed and different strategies to improve its reliability are discussed.

## 5.1 RELATED WORKS

Since low-cost wireless equipment was made available at the beginning of the 2000s, wireless community networks [129] and WISPs [130] proliferated. Rapidly these types of networks attracted the attention of the research community studying a variety of aspects related to routing, scalability, security, measurements, testbeds, topologies, performance, usage patterns, evolution, and mobility. See e.g. the Ph.D. thesis from Neumann [131] and Yahel [132] and the references therein for wireless community networks and WISPs, respectively. Soon these technologies were considered good candidates to provide Internet access to developing and rural regions. In this context, some works related to economy and planning similar to ours can be found. In Maccari et Al. [133] the authors investigate the economic feasibility of the growth of a wireless community network. In Cameron et Al. [134] a WISP backhaul optimization model is formulated in order to minimize energy consumption. Examples of real use cases include Potch et Al. [135] which describes the WISP planning in a rural region of Northern California; Hameed et Al. [136] where the implementation of a WISP in mountainous areas in Pakistan is discussed; and Mubaraq et Al. [137] where there is an economic study of a WISP in a specific district in the state of Kerala in India.

In the context of 5G/6G cellular networks, there is a large number of works dealing with economic and topology planning models [138]. E.g. in Oughton et Al. [139] an optimal network planning and cost assessment tool is developed for 5G networks. In Yaghoubi et Al. [88] an optimization model is used to formulate a backhaul design maximizing reliability for a channel model that includes rain attenuation. However, 5G has a specific focus on increasing user performance in dense areas, but no provision for under-served rural areas [140]. Some effort in this direction is ongoing for 6G [141] but it is far from being a viable present solution.

A promising area of research involved in reducing digital divide revolves around Community Cellular Networks (CCNs), a concept that leverages many concepts used in WCN but to deploy Cellular Networks in underserved areas. In Barela Et Al. [142] [143], the authors document the experience and challenges encountered in deploying a CCN in the Philippines. In Guzman Et Al. [144], a wireless backhaul using Line of Sight (LoS) links, is evaluated as an alternative to satellite backhaul. In

Heimerl Et Al. [145] a system to operate CCNs at large, called CommunityCellularManager, is presented and its impact is evaluated through a multi-year deployment in the Philippines.

Another hot research topic and a promising technology for closing the digital divide is given by low-orbit satellites, as they are capable of providing broadband connection with global coverage at a competitive price. However, their global performance, cost-effectiveness, and manageability are still to be demonstrated, especially in underserved areas with low income. In the near future, the integration of satellite networks and community networks is a promising option, and it is already part of the research agenda [146].

## 5.2 TOWARDS SMART WIRELESS COMMUNITY NETWORKS

This research analyzed the state of WCNs from a technological and research point of view, with three main contributions: first, we summarized the path of WCNs in the last two decades, focusing on the design of two generations of wireless WCNs that were analyzed in the literature, with their pros and cons; Second, we described the emergence of a third generation of WCNs (3G-WCN) which, in addition to the evolution of Wi-Fi, will incorporate recent advances in blockchain and AI, and we discuss the directions in which these technologies can be applied. Third, we argued that WCNs can continue to be a key instrument to overcome the digital divide if they evolve, providing added value to their users beyond connectivity. We outlined the challenging research items that need to be addressed to make WCNs evolve into smart community networks based on past experiences and growing interest in this theme.

As a tangible example, we simulated a 3G-WCN and study its penetration in regions that are now under-served. We used an innovative approach mixing open data and GPU-based visibility analysis to estimate the fraction of households that could be reached using 3G-WCN with state-of-the-art technology. This had a double goal of showing the feasibility of a CN and of introducing new methodologies for high-impact research.

### 5.2.1 From Roofnets to Large Scale WCNs

With the diffusion of IEEE 802.11, the concept of mesh networks was introduced in ICT research. While there is no specific definition, a wireless mesh network is generally described as an infrastructure for which:

- The nodes are stationary (as opposed to a mobile ad-hoc network);
- Nodes are both generators/receivers of traffic and routers of other nodes' traffic;
- There is no planning, and the network is self-healing: nodes can be added or removed, and the protocols must make new nodes reachable and/or *route around* failures without manual reconfiguration.

The first generation of WCNs (1G-WCN) used commercial off-the-shelf (COTS) 802.11b/g access points mounted on house roofs or terraces, as in the MIT Roofnet [147].

The advantage of this design was an extremely simple set-up, with only one device to be mounted and configured. Morover the network was purely non hierarchical with a flat mesh topology in which any node was connected to all the reachable neighbors. However the performance of this architecture was extremely low. Omnidirectional antennas can cover short distances (up to hundreds of meters), and the network capacity

scales sub-linearly with the number of nodes, as the interference is hard to mitigate when the spatial density increases. As soon as 802.11 became popular, the unlicensed bands started to be crowded, and the idea of creating general-purpose, large-scale mesh networks with this design faded.

#### 5.2.1.1 Motivations and Design of a Modern CN

To be usable in under-served areas with low house density, the wireless links of a CN need to span across several km, which encouraged a second generation of community networks (2G-WCN). The single COTS device on the roof is replaced by a set of outdoor ISP-grade point-to-multipoint devices using Power over Ethernet (PoE) and mounted on one (or more than one) pole on the house's roof (as depicted in Fig. 5.1). The devices create links spanning several km, powered by 802.11ac with up to gigabit performance (see Tab. 5.1). They need to be connected to a router that will take care of the routing of packets, plus a home access point to provide connectivity to the terminals inside the house. 2G-WCNs enable networks made of hundreds [92] or even thousands of nodes [148] covering very large regions.

Such a CN is a techno-social artifact: an unplanned distributed network that adapts to the addition of new nodes, mirrored in a distributed community that grows spontaneously one person after the other. The network is still, in principle, non-hierarchical. However, nodes placed on better locations, such as location with great visibility or easy accessibility, quickly becomes more important than others and thus the network often becomes semi-hierarchical, with a two-layer architecture. The owners of the nodes, however, can still organize in agile peer-to-peer communities with very lightweight coordination to organically expand the network at needs. WCNs are generally treated as a not-for-profit initiative held in commons, often with no legal entity, which makes them easy to bootstrap, the exact opposite of the *all-or-nothing* model we mentioned. This design sparked a body of interdisciplinary research in ICT [149], social systems analysis [150], network economy [151], and many more.

However, 2G-WCNs design has three big drawbacks. The first is the complexity of the architecture, which now requires configuring at least three devices with a non-trivial network and physical set-up, as shown in

| gen. | standard 802.11 | antenna type | band GHz | throughput Mbps |
|---|---|---|---|---|
| 1 | b,g | omni. | 2.4 | $< 20$ |
| 2 | n,ac | sector, ptp | 2.4, 5 | 20∼300 |
| 3 | ac,ad,ax | ptp, beamf. | 5, 60 | 100 ∼1000 |

**Table 5.1:** Link characteristics for different generations of WCNs.

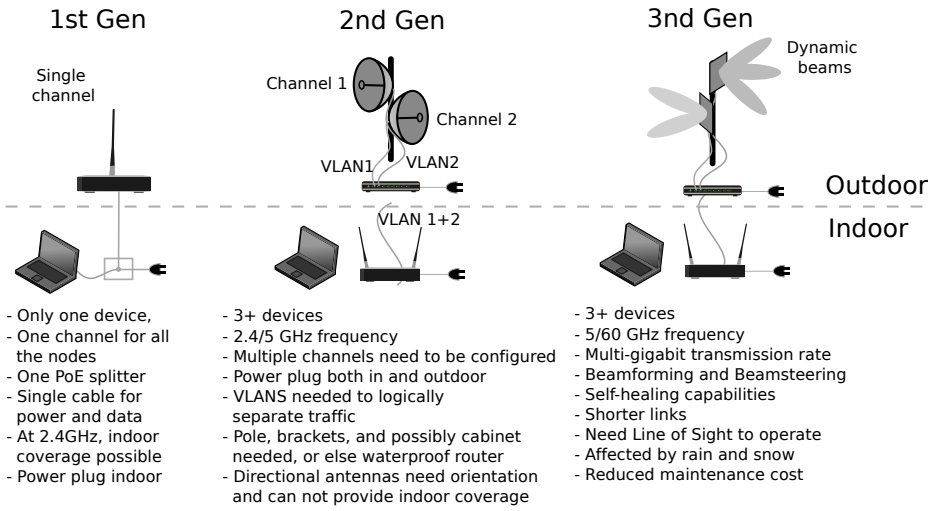



**Figure 5.1:** Comparison between CN generations. © 2023 IEEE

Fig. 5.1. The second is the increased cost, which can reach several hundred euros per node (with a large variability). The third and most critical is the loss of self-healing capacity, as directive antennas need to be re-pointed or added when the topology is modified. With this network design, every time the conditions change (a new link is created or some link fails), a trained person needs to climb onto a roof (in someone else's house), physically modify some existing nodes, and reconfigure the devices. C onsequently, 2G-WCNs strongly rely on the initiative of some technically skilled members of the community to be maintained. Their technical structure tends to centralize because a few, better-equipped nodes are easier to maintain, and the social structure, while still being perceived by the community as a peer-to-peer distributed one, inevitably degenerates into a highly centralized and fragile one depending on a few skilled individuals. When the network grows too much to be maintained by a small voluntary team, its performance and reliability degrade, and the community fades. This is a structural limitation of current WCNs that has been measured analytically in one case [150] and led to the decline of some initiatives[4].

Future WCNs will have to improve in two directions to make the community less dependent on a few individuals: i) improve the manageability; ii) provide added value beyond connectivity to enlarge the user base. We foresee two possible ways to overcome this limitation, which introduces very interesting research challenges. The first is a 3G-WCN design based on new technologies that can make the network easier to maintain. In Sect. 5.2.2, we explain the technological building blocks and provide data to assess its feasibility. The second requires better integration with services that can make WCNs attractive beyond Internet connectivity, provide

[4]See the "Report on the Governance Instruments and their Application to WCNs" of the netCommons Project: https://netcommons.eu/?q=content/deliverables-page.

added value to the participants, and thus create new economic incentives to make WCNs self-sustainable (we cover it in Sect. 5.2.3 and 5.2.4).

### 5.2.2 3rd Generation Community Networks

The advent of 5G encouraged research on communications at high frequencies such as mmWave (beyond 24 GHz) and Terahertz (beyond 1 THz) communications, where there are large unused portions of the spectrum that can be used to provide high throughput links. Moreover, high frequencies need smaller antennas, so we can have Massive Multiple Input Multiple Output (M-MIMO) antenna arrays made of tens of elements to enable beamforming. A MIMO device covering a sector of 120° can create a narrow beam aimed in the direction of the receiver, achieving a transmission (and reception) gain close to the one achievable with a directional antenna and reducing the overall interference. The beam is steerable in a dynamic way, so the advantage of having a point-to-point link does not come with the disadvantage of manual aiming, as with directive antennas. However, at high frequencies, the propagation of signals is challenging. The communications happen primarily in line-of-sight as obstacle penetration is impossible, and fading due to rain, foliage, or even atmosphere is way more impacting than at the unlicensed spectrum used by 802.11n/ac.

While these techniques were primarily intended to improve radio access in mobile 5G networks, mesh networks at high frequencies are the next step. For instance, 5G supports backhauling via mesh networks (referred to as Integrated Access and Backhaul, which is receiving considerable attention from the communications research area). In contrast, in the unlicensed spectrum (60GHz), commercial mesh devices are already available, like the products that support the Facebook/Meta Terragraph mesh network, providing gigabit performance.[5]

Given these recent advances, we can imagine a forthcoming 3G-WCN design using higher frequencies than the classical 2.4/5 GHz spectrum, from 24 GHz up, and based on two building blocks: i) MIMO devices that create links in line-of-sight with steerable beams of length lower than one km to create dense local meshes; ii) a minimal number of long links spanning several km, using directional antennas to connect the otherwise disconnected meshes. 3G-WCN mesh nodes look similar to the second one. Still, the large majority of the antennas do not need to be pointed, and once mounted and configured, they don't need human intervention to create new links, which makes their management way easier than 2G-WCN mesh nodes and recovers the self-healing capacity. In contrast, link length is reduced substantially, and line-of-sight is required. Thus, we must answer the question: is it feasible to achieve widespread population coverage in rural areas with a 3G-WCN?

[5] See https://terragraph.com/.

#### 5.2.2.1 Simulating The Coverage of 3G-WCNs in Rural Areas

To evaluate the possible coverage of 3G-WCNs, we adopt a data-based methodology similar to the one used in the other contribution of this chapter, that can be fully re-used in this, or other contexts. We start from a 3D surface obtained by open data from public administration that represents a certain municipality, we identify buildings using open maps, and we use open demographic data to obtain the number of households for every *census section*, which is a subset of the area of the municipality. Using the 3D shape, we estimate the volume of the buildings, exclude the ones smaller than a certain threshold, and assign to each building a number of households proportional to the volume, up to the number of households in the area. Of the total number $N$ of buildings, we randomly select a fraction $\rho$ of buildings, with a probability distribution that follows the number of households per building. We ideally place an antenna on top of the building roofs, assuming a 2 m tall pole, and we then follow a 2-step heuristic: in step 1, we use visibility analysis to compute the line-of-sight between every couple of buildings in the area, setting a maximum link length of 600 m. We extrapolated this value from the specifications of the Terragraph hardware as published by Facebook/Meta[6], which reports links up to 450 m (600 m) with a throughput of roughly 1 Gb/s (200 Mb/s).

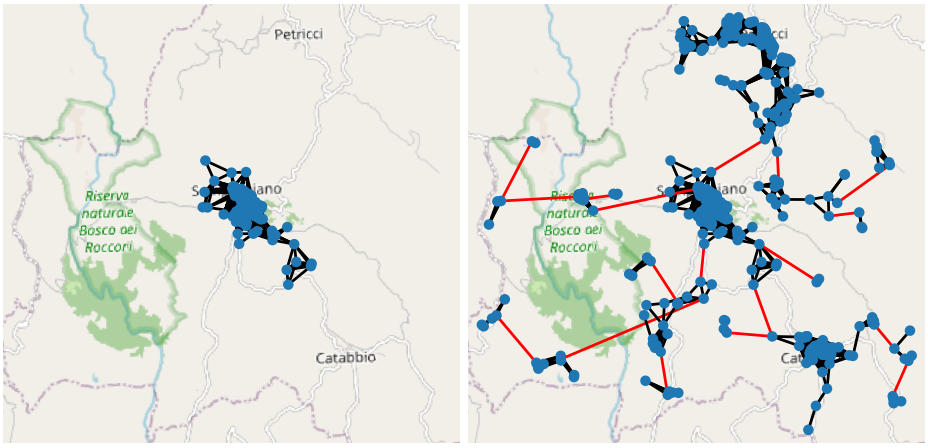
Step 1 creates several disconnected clusters, which we try to connect with longer links. In step 2, for each couple of clusters larger than one node, we search for a feasible line-of-sight link of length up to 4 km. This distance is instead extrapolated by the data-sheet of the point-to-point Ubiquiti airFiber 60XG[7], certified as compatible with Terragraph. We consider the resulting graph, and we count the nodes $M$ in the largest connected component of the network graph. We evaluate a coverage metric $\lambda$ given by $M$ divided by the number of potential nodes in the network ($\lambda = \frac{M}{N\rho}$) that provides an estimation of the coverage of wireless WCNs in the selected areas. If $\lambda = 1$, then all the nodes we randomly chose are part of the connected component of the mesh network, which means that a network connecting all the nodes can be successfully realized. If $\lambda < 1$, then not all nodes can be reached by the mesh network, and it is essential to evaluate how close $\lambda$ is to 1.

We use data from 10 rural areas in Italy that we selected among those considered digitally divided (the fastest wired technology that reaches *some* houses in the area is ADSL with legacy copper cables). Their size and the number of houses is reported in Tab. 5.2. Given the large number of links for which we need to test line-of-sight (up to the order of $10^6$), we use an NVIDIA GPU to parallelize the task.

For each value of $\rho$, and each area, we repeat 30 Monte Carlo simulations with a different choice of the buildings, and we report $\lambda$ Vs $\rho$ in Fig. 5.3.

[6] See the Terragraph Mesh Paper, Throughput Vs Range image, page 11: https://terragraph.com/docs/whitepapers. We use the 99.6 availability curve.

[7] The device is a point-to-point device capable of up to 6Gb/s at 60GHz, see https://dl.ui.com/ds/af60-xg_ds.pdf.




**(a)** short links only **(b)** long+short links

**Figure 5.2:** Two instances of the same network in Semproniano that attempt to connect all the buildings ($\rho = 1$), long links are colored in red. © 2023 IEEE

**Table 5.2:** Summary of data when trying to connect all the buildings in the area ($\rho = 1$, $N$ potential nodes).

| Village | Area size ($km^2$) | # of nodes (N) | # of links (potential) | # of links $< 450$m | # of links $> 600$m |
|---|---|---|---|---|---|
| Borgo a Mozzano | 72.2 | 2174.5 | 99398.4 | 59239.1 (60%) | 3.5 |
| Castel del Piano | 67.8 | 1337.8 | 168956.3 | 96733.3 (57%) | 0.6 |
| Magliano in Toscana | 250.7 | 1270.6 | 30299.3 | 12861.4 (42%) | 23.2 |
| Porcari | 17.9 | 1066.4 | 59083.3 | 30518.5 (52%) | 0.7 |
| Roccalbegna | 124.9 | 482.1 | 17128.7 | 13027.7 (76%) | 0.5 |
| Sambuca Pistoiese | 766.9 | 1449 | 9577.4 | 6413.1 (67%) | 6.5 |
| Santa Fiora | 63.5 | 1383.0 | 91653.3 | 48452.3 (53%) | 4.0 |
| Semproniano | 81.7 | 711.4 | 24483.3 | 14966.1 (61%) | 6.5 |
| Stazzema | 82.1 | 347.3 | 2367.5 | 2056.3 (87%) | 3.8 |
| Villa Basilica | 36.6 | 574.4 | 13962.9 | 11291.8 (81%) | 2.1 |

Every point in Fig. 5.3 reports the average $\lambda$ over 300 runs (10 areas times 30 runs) and the 95% confidence interval. The blue line is the value of $\lambda$ when only step 1 is applied. Even without long links, roughly 50% of the houses can be connected. The red line is obtained with steps 1 and 2. It grows from 66% to 85% and shows that this technology can reach a very large proportion of the population. Note that $\lambda$ grows with $\rho$, so some houses are key to connecting otherwise disconnected areas. With a random choice of the buildings, the probability of choosing these key houses increases with $\rho$ (and thus $\lambda$ increases). Still, with minimal network engineering, we believe $\lambda$ can be increased even for lower values of $\rho$: In a real network, the community may set up nodes in strategic positions, or they may decide to use higher poles in the nodes that can not be reached. However, these simulations confirm that the current technology can enable 3G-WCNs with a widespread population coverage in rural areas. In Fig. 5.2 two instances of a 3G-WCN in Semproniano are shown, highlighting how a small number of longer links can dramatically improve coverage.

Tab. 5.2 provides more details on the areas and on the results for the case $\rho = 1$ and shows that the density of the potential links (those that can be realized in line-of-sight) is extremely high. The average degree ranges from 13 to 252, which provides an extremely high diversity of paths between nodes and thus, resilience to failures. The number of links that can provide gigabit performance is the majority of the potential ones (except in one case), and, most of all, the number of long links that require manual reconfiguration upon failure is extremely low (less than 2 every 100 nodes). We can thus say that 3G-WCNs would be much easier to maintain than 2G-WCNs, and finally more reliable in case of failures.

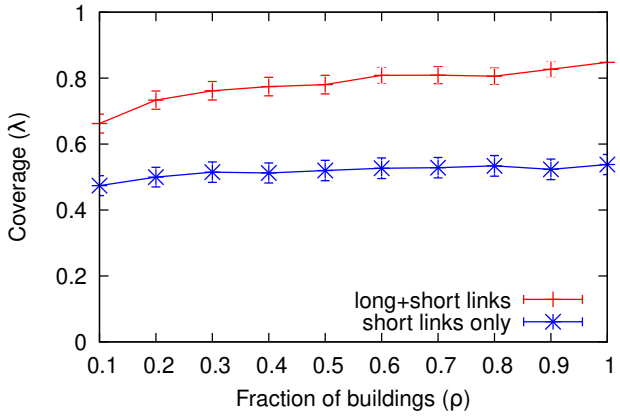



**Figure 5.3:** Coverage of a CN in 10 rural areas of Italy. © 2023 IEEE

### 5.2.3 From WCNs to Smart WCNs

Smart and sustainable cities (or rural communities), as ITU-T defines[8], need interconnected information to understand and control their operations and optimize their limited resources. Information and Communication technologies can improve the quality of life, the efficiency of urban operations and services, and competitiveness while ensuring the needs of present and future generations concerning economic, social, environmental as well as cultural aspects. This is achieved by comprehensive environmental sensing: data on public infrastructures such as energy and water consumption, pollution, weather conditions, security, and safety, combined with data processing to automate well-informed, proactive, and efficient decisions with transparency and accountability, making a community smart. The nodes of WCNs are strategically placed close to data sources and sinks (sensors and actuators on roofs) together with computing nodes with storage capacity (below the roof). This split physical design depicted in Fig. 5.1 is a peculiarity of WCNs that help environmental monitoring practices. However, to support smart applications, WCNs need to provide a faster, more reliable, and more easily manageable network, minimizing the dependency on manual interventions that increase downtime and lower reliability. To make a parallel, we can see 3G-WCN as the innovation brought to mobile networks by 5G. They both improve performance and reliability and enable future added-value applications that are interdependent, co-located, and vertically integrated. Smart applications encourage participation and economic investment, fundamental to making a not-for-profit CN infrastructure sustainable in the long term, especially today that even for traditional telecommunication companies providing bare connectivity is less and less a profitable/sustainable activity.

In this section, we outline some of the promising applications that 3G-WCNs can enable.

**SMART APPLICATIONS FOR WCNS** Organizations and people collaborate in WCNs to achieve a shared goal. Thus, they are generally willing to collect and share their data in a privacy-respecting way to improve the operation and development of their neighborhood. In WCNs, users already have network devices on their roofs and are excellent candidates for data collection and participation in smart city initiatives. Complementary data add detail and quality to environmental sensing, informing city-level and citizen-level decisions with more relevant data. That data pool can feed edge processing for informed and adaptive decisions and actuation. In fact, pooling brings scaling advantages to networking, data storage or computing, and co-location facilitates mutual support for smart decisions. Two examples are the use of data and applications for smart farming, which has been enabled in rural areas thanks to the presence of a CN

[8]See the ITU-T Recommendation "Overview of key performance indicators in smart sustainable cities", https://www.itu.int/rec/T-REC-L.1600-201606-I

[152], or initiatives that use the CN infrastructure to monitor air or noise pollution[9].

**EDGE INTELLIGENCE IN WCNS** Data become actionable when stored and processed. WCNs offer local clouds with storage and processing of large volumes of locally relevant data, accessible at minimal cost and latency, away from the privacy and confidentiality risks of large cloud providers [153]. Tightly coupled with cloud storage is the opportunity of federated machine learning where data is processed near the generation point, and only locally trained models are transferred to a coordinating entity. This procedure has significant advantages over traditional centralized ones, such as data privacy and savings in energy and network bandwidth, and works for both city and citizen-level control decisions and resources. Edge data and intelligence is a key ingredient for smart applications. Even the network infrastructure can benefit from that, allowing more automated self-managing networks. A big challenge is the development of a multi-layer software stack and network architecture, re-usable in different networks, similarly to the Open Radio Access Network (O-RAN) now under development for 5G[10].

**VALUE TRANSACTIONS ON WCNS** The sustainability of a CN depends on its economy, that is, how people share the cost of the infrastructure and its maintenance. Blockchains are suitable instruments to play this role in WCNs as they provide accountability and transparency and are designed to work in communities of people with limited reciprocal trust. Permissioned blockchains, running on a set of small edge nodes, have been deployed and experimented in WCNs as tools for network traffic metering and cost-sharing [154]. Once the Blockchain is in place, the flat and cooperative governance of WCNs suites its usage as a form of alternative currency for reputation and value in exchange b data, work, and smart services.

**WCNS AS ENABLERS OF OTHER NETWORKS** Finally, we mention that 3G-WCNs can become the backbone for other networks as a distributed internet exchange that adds even more value to the community. It is the case with WCNs acting as Internet traffic exchanges, enabling mobile access or IoT applications that need a capillary distribution system.

### 5.2.4 Challenges emerging from Real Cases

CN solutions have open challenges due to their peculiar distributed, bottom-up infrastructure. As such, we need to keep in mind that centralized, top-down solutions for network management and economic

[9] See, for instance, the SEA-HAZEMON project in WCNs in Thailand and South Asia https://interlab.ait.ac.th/HAZEMON/.

[10] 5G is not a monolithic network as previous generations, and its interfaces are open ones, see: https://www.o-ran.org/

sustainability that are widespread in other networks can not be easily applied. In this section, we mention some research challenges that are peculiar to WCNs, coming from the observation of real cases.

**DISINTERMEDIATION:** If a smart CN becomes a key societal asset, the dependability on its operational status increases and thus the governance and management become more complex. The automation and verifiability of an increasing volume of information to deal with become a need to make sure the network operates smoothly. This is the case of network routing and traffic analysis to optimize and deal with traffic anomalies, attacks, or faults. Distributed ledger technologies (DLT) and blockchains bring technological solutions to implement networks that generally exchange verifiable data and value. They allow a trusted and verifiable accounting required to support the representation, management, and transfer of valuable information, such as funds. As a result, smart WCNs can have truthful and verifiable information about service-level agreements, commitments, property, investment, payments, penalties, etc., recorded in a trusted manner and accounted for automatically by code in smart contracts that automatically execute, control or document relevant events and perform actions according to the terms of a contract. [154]. In addition, all these DLT applications can run on servers owned and managed by the smart CN, becoming an economical substrate for further smart CN applications, where network managers and end-users develop economic schemes for resource sharing and service provision and retribution. The scalability of such a system, the kind of blockchain, and how to map network resources to transactions in a verifiable way are still open research issues.

**ECONOMIC:** Once analyzed the feasibility of 3G-WCNs in terms of coverage, economic feasibility comes into play. With today's prices, a mmWave node can cost up to 2000 dollars, but as high-frequency technologies become widespread, we expect the price of devices to reach the price of 2G-nodes, about 200-300 dollars per node. Beyond radio devices, there are several other costs: upstream bandwidth, fixed network routers, and the computing and storage needed for smart applications. When a smart CN becomes a key asset for a whole community, its ecosystem must include individuals, IT companies, small telecom operators, cloud providers, non-IT companies or organizations from any related sector, and public organizations [155]. In the end, a smart community network will thrive if the infrastructure performs well and all providers and consumers have a positive balance in terms of investment and return value, so economic sustainability becomes fundamental.

However, the interplay between common goods and for-profit activities is highly complex to manage. Individuals need incentives for voluntary actions that naturally result in desired collective direction. Commercial activities use profit-oriented economic models to compensate for risk-taking, and the two models may conflict. One viable approach is to consider the

network as a shared infrastructure owned by the community, that develops in a cost-oriented, not-for-profit model, equivalent to the model of Internet Exchange Points [151] and then allows commercial initiatives on the network for services [156]. Again, machine learning and blockchain could be the bridge between the two worlds, as models based on data processing combined with incentive schemes approved by the community can be implemented as smart contracts that, in exchange for actions, result in the automated generation of community currency transfers that can be spent to buy services offered in the network [154].

**TECHNOLOGICAL:** Different technologies offer different sweet spots to satisfy network needs. Therefore, there is a need for a combination of technological solutions, including software tools to manage the integration of traffic management (e.g. routing), data collection and processing, and smart applications on top.

This is particularly true for WCNs that often operate in market failure situations, in which the traditional *one-size-fits-all* connectivity model is not applicable[11]. 3G-WCNs must diversify, combining diverse connectivity technologies with multiple devices and service providers to offer the most cost-effective technology for each location, node, and person. That implies combining wireless, cellular, and wired connectivity [155] while managing location sharing, and more automated interconnection with higher amounts of traffic and economic compensation, for more resilient critical networks and smart applications. In some cases, these technologies are operated by different entities or third parties (including public and private for-profit companies) [156] as local internets. Managing such a heterogeneous network is an open challenge, and the smart aspects bring an additional challenge related to data management, data integration, smart data processing, and data protection. Vendor silos with limited interoperability and lack of standards on 3G-WCNs are limiting factors.

Finally, machine learning is a pervasive technology that is also now being successfully applied in the field of networking. Federated learning, a form of distributed ML, shows it is possible to efficiently utilize network resources, automating management for IoT users and the operating costs for cellular operators [157]. Its application to a distributed network has been tested for traffic classification or anomaly detection, but we need more evidence on its scalability and applicability in real cases [158].

**GOVERNANCE:** As mentioned, WCNs tend to develop in an organic and decentralized manner, developing with little planning compared to most large operator networks. That is similar to how the Internet has developed by setting up local networks that expand and interconnect. WCNs require

[11] One promising large-scale connectivity option is given by low-orbit satellites. However, their global performance, cost-effectiveness, and manageability are still to be demonstrated. Satellite networks are not yet an alternative to CN's communications needs, even if their integration with WCNs is already part of the research agenda. See the recent discussion inside IETF on Satellite-Integrated WCNs, https://datatracker.ietf.org/meeting/114/session/gaia

agreements for inter-operation and standards, similar to the global Internet. Coordinating this evolution while keeping a network operational and avoiding single points of failure [150] is an open challenge.

## 5.3 WIRELESS BACKHAUL DESIGN FOR RURAL NETWORKS

In this study, we investigated the deployment of networks by WISPs in rural areas setting out a detailed economical model considering the cost of the fixed initial assets, the so-called Capital Expenditure (CapEx), and the recurring expenses required to sustain the network, the Operational Expenditure (OpEx). Using open data provided by public administrations we also estimated the demand for connectivity in rural areas, which provided realistic constraints for network design, and the cost of the available connectivity options in these areas.

We focused on a model of a WBN because the main cost of wired connectivity in a rural area is due to the last- and mid-mile between a fiber-connected building (generally in the center of a nearby town) and all the houses that are scattered around the territory. From a fiber-connected location, we modeled the deployment of aerial fiber to a set of gateway nodes, from each gateway we design a multi-hop wireless backbone made of relay nodes that brings connectivity close to the users, and finally each user connects wirelessly to any relay node. We proposed a methodology to choose gateway locations and to plan the network topology. We dedicated particular attention to the robustness of the backbone wireless network, showing that with minimal cost increase it can be made resistant to the failure of several nodes and edges, adding value to the network architecture. We applied our model to ten digitally divided municipalities in central Italy, exploiting the availability of geographical and demographic open data in public repositories.

To the best of our knowledge, this was the first study proposing:

- a detailed technical and economic model for network deployment in rural areas with a mixed wireless/wired approach,
- a methodology to design a reliable backhaul topology and node placement using geographical data, with a realistic estimation of the downtime due to hardware failure,
- the application of our model in a real-case situation based on open data, and available connectivity options.

### 5.3.1 Reliable Wireless Backhaul Design

We briefly describe the design of a router node depicted in Fig. 5.4a, which has an indoor part and an outdoor part. Outdoors there is one pole on which wireless devices (*devices*, from now on) are mounted, these are ISP-grade radio devices that create point-to-point or point-to-multipoint links, which we assume operate in the 5GHz ISM bandwidth with 802.11ac. This configuration has been used in real mesh networks made of hundreds of nodes studied in the literature [133], [159], but our model could be easily modified to use different technologies such as mmWave communications.

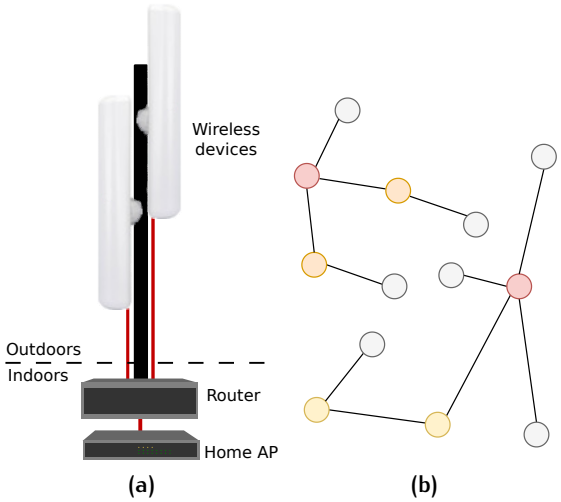



**Figure 5.4:** a) Graphical depiction of a relay node with two radios and one router. b) Example topology made of two gateways (red), four relay nodes (yellow), and seven leaf nodes (white). © 2022 IEEE

A router that takes care of packet routing with some standard routing protocol like Open Shortest Path First (OSPF), Optimized Link State Routing Protocol (OLSR), or even a centralized one based on Software Defined Networking (SDN). Inside the house of the subscribers, there is a simple 802.11 Access Point. If only one wireless device is on the roof, the router is not present.

Let us consider a rural area in which there is a set of households that need connectivity. These households are spread on a set of physical locations (buildings), and each building is a potential node $n$ of the network. The realistic way we choose $\mathcal{V}$ is explained in Sect. 5.3.2.1, for the time being, we can assume $\mathcal{V}$ is given. Exploiting recent ray-tracing techniques and the availability of open data for the terrain, we assess the presence of line-of-sight between the roofs of every couple of nodes (assuming a pole of 2 meters) and create a visibility graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ made of all the nodes and all the potential links. Details on this process can be found in past works that exploit the same data and methodology [160] [10].

For scalability reasons, we can not connect every node to a single gateway, so the network must be a clustered network with one or at most two gateways per cluster. A WBN is thus made of three types of nodes, graphically depicted in Fig. 5.4b. The first type is the gateway node, which is fiber connected to some ISP core network. We call $\mathcal{B}$ the set of gateway nodes. The second type is the relay node, which is used to create a wireless backhaul that brings connectivity close to the users. Relay nodes route the user traffic, and we call $\mathcal{R}$ their set:

$$\mathcal{R} = \{n \in \mathcal{V} \mid \delta(n) > 1 \wedge n \notin \mathcal{B}\} \tag{5.1}$$

where $\delta(n)$ is the degree of node $n$ in the resulting network graph (i.e. the number of neighbors). The last type of node is the leaf node, that is

connected to the backhaul with a single edge, so it does not perform any routing operation. These are defined by the set $\mathcal{L}$:

$$\mathcal{L} = \{n \in \mathcal{V} \mid \delta(n) = 1 \wedge n \notin \mathcal{B}\} \tag{5.2}$$

All nodes provide connectivity to the subscribers that are assigned to the buildings where they are placed.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the visibility graph. The Reliable Wireless Backhaul Design Problem (RWBDB) consists in finding a set $\mathcal{B} \subset \mathcal{V}$ of gateway nodes and a set of edges $\mathcal{E}' \subseteq \mathcal{E}$ that interconnect all the nodes in $\mathcal{V}$ to a gateway in $\mathcal{B}$ with a multi-hop path, providing a configurable resilience to failures.

Ideally, we could formulate an optimization problem that contains all the constraints and finds an optimal solution. However, this approach can scale up to tens of nodes (like in previous works [88]) while in our setting we have thousands of nodes and tens of thousands of potential edges. The aim of our work is not to simplify the problem to make it tractable with some theoretical formulation to achieve some global optimum, nor to introduce new approximation algorithms to marginally improve the algorithmic state of the art. Rather, it is to provide a readily usable solution for network planning and engineering on arbitrarily large networks, to study the feasibility of the mixed mesh/wired approach. Thus, we use state-of-the-art algorithms customized to fit our specific problem and the real data we own, and we make data and source code available for further research that will improve and refine the algorithmic aspects.

We propose a heuristic *divide-et-impera* approach in four steps: i) we partition $\mathcal{G}$ into $k$ clusters; ii) we find the set of gateways $\bar{D}_i$ for each cluster. We consider the base case in which $|\bar{D}_i| = 1$ or a more robust design in which $|\bar{D}_i| = 2$; iii) we elaborate a strategy to select a $\mathcal{E}_i^* \subset \mathcal{E}$ in order to connect each node in the cluster to a gateway; iv) we ensure each cluster has a failure resistant topology. In the following subsections, we will go through the details of each step of the process.

#### 5.3.1.1 Graph Partitioning

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ there are a plethora of clustering algorithms that partition $\mathcal{V}$ in $k$ subsets $\mathcal{V}_1, \ldots \mathcal{V}_k$. In our specific case each cluster corresponds to one or at most two gateways, and considering that the cost of the gateway is a large part of the CapEx due to the fiber connection, we need to minimize the number of clusters. The gateway is connected to the cluster via wireless links, and given the limited amount of devices we can place on a single gateway, the wireless connection of a gateway is the bottleneck of the cluster. Therefore, our design tries to achieve the smallest number of clusters connecting roughly the same amounts of households, in order to obtain the maximum usage of the bottleneck.

Each node in the network can serve more than one subscriber (one node corresponds to a building, and in a building we can have more than one household), thus we need to solve the following problem: Given a graph

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a weight function $s(n)$ that is the number of subscribers in building $n \in \mathcal{V}$, the k-way graph partitioning can be defined as the problem of partitioning $\mathcal{V}$ in $k$ subsets $\mathcal{V}_1, \dots \mathcal{V}_k$ such that:

1. $\cup_i \mathcal{V}_i = \mathcal{V}$
2. $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for $i \neq j$
3. $\sum_{n \in \mathcal{V}_i} s(n) \simeq \frac{1}{k} \sum_{n \in \mathcal{V}} s(n)$
4. the number of edges with end-points in two different subsets is minimized

The well know *METIS* graph partitioner [161] is one of the few algorithms that enables us to have clusters of the same (or at least similar) size. It uses the Kernighan–Lin algorithm [162] with complexity $O(|\mathcal{E}|)$.

After computing all the sets $\mathcal{V}_i$ we extract their associated subgraph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, where $\mathcal{E}_i$ is defined as:

$$\mathcal{E}_i = \{(v_j, v_k) \in \mathcal{E} \mid v_j, v_k \in \mathcal{V}_i\} \tag{5.3}$$

#### 5.3.1.2 Gateway Selection

Given a certain $\mathcal{V}_i$ we need to identify a node, or a group of nodes $\bar{D}_i$ that are in the best position for being the gateways of the cluster. Gateways are fiber connected, thus expensive, and they represent a critical point of failure for all nodes in $\mathcal{V}_i$. We use two designs, a first one with one gateway per cluster and a second with more than one, so if one gateway fails, the nodes in the cluster have a backup. In the paper, we consider the two-gateway design ($|\bar{D}_i| = 2$) but the process can be generalized to more.

Since the hop-count from a node to the gateway influences the end-to-end delay, we choose $\bar{D}_i$ minimizing the average distance to all nodes in $\mathcal{V}_i$. Given a node $y \in \mathcal{V}_i$, the distance of a candidate group $D$ of nodes to $y$ is defined as follows:

$$d(y, D) = \min_{n \in D} d(y, n)$$

Since in our problem each node can represent multiple subscribers we weigh the distance to that specific node $y$ by the number of subscribers $s(y)$ associated with it and we use the concept of Centrality [163], and in particular of *group closeness centrality*. The set $\bar{D}_i$ with the highest group closeness centrality is the group (among all groups $D$ of the chosen dimension) that maximizes the inverse of the sum of the weighted distance to all the other nodes of the network, as follows:

$$\bar{D}_i = \text{argmax}_D \frac{1}{\sum_{y \in \mathcal{V}_i} s(y) d(y, D)}$$

If we want only one gateway per cluster ($|\bar{D}_i| = 1$) computing centrality requires computing the shortest path between any couple of nodes, so

it has polynomial complexity with the size of $\mathcal{V}_i$, and it takes negligible time with thousands of nodes. When $|\bar{D}_i| > 1$ the problem becomes combinatorial and NP, but with two gateways we can still find the optimal solution in minutes.

#### 5.3.1.3 Distribution Tree Design

To connect each node of a cluster to its gateway, a subset of the available edges of $\mathcal{G}_i$ is sufficient. In principle, all edges could be used, but in practice, each node has a limited amount of devices, so economical and technical constraints lead us to minimize the number of edges. At the end of this step we have a set of graphs $\bar{\mathcal{G}}_i = (\mathcal{V}_i, \bar{\mathcal{E}}_i)$ so that the subset of edges $\bar{\mathcal{E}}_i \subseteq \mathcal{E}_i$ creates a Shortest Path Tree (SPT) allowing each node in $\mathcal{V}_i$ to reach a gateway. This problem has to be treated separately for the one gateway or the two gateways scenario.

**TREE DESIGN WITH ONE GATEWAY PER CLUSTER** In this scenario, the optimal tree is the SPT computed with classical Dijkstra's algorithm. We use the link length as the weight for the shortest path, as it is computationally easier to handle, we performed several tests with a weight that is more realistic due to the non-linearity of the path loss of wireless signals, and noticed only minimal differences. Note that the SPT tends to produce many nodes that directly connect to the gateway, which is convenient to reduce the traffic that needs to be relayed on multi-hop paths, thus occupying more resources.

**TREE DESIGN WITH MULTIPLE GATEWAYS PER CLUSTER** The straightforward extension of the previous design would be to compute a multi-rooted SPT, with roots corresponding to the gateways. However, this would not fully exploit the presence of multiple gateways. The failure of a gateway, in fact, would disconnect the leaf nodes attached to it from the wireless backhaul and thus, from the other gateway.

Thus, we use the multi-rooted SPT approach as a benchmark but we propose a multi-gateway design that avoids leaf nodes directly attached to the gateways named Shortest Path without Attached Leaves (SPAL). Given $\mathcal{V}_i$, and a choice of gateways $\bar{D}_i$ we search for a forest $F(\mathcal{V}_i, \bar{\mathcal{E}}_i)$ that minimizes the distance between each node and one gateway, where no edge directly connects leaf nodes to gateways. More formally we search for $\bar{\mathcal{E}}_i$ that minimizes:

$$min \sum_{n \in \mathcal{V}_i} d(n, \bar{D}_i) \tag{5.4}$$

constrained to:

$$(n, d) \notin \bar{\mathcal{E}}_i \; \forall \; n \in \mathcal{V}_i \mid \delta(n) = 1, \; \forall d \in \bar{D}_i \tag{5.5}$$

Through some graph manipulations, SPAL can be reduced to the class of well know Shortest Path with Forbidden Paths (SPFP) [164] problems, specifically to Elementary SPFP (ESPFP) which does not allow cycles and

produces a tree topology [165]. Many algorithms have been proposed for the solution of these problems, for this work we took advantage of an open source implementation of ESPFP from Elkael et Al. [166], which produces the desired minimal subset of edges $\bar{\mathcal{E}}_i$.

Specifically, we use the function $\bar{\mathcal{E}}_i = \text{ESPFP}(\mathcal{V}_i, d, P)$ by Elkael et Al. [166] that solves the generic ESPFP problem. The function takes in input a graph $\bar{\mathcal{G}}_i = (\mathcal{V}_i, \mathcal{E}_i)$, a destination node $d$ and the set of forbidden paths $P$ and yields the set of edges $\bar{\mathcal{E}}_i \subseteq \mathcal{E}_i$ that defines the tree topology. We adapt this function to solve our specific problem in Algorithm 5.1.

We first add to the graph a fictitious destination $d^\star$ connected to each gateway $d \in \bar{D}_i$ (Line 5), so the result will be a single tree we can map to a forest of trees in our graph. Then for each node $n$, we introduce a fictitious leaf node called $n^\star$ connected to $n$ (Lines 8,9). We add to the list of forbidden paths $P$ the paths $\{n^\star, n, d, d^\star\}\ \forall d \in D_i,\ \forall n \in \mathcal{V}_i$ (Line 11). We can then apply ESPFP to this manipulated graph, remove all the fictitious edges from the solution and obtain the solution to our problem (Line 14).

Note that the failure of a gateway may still disconnect some subscribers that are allocated to the gateway itself, or due to the fact that the failure of a gateway may disconnect some relays from the backhaul. This is however mitigated by the graph augmentation described in the next section.

**Algorithm 5.1** Reduction from SPAL to SPFP

**Require:** $\mathcal{G}_i(\mathcal{V}_i, \mathcal{E}_i)$ (Visibility Graph), $\bar{D}_i$ (Gateway Nodes)

1: **procedure** SPAL($\mathcal{G}_i$, $\bar{D}_i$)
2: F = {}
3: $\mathcal{V}_i = \mathcal{V}_i \cup d^\star$ ▷ Create and add the fictitious gateway
4: **for** $d \in \bar{D}_i$ **do**
5: $\mathcal{E}_i = \mathcal{E}_i \cup (d, d^\star)$
6: **end for**
7: **for** $n \in V$ **do**
8: $\mathcal{V}_i = \mathcal{V}_i \cup n^\star$ ▷ Create and add the fict. node
9: $\mathcal{E}_i = \mathcal{E}_i \cup (n, n^\star)$
10: **for** $d \in \bar{D}_i$ **do**
11: $F = F \cup \{n^\star, n, d, d^\star\}$
12: **end for**
13: **end for**
14: $\bar{\mathcal{E}}_i$ =ESPFP$(G, d^\star, F)$ ▷ Solve ESPFP
15: **for** $d \in \bar{D}_i$ **do**
16: $\bar{\mathcal{E}}_i = \bar{\mathcal{E}}_i \backslash \{(d, d^\star)\}$ ▷ Remove the fict. gateway
17: **end for**
18: **for** $n \in V$ **do**
19: $\bar{\mathcal{E}}_i = \bar{\mathcal{E}}_i \backslash \{(n, n^\star)\}$ ▷ Remove the fict. source
20: **end for**
21: **return** $\bar{\mathcal{E}}_i$
22: **end procedure**

#### 5.3.1.4 Graph Augmentation

The previous step produces a backhaul network in which every node has a path to one gateway using the minimal number of edges (a tree). The outcome is a topology in which the failure of a single link close to the gateway could disconnect large portions of a cluster, and even in the case of multiple gateways there is no guarantee that nodes have another path to the other gateways. In order to increase the reliability of the network, additional edges must be added. This is a *Graph Augmentation Problem*, which can be formalized as follows. Given $\bar{\mathcal{G}}_i$ we want a set of edges $\mathcal{E}_i^* \subseteq \mathcal{E}_i$ such that:

- $|\mathcal{E}_i^*|$ is minimal and $\bar{\mathcal{E}}_i \subseteq \mathcal{E}_i^*$
- $\mathcal{G}_i^* = (\mathcal{V}_i, \mathcal{E}_i^*)$ is 2-edge-connected

A 2-edge connected graph tolerates the failure of 1 edge without disconnecting any node. Like other graph combinatorial problems, also this problem has been proven NP-Hard. For this reason, a heuristic is needed to solve it [167]. This heuristic finds an approximated solution $\mathcal{E}_i^*$ with log-linear complexity.

Augmenting the whole wireless backhaul, however, would be too costly, as it would approximately double the number of devices. Moreover, no commercial Internet Service Provider (ISP) guarantees a fault-tolerant connection to its subscribers. For this reason, we augment only the core of the network (which corresponds to the sub-graph made of the relays and gateways) to make it 2-edge-connected. In the scenario with two gateways we first run the ESPFP which creates two trees. Each tree is a separate partition of nodes in $\mathcal{V}_i$, which is augmented. Finally, we connect the two partitions with the shortest link between relay nodes if available, or the shortest link between any two types of nodes.

Note that not every graph can be transformed in a 2-edge connected graph, and that's why in Sect. 5.3.3 we evaluate the robustness of the resulting graphs.

### 5.3.2 Link Dimensioning and Economical Modelling

The process we described so far uses as input an estimation of the demand, in terms of what buildings will be served, and how many subscribers per building are present. Here we describe two original contributions of the paper: how to estimate the demand and how to estimate the cost of creating and operating the network (the CapEx and OpEx, respectively).

The final goal is to compute a monthly cost per subscriber, given by:

$$S_c = \frac{1}{60\, n_s} (C + 5 \cdot O) \tag{5.6}$$

where $C$ is the overall CapEx, which we amortize in 5 years (and thus in 60 months), $O$ is the yearly OpEx and $n_s$ is the number of subscribers:

$$n_s = \sum_{n \in \mathcal{V}} s(n)$$

Note that a 5 years amortization time for a network serving thousands of people is very short compared to the life of competitors' technologies, such as fiber connections. However, wireless technologies are subject to a rapid development and thus, we chose this short and challenging target, assuming that every 5 years the infrastructure needs to be renewed.

#### 5.3.2.1 Demand Model

To construct a realistic backhaul network, we first have to define $\mathcal{V}$ by selecting a subset of the buildings in the analysis area where the subscribers are located. This could be done by randomly selecting a given number of buildings, but it would lead to a selection that is not representative of the geographical distribution of the population. For this reason, we have retrieved the population and household census from the Italian Institute of Statistics (ISTAT), which provides the number of households and inhabitants for each census section. A census section is a polygon of variable dimension, which in the areas under analysis has an average surface of 4.52 km$^2$ and contains on average 70 buildings. Based on this data-set, and the position and size of the buildings extracted from Open Street Map (OSM) we assign to each building a probability $w_i$ of having a household that could be a subscriber of the WISP. This allows us to perform a weighted sampling with replacements of $n_s$ subscribers distributed over $|\mathcal{V}|$ buildings, which are representative of the household distribution, and perform Monte Carlo simulations. Note that the same building can be extracted multiple times, this corresponds to a single building with multiple households ($s(n) > 1$). We refer to the ratio between $n_s$ and the total number of households in the area as the subscriber ratio, which is a configuration parameter. Let $\mathcal{C} = \{c_j\}$ be a set of census section polygons for the municipality under analysis and let $h_j$ be the number of households in the polygon, both obtained by ISTAT. Let $\mathcal{P} = \{p_i\}$ be a set of polygons representing buildings, obtained by the OSM data-set. We use highly precise morphological open-data[12] to compute the volume $v_i$ of each building $p_i$, we remove buildings smaller than 100m$^3$, as they have a low probability of being inhabited.

**MATHEMATICAL FORMULATION FOR $w_i$** Since OSM buildings can span different census sections, we first need to calculate the volume occupied by the building $p_i$ in each census section $c_j$. We call $V_i^j$ the volume of the building $p_i$ in the section $c_j$:

$$V_i^j = \frac{\cap(c_j, p_i)}{|p_i|} \times v_i \tag{5.7}$$

[12] See the LiDAR datasets released by the Italian Ministry of Environment https://www.mite.gov.it

where $\cap(c_j, p_i)$ is a function that returns the area of the intersection of two polygons and $|p_i|$ is the area of the building. Eq. (5.7) takes into account the fact a building may lay across two different census sections. We then normalize $V_i^j$ over the volume of all buildings in the census area, and multiply it for the number of households in the area:

$$H_i^j = \frac{V_i^j}{\sum_{p_k \in \mathcal{P}} V_k^j} \times h_j. \tag{5.8}$$

$H_i^j$ is the number of households we expect to live in building $p_i$ that pertain to the census area $c_j$. We need to sum this value over all the possible census sections, and then normalize again over the total number of households in all census sections. We obtain a set of weights $\{w_i\}$, which represents the probability of having a household in a given building $b_i$ in the municipality under analysis.

$$w_i = \frac{1}{\sum_{c_k \in \mathcal{C}} h_k} \sum_{p_j \in \mathcal{C}} H_i^j \tag{5.9}$$

#### 5.3.2.2 Estimating the CapEx

To compute the CapEx of a WBN we assume that leaf nodes are equipped with a single device, whose cost is $a_l$. Relay nodes and gateway nodes can have multiple wireless devices of the same cost, each one with a beamwidth $\beta$. They are also equipped with a router, whose costs are $r_r$ and $r_{gw}$ respectively. Every node has a fixed cost for the physical installation, which includes also the home router.

Tab. 5.3 reports the cost of the devices. Note that the costs are intentionally higher than the market price as they are intended to include also other accessories such as uninterruptible power supplies and power over Ethernet switches. The costs are extracted from data sheets, works in the literature, and interviews with members of the guifi.net community network based in Catalonia, which offers connectivity in rural areas.

| Symbol | Value | Description | Source |
|---|---|---|---|
| $d_l$ | 300€ | Deployment cost of a leaf node | $*$ |
| $a_l$ | 100€ | Cost for a leaf radio | [35] |
| $d_r$ | 1000€ | Deployment cost of a relay node | $*$ |
| $a_r$ | 200€ | Cost for a single 120° PtMP radio | [35] |
| $r_r$ | 500€ | Cost for a relay router | [168] |
| $d_{gw}$ | 10000€ | Deployment cost of a gateway node | $*$ |
| $r_{gw}$ | 5000€ | Cost of a gateway router | [168] |
| $cap_f$ | 6000€ | Cost to deploy aerial fiber (per km) | [169] |

$*$ Values obtained by interviews, similarly as in [170].

**Table 5.3:** CapEx costs

Given a gateway $b \in \mathcal{B}$ its cost $C_{gw}(b)$ depends on the number of the radio devices $n_a(b)$, times the cost of a radio $a_r$, plus the cost of the router $r_{gw}$ and the installation cost $d_{gw}$:

$$C_{gw}(b) = d_{gw} + r_{gw} + a_r \cdot n_a(b) \tag{5.10}$$

Similarly to the gateway node, the cost $C_r(r)$ of a relay node $r \in \mathcal{R}$ depends on the number of radio devices $n_a(r)$ times the cost of a radio $a_r$ plus some fixed costs for the router and the deployment ($r_r$ and $d_r$):

$$C_r(r) = d_r + r_r + a_r \cdot n_a(r) \tag{5.11}$$

The cost of a leaf node is given simply by the sum of the cost of one radio $a_l$ and the physical deployment $d_l$.

Finally, we estimate the cost for the deployment of aerial fiber to the gateways $C_f(\mathcal{B})$ from the closest Point of Presence (PoP) of some operator. Since traditional operators are present in these areas we assume that the PoP is a point $p_0$ in the center of the municipality. Given the street graph, we compute the Steiner tree connecting all the gateways to the PoP along existing public streets, we sum the length (in km) of every arch of the tree and we multiply it by the cost of a km of aerial fiber $cap_f$. Eq. (5.12) shows the composition of the CapEx of the network which is the sum of the cost of gateways, relays, and leaf nodes plus the cost of the aerial fiber backhaul:

$$C = \sum_{b \in \mathcal{B}} C_{gw}(b) + \sum_{r \in \mathcal{R}} C_r(r) + (d_l + a_l)|\mathcal{L}| + C_f(\mathcal{B}) \tag{5.12}$$

**DETERMINING $n_a()$** Once we have a desired topology, we need to estimate the number of wireless devices per relay in order to satisfy two constraints: the coverage of the neighbor nodes and the overall capacity required to route the traffic. The first number is given by elementary geometrical considerations based on the position of the neighbors, so that if the total angle that must be covered by relay $r$ is $\phi(r)$ then we need at least $\left\lceil \frac{\phi(r)}{\beta} \right\rceil$ devices. The second number requires a more elaborated reasoning. Let us call $sp(r)$ the number of shortest paths that go from a subscriber (there can be more than one subscriber per node) to a gateway and pass through relay $r$. We call $c_{sub}$ the minimum guaranteed capacity (in Mb/s) per subscriber, so that $sp(r)\, c_{sub}$ is the required minimum incoming capacity at $r$. Since traffic is relayed to the gateway the sum of incoming and outgoing required capacity at $r$ is $2\, sp(r)\, c_{sub}$. The capacity of a link depends on the Modulation and Coding Scheme (MCS) negotiated with the other end of the link, for which we assume an average value $c_{ch}$ (in Mb/s). Then, the required number of devices is given by:

$$n_a(r) = \max\left( \left\lceil \frac{2 \cdot sp(r) \cdot c_{sub}}{c_{ch}} \right\rceil, \left\lceil \frac{\phi(r)}{\beta} \right\rceil \right) \tag{5.13}$$

Through Eq. (5.13) our model takes into account two factors: the need to add devices to cover a wider angle, or the need to add devices to provide more capacity in a specific direction.

Note that these additional devices could be point-to-point with a narrow beamwidth, to reduce overall interference. However, this is an approximation of the minimum number of radios given the total capacity needed in the relay, since our planning does not include device orientation. It holds on average, but in the real world, installers may take different decisions.

The number of devices for the gateway is determined by the same Eq. (5.13), with the only difference that the multiplier 2 is removed, as the gateway does not rely on wireless devices but on fiber. We call $n_a = \sum_v n_a(v)$ the total number of wireless devices in the network.

#### 5.3.2.3 Estimating the OpEx

As shown in Eq. (5.14), the yearly OpEx is made of three different parts: $O_w$ is the cost of leasing the needed wholesale capacity at the closest Internet Exchange Point (IXP) (see Waites et al. [171] for a description on the role of IXPs in rural connectivity); $O_t$ is the cost of the transit from the closest PoP of some operator to the IXP; $O_m$ is the cost of maintenance of the backhaul:

$$O = O_w + O_t + O_m \tag{5.14}$$

The basic costs we consider for the calculation are reported in Tab. 5.4. The total capacity that the WISP needs to contract is given by the minimum guaranteed capacity provisioned to each subscriber ($c_{sub}$, in Mb/s), times the number of subscribers ($n_s$). We consider a yearly price for wholesale connectivity given by $op_w$ (see Tab. 5.4) with a minimum unit of 1 Gb/s.

$$O_w = \left\lceil \frac{c_{sub}}{1000} \cdot n_s \right\rceil \cdot op_w \tag{5.15}$$

The cost for the transport of the connectivity from the PoP to the regional IXP equals $op_t^{10}$ if the transport is up to 10 Gb/s or $op_t^{100}$ if it is between 10 and 100 Gb/s. We then have:

$$O_t = \begin{cases} op_t^{10} & \text{if } \frac{c_{sub}}{1000} \cdot n_s < 10\ Gb/s \\ op_t^{100} & \text{otherwise} \end{cases} \tag{5.16}$$

To estimate the yearly maintenance cost of the network, we take into account the failures of both routers and radio devices deployed in the wireless backhaul. For both, we have found realistic mean-time-to-failure (mttf) and mean-time-to-repair (mttr) values, which respectively express the average life of a device and the average time needed to repair/replace it after a failure.

We can calculate the yearly cost of maintenance as the number of devices divided by the mttf (which yields the number of yearly failures) times the cost of the intervention, plus the cost of a new device. This is detailed in Eq. (5.17) and is made of four terms:

$$\begin{aligned} O_m = &\frac{|\mathcal{B}|}{\text{mttf}_r}(\text{mttr}_r \cdot m_u + r_{gw}) + \frac{|\mathcal{R}|}{\text{mttf}_r}(\text{mttr}_r \cdot m_u + r_r) \\ &+ \sum_{r \in \mathcal{R}} n_a(r) \frac{1}{\text{mttf}_a}(\text{mttr}_a \cdot m_u + a_r) + \frac{|\mathcal{L}|}{\text{mttf}_a}(\text{mttr}_a \cdot m_p + a_l) \end{aligned} \tag{5.17}$$

The first term takes into account the failure of gateway routers, which are one per gateway, the following term takes into account the failure of routers on relay nodes, the third term takes into account the failure of radio devices on relays, the last one takes into account the failure of leaf nodes. The reason we separate leaf nodes from relay nodes is that most user contracts allow to delay the technical repair until the next working day, while relay nodes need to be repaired as soon as possible as they can impact many users, moreover, even if we have a fully redundant backhaul network, correlated failures could disconnect large portions of the topology. Thus, we use $m_u$ in Eq. (5.17) for relays and $m_p$ for the leaf nodes.

Some other components of the OpEx could be included to refine the model, such as energy consumption and software licenses. The former is very hard to estimate because we miss the data on energy consumption of commercial hardware. However, since our model uses a precise estimation of the required transmission power, it could be enriched with an experimental campaign to extract these data. The latter is a cost that depends on the chosen technology, we notice that open-source commercial-grade options for network management exist (see for instance the openwisp.io platform), so this cost can be compressed. Other works that focus on cellular networks, including both licenses and energy estimate the OpEx to be 10% of the CapEx [139], and this is in line with the data we report in Sect. 5.3.3.

### 5.3.3 Experiments and Results

We evaluate our model on ten rural municipalities in central Italy. The areas have been chosen, among the ones for which the morphological data were available, for their degree of digital division. In fact, for all the areas the average download speed of traditional (xDSL) broadband connectivity was below 30 Mb/s. The municipalities have on average an area of 83 $km^2$, 1558 households, and 3110 buildings. We assume the use of 802.11ac,

| Symbol | Value | Description | Source |
|---|---|---|---|
| $m_u$ | 200€/$h$ | Unplanned maintenance cost | $*$ |
| $m_p$ | 50€/$h$ | Planned maintenance cost | $*$ |
| $\text{mttf}_r$ | 22.8$y$ | Mean time to failure of a router | [172] [168] |
| $\text{mttf}_a$ | 11.4$y$ | Mean time to failure of a radio | [35] |
| $\text{mttr}_r$ | 2$h$ | Mean time to repair of a router | [172] |
| $\text{mttr}_a$ | 4$h$ | Mean time to repair of a radio | $*$ |
| $op_w$ | 1680€/y | 1Gb/s wholesale at the local IXP | [170] |
| $op_t^{10}$ | 31200€/y | 10Gb/s of transport to local IXP | [173] |
| $op_t^{100}$ | 55200€/y | 100Gb/s of transport to local IXP | [173] |

$*$ Values obtained by interviews.

**Table 5.4:** OpEx costs

| Symbol | Value | Description |
|---|---|---|
| $\beta$ | 120° | Beamwidth of relay antenna |
| $c_{ch}$ | 360 Mb/s | Channel capacity at MCS8 |
| $c_e$ | 0.84 | 802.11ac MAC efficiency [174] |
| $p_m$ | 30 dBm | Maximum EIRP due to regulations |
| $f$ | 5.8Ghz | Tranmission Center Frequency |
| $g_r$ | 19 dBi | Relay antenna gain |
| $g_l$ | 27 dBi | Leaf antenna gain |
| $s_{ch}$ | 40 Mhz | Channel width |
| $c_{sub}$ | 7.2 Mb/s | Minimum Guranteed Capacity |
| $c_s$ | {50,100,200} | Maximum Cluster Size |
| $s_r$ | {0.25,0.5,1} | Subscribers ratio |

**Table 5.5:** Parameter used in the Experiments

as is the newest Wi-Fi standard that is widely supported by commercial hardware. However, any other kind of wireless standards, such as 802.11ay or 802.11ax could be used by modifying the configuration parameters, which are reported in Tab. 5.5. Among them we mention the transmission frequency $f$, the channel bandwidth $s_{ch}$ and the link capacity $c_{ch}$ we used in Eq. (5.13). We consider two kinds of devices: a sectorial antenna with beamwidth $\beta$, gain $g_r$ and cost $a_r$ for the relay nodes; and a more directive antenna with gain $g_l$ and cost $a_l$ for the leaf nodes. For both devices, the maximum transmission power (including the antenna gain) has been set to $p_m$ according to local regulations. Finally, the minimum guaranteed capacity per subscriber ($c_{sub}$) has been set to 7.2 Mb/s as an xDSL offer. The experiments have been executed by varying the cluster size $c_s$ and the fraction of served households $s_r$. Each combination of parameters has been run 50 times in ten different areas with a different random seed and thus a different set of subscribers (500 runs per configuration). Images report 95% confidence intervals as error bars.

This section reports the results with only one gateway per cluster, which helped us to calibrate parameters, Sect. 5.3.4 reports the results with the multi-gateway scenario.

#### 5.3.3.1 CapEx and OpEx

We first show the cost of aerial fiber $C_f(\mathcal{B})$ with respect to the cluster size $c_s$ and the subscribers ratio $s_r$, dividing this metric by the number of subscribers to make it comparable in different configurations:

$$F_c = \frac{1}{n_s} C_f(\mathcal{B}) \tag{5.18}$$

Fig. 5.5a shows that $F_c$ monotonically decreases with both the cluster size $c_s$ and the subscribers ratio $s_r$, as larger clusters require less gateways. Clusters smaller than 200 subscribers are hardly economically sustainable at low penetration, as the CapEx needed to deploy the aerial fiber backhaul alone can reach 500 euros per subscriber. On the other hand, larger clusters

may not scale due to the limited availability of independent channels in the 5 GHz band. Fig. 5.5b shows a boxplot of the number of devices per gateway with $c_s = 200$, which is always below 5 and concentrated on 4 (whiskers are 1st and 99th percentiles). Using 802.11ac this is well below the number of 12 independent channels available at 40 MHz width. Fig. 5.5b also shows the same number for relays, which is generally lower. A very small number of outlier nodes exist with a higher number of devices (less than 0.1% of the relays have more than 3 devices), which is an effect of the automated design process, in the real world these relays would be manually split in more than one node to make them less critical. The small average number of devices per node, and the use of PtP/PtMP links, reduce the chances of saturating the spectrum. For this reason, we then set the maximum number of subscribers per cluster to $c_s = 200$ and from now on we show results for this cluster size, without excluding that in future works we can better model the channel allocation and increase this number.

Fig. 5.6 shows the overall CapEx per subscriber divided by each cost type considered in Sect. 5.3.2.2:

$$C_c = \frac{C}{n_s} \tag{5.19}$$

We observe a similar decreasing trend due to the reduced (relative) cost of the aerial fiber (as the clusters get saturated), and the reduction of the ratio between relays and nodes, as their capacity is better exploited. The total upfront investment fluctuates between 923€ and 539€ per subscriber, a value which could be either advanced by the subscriber or amortized over a fixed amount of years.

Fig. 5.7 shows $S_c$ (as in Eq. (5.6)) which corresponds to an ideal monthly recurring cost for subscribers including OpEx and CapEx amortized over 60 months (5 years). We will compare this figure with existing offers in Sect. 5.3.5.

### 5.3.3.2 *Backhaul Network Reliability*

The graph augmentation process described in Sect. 5.3.1.4 produces a bi-connected backhaul network, resistant to the failure of one edge. However, certain network topologies do not allow a fully bi-connected backhaul, and more than one link can fail at the same time when one radio fails. For this reason, we run experiments to quantify the robustness of the backhaul network to the failure of $j$ radios. As a first step we evaluate the probability of $j$ simultaneous radio failures, assuming that the time to failure (repair) of antennas is exponentially distributed with rate 1/mmtf (1/mttr). The probability of being in state $\pi_j$ (having $j$ failed radios) can be computed as the stationary distribution of the birth/death process in Fig. 5.8, given by:

$$\pi_j = \frac{\frac{\alpha^j}{(n_a-j)!}}{\sum_{j=0}^{n_a} \frac{\alpha^j}{(n_a-j)!}}, \; j = 0, \cdots, n_a \tag{5.20}$$

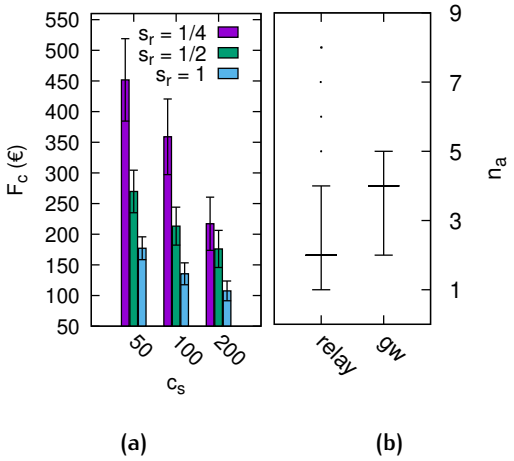



**Figure 5.5:** a) Cost of fiber divided by number of subscribers; b) Boxplot of the number of devices per gateway with $c_s = 200$. © 2022 IEEE

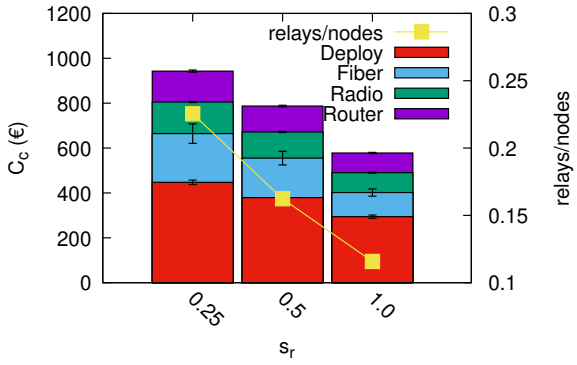



**Figure 5.6:** Overall CapEx when $c_s = 200$ (left axis) and fraction of relays in the network (right axis). © 2022 IEEE

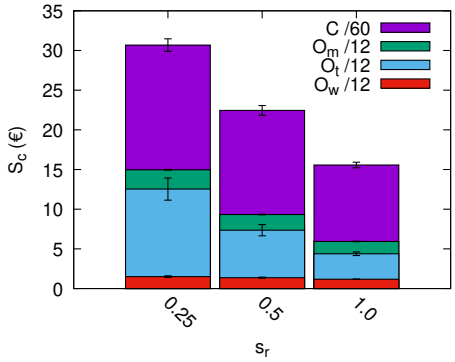



**Figure 5.7:** Lower bound on the subscription price assuming the CapEx gets amortized on 60 months. © 2022 IEEE.

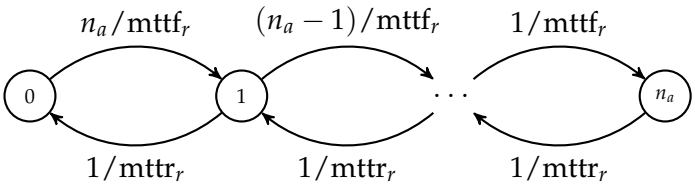



**Figure 5.8:** Markov Chain whose state variable $j$ represents the number of failed radios.

where $\alpha = \frac{\text{mttr}_r}{\text{mttf}_r}$. We consider a very pessimistic scenario (compared to the realistic numbers used in Tab. 5.4): mttf $= 3(y)$ and mttr $= 48(h)$ corresponding to radios that fail every three years and 48 hours to repair one of them, so that $\alpha = 0.00183$.

Fig. 5.9 reports the number of hours per year we expect to be in the state of having exactly $j$ failed radios, for clusters of maximum 200 subscribers. For 97% of the time the network is in state 0 or 1 failures, that do not affect the connectivity of the backhaul. Having more than 4 failures happens less than one hour per year, and thus, we set the maximum $j$ to 4. Given the bi-connected relay graph of a single cluster $\mathcal{G}_i^*$ we remove $j$ radios chosen at random among the ones on the relay nodes. We then remove from the affected relays $\lceil \frac{\delta(r)}{n_a(r)} \rceil$ edges at random for each failed radio and we compute $R_j$: the fraction of subscribers that still have a multi-hop path to the gateway after $j$ failures. Fig. 5.10 then reports $R_j$ averaged on all clusters after 100 repetitions (10 runs for 10 areas). The curves don't start at value one because we take into account the subscribers that don't have line-of-sight with any relay from the very beginning and are unreachable from our network. The trend of the curve is almost linear, indicating that the backhaul is very robust: Every failure disconnects the subscribers that had a direct link towards the failed radio, but there is not a catastrophic effect, typical of graph percolation analysis. This is expected to happen for larger values of $j$, that are so rare that we don't consider them. For 97% of the time the network is very close to its maximum connectivity.

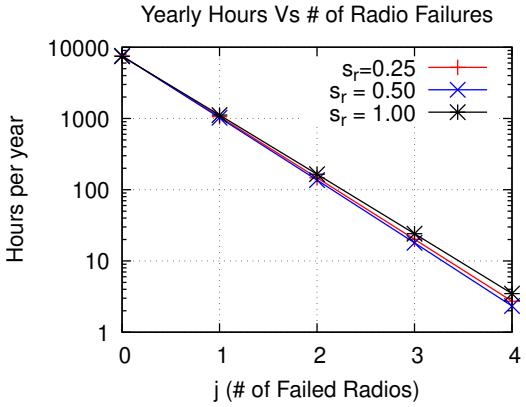



**Figure 5.9:** Yearly hours w.r.t the number of contemporary failed radios.

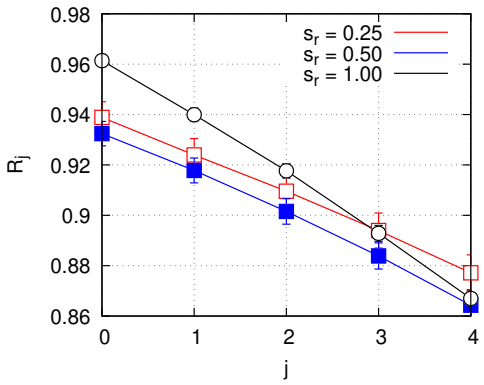



**Figure 5.10:** Ratio of connected subscribers w.r.t the number of failed radios, computed on clusters only on relay nodes

#### 5·3·3·3 *Capacity for Subscribers*

To estimate the performance provided to the subscribers, we first need to compute the maximum capacity of each link. Then, we have to aggregate the capacities onto the path from each subscriber to the nearest gateway to obtain the capacity we can provision to each subscriber. We use two capacity metrics: the average minimum and maximum bit-rate per subscriber ($\bar{c}_{min}$ and $\bar{c}_{max}$, respectively).

**LINK CAPACITY MODELING** Given two nodes we compute the pathloss using the Free Space Path Loss (FSPL) model (note that, as explained in Sect. 5.3.1 our links are deterministically in line-of-sight and we can use higher poles to avoid partial obstruction of the Fresnel zone). We then compute the received power by considering the antenna gains and the power of the devices. Finally, we employ the 802.11ac channel model, described in Sect. 2.2.5.2, that maps the received power to the negotiable bit-rate as per the data sheets of the devices. They range between $bw(-72\,\text{dBm}) = 400\,\text{Mb}/s$ for the coding scheme MCS9 [13] and $bw(-96\,\text{dBm}) = 30\,\text{Mb}/s$ for the coding scheme MCS0, and we multiply this value for a MAC efficiency parameter $c_e = 0.84$ [174]. In our topologies we have 96% of links using MCS9 while the lowest value is given by MCS6 (270 Mb/s), which validates the assumption of using a fixed link capacity $c_{ch} = 360$ Mb/s made in Eq. (5.13).

**CAPACITY PER SUBSCRIBER** In order to compute the capacity metrics from a subscriber $s$ (hosted on node $n_t$) to a gateway $g$, let $n_0, n_1, n_2 \dots n_t$ be a sequence of nodes on the shortest path from $g = n_0$ to the destination node $n_t$ and $sp(n)$ be the number of shortest paths from $g$ to all subscribers that pass through $n$. We call $\Delta(n)$ the set of the neighbors of $n$ with $|\Delta(n)| = \delta(n)$. We call $r(n_i, n_j)$ the negotiated bit-rate on the link $n_i \rightarrow$

[13] We consider a $s_{ch} = 40$Mhz channel width, 2 MIMO streams and guard interval 400 $ns$, as per device specifications. We use datasheets for the Microtik mANTBox 19s and LHG XL5 ac, see https://mikrotik.com/products/group/wireless-systems, accessed April 7th 2023.

$n_j$ (based on the $bw()$ function introduced earlier). Then we derive the average bit-rate per neighbor of node $n$:

$$\hat{r}(n) = \frac{c_e}{\delta(n)} \sum_{n_j \in \Delta(n)} r(n, n_j) \tag{5.21}$$

The minimum bit-rate per shortest path on node $n$ based on the number of devices and the number of paths passing through the node is:

$$r_{min}(n) = \frac{\hat{r}(n) \cdot n_a(n)}{2 \cdot sp(n)} \tag{5.22}$$

Note that in our model we do not assign a specific orientation to the devices, so we can not effectively compute the number of edges per device, and we must rely on an average per node. On the path from $g$ to $s$ if $n = g$ then Eq. (5.22) is modified removing the 2 at the denominator. The minimum capacity for subscriber $s$ is the bottleneck on the path from the gateway:

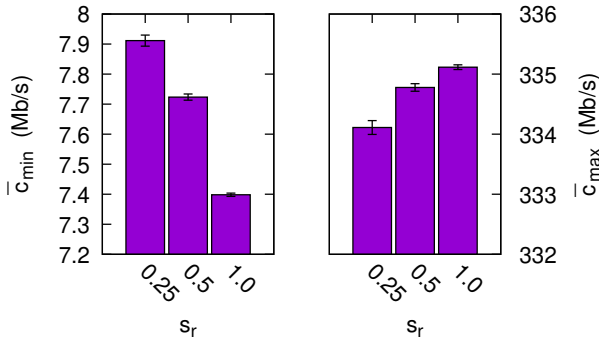
$$c_{\min}(s) = \min_{n_i \in \{n_0 \ldots n_t\}} r_{min}(n_i) \tag{5.23}$$

and $\bar{c}_{min}$ is the average on all subscribers. To compute $c_{max}(s)$ we use:

$$c_{\max}(s) = \min_{n_i \in \{n_1 \ldots n_t\}} r(n_{i-1}, n_i) \tag{5.24}$$

where $\bar{c}_{\max}(s)$ is the average on all subscribers.

**RESULTS** Fig. 5.11a shows $\bar{c}_{min}$, the value is slightly greater than the expected one ($c_{sub}$) as Eq. (5.13) often overprovisions the number of the radios due to geographical constraints or the *upper integer* operator ($\lceil\ \rceil$). Fig. 5.11b shows $\bar{c}_{max}$ and confirms that the maximum capacity we can provide to subscribers is high, as the links are short enough to be efficient. Overall these results confirm our design is sound.




**(a)** Effective average minimum capacity; **(b)** Effective maximum capacity

**Figure 5.11:** Capacity metrics per subscriber.

### 5.3.4 Analysis for multiple gateways

The two gateways topology detailed in Sect. 5.3.1.3 comes at a cost, but it increases the resilience and improves the performance of the network. This section analyzes the effects in terms of Reliability, Cost, and Capacity.

#### *5.3.4.1 Reliability*

To estimate the network robustness we consider the network graph made of all clusters and gateways, we randomly remove $j \in [1,4]$ gateways and we call $R_j$ the fraction of subscribers that still have a multi-hop path to some gateway. We average $R_j$ over 100 repetitions (10 runs with different random seeds, per 10 areas). We compare the single gateway scenario and the multi-gateway using both multi-tree SPT and SPAL in Fig. 5.12.

It is clear that the addition of a second gateway dramatically improves the general reliability of the network. With the failure of one gateway, the single gateway strategy loses 32% of the subscribers, while the multi-gateway strategy with SPT loses only 8% and the SPAL strategy the loss is close to 0. After the failure of additional gateways, $R_j$ maintains the same trend, with a difference of roughly 30% between one and two gateways with SPT and 10% between SPT and SPAL.

#### *5.3.4.2 Capacity*

Fig. 5.13 shows the capacity $\bar{c}_{min}$ (as in Sect. 5.3.3.3) upon the failure of a gateway. The first group of bars reports the average $\bar{c}_{min}$ in a network with two gateways per cluster without failures, to be compared with Fig. 5.11a. Since the bottleneck is given by the wireless access to the gateway reducing the number of subscribers per gateway increases the minimum capacity per subscriber by roughly 50%. The second group of bars reports $\bar{c}_{min}$ when we force the failure of one gateway in each cluster. The average capacity halves, as the network is dimensioned for two gateways and the failure of one triggers re-routing of all the traffic to the other. Still, the reliable design is capable of providing 75% of the capacity compared to a topology that was optimized for one gateway only (see Fig. 5.11a).

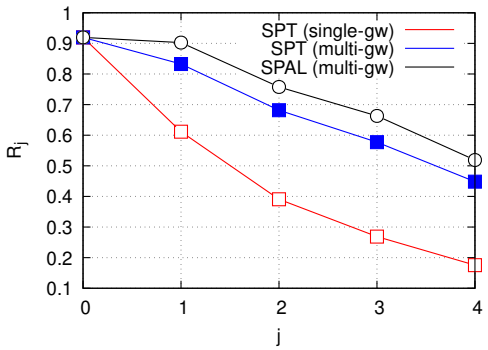



**Figure 5.12:** $R_j$: ratio of connected subscribers w.r.t the number of failed gateways, computed on the whole network

We don't report $\bar{c}_{max}$ since the maximum capacity per subscriber only depends on the slowest link on every path to the gateway, and the multi-gateway design does not practically affect the average value.

#### 5.3.4.3 *Cost*

Doubling the number of gateways clearly comes at a cost in term of both CapEx and OpEx since we account 10000€ for the deployment of each gateway, and more fiber deployment for a network can have up to a dozen gateways in the most populated areas. The most impacted metric is the deployment cost in $C_c$, shown in Fig. 5.14a, which increases from the results reported in Fig. 5.6 by roughly 35%. Other components of the CapEx, such as the Radio, Router and Fiber costs hardly differ from the single gateway design. Finally, Fig. 5.14b shows that the multi-gateway reliable design increases the subscription cost by around 8.20€, 6.9€ and 5.6€ for $s_r$ 0.25, 0.5 and 1, respectively.

3

### 5.3.5 Comparison with Available Offers

We conclude our analysis showing a price comparison with available commercial offers in the areas. Tab. 5.6 shows the comparison of WBN for two values of $s_r$ with the monthly price of xDSL offers [175] and Starlink satellite connectivity[14] [176]. Speed measurements for the xDSL offers come from AGCOM (the Italian Regulator on Communications[15]) while for Starlink they represent the maximum capacity advertised by Starlink itself. Some details are important to notice: with one gateway and 0.25 penetration the cost is roughly 10% below the lowest price, and with 0.5 penetration it is 34% lower. The two gateway design is still convenient with $s_r = 0.5$. This leaves a very high margin to include the interests of a loan (if the WISP can not afford the initial upfront cost) and also some profit if the WISP is not a community initiative but a for-profit one. It is also important to note that existing ISPs have been publicly financed in

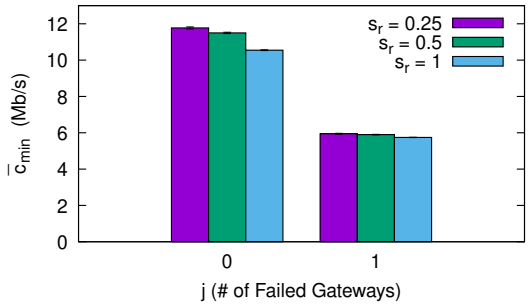
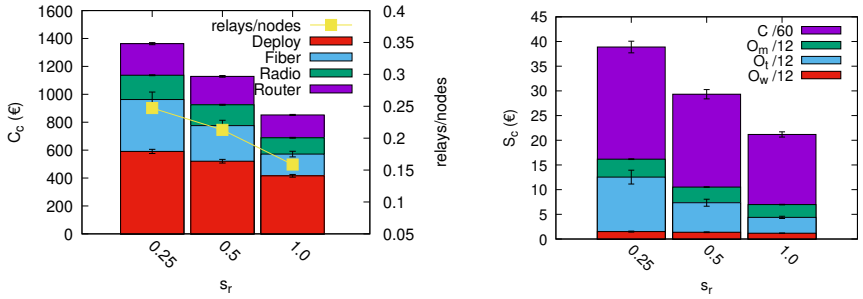



**Figure 5.13:** Effective average minimum capacity for subscriber

[14] Note that during the publication of the manuscript the price for Starlink's equipment and subscription have dropped from 719€ to 300€ and from 99€ to 50€ respectively.

[15] https://maps.agcom.it/




**(a)** Overall CapEx when cs = 200 (left axis) and fraction of relays in the network (right axis)

**(b)** Lower bound on the subscription price assuming the CapEx gets amortized on 60 months

**Figure 5.14:** Cost analysis for a 2-gateway-redundant SPAL topology.

the latest years to extend their infrastructure in remote areas, while we assume the WISP takes all the costs. The cost of the satellite solution is simply incomparable to the other ones.

If we look at the per-Mb/s cost, we also see that the WBN model provides better efficiency than the other models and that the multi-gateway solution offers the best trade-off in terms of minimum guaranteed capacity per user for $s_r = 0.5$. Note however that while our model roughly respects a contention ratio of 50 between maximum and minimum capacity, we do not know the real performance of the other operators. For the xDSL we can only report the minimum negotiated capacity of the market offers available in the area without knowing the real performance and there is no information on the minimum capacity for the satellite offer.

Finally, we report the number of households that we could not connect to the visibility graph and the number of households that are declared impossible to serve by the telecommunication ministry, and we see that even without any specific means to increase penetration (high trellises, or nodes placed on strategic positions to increase coverage) a WBN can serve a higher percentage of the population.

| | WBN | | | | xDSL | SAT |
|---|---|---|---|---|---|---|
| | single-gateway | | multi-gateway | | | |
| | 0.25 | 0.5 | 0.25 | 0.5 | | |
| Upfront cost | 942.9 | 787.2 | 1363 | 1129 | 480 | 324 |
| Recurring cost | 14.9 | 9.3 | 16.2 | 10.5 | 29 | 50 |
| Monthly cost (5y) | 30.7 | 22.5 | 38.8 | 29.3 | 34 | 55.4 |
| Unconnected Households (%) | 8 | 9 | 6 | 7 | 18 | - |
| Max Speed (Mb/s) | 334.1 | 334.8 | 335.5 | 335.8 | 12.6 | 250 |
| Min Speed (Mb/s) | 7.9 | 7.2 | 11.7 | 11.3 | 7.2 | - |
| Monthly Max (€/Mb/s) | 0.09 | 0.07 | 0.11 | 0.09 | 2.6 | 0.4 |
| Monthly Min (€/Mb/s) | 3.9 | 3.1 | 3.3 | 2.6 | 4.7 | - |

**Table 5.6:** Comparison of single and multi gateway designs with $s_r = [0.25, 0.5]$ with available commercial offers.

# 6 CONCLUSIONS

In this dissertation, we delve into the challenges associated with expanding broadband infrastructure in two contrasting yet interconnected settings.

In urban landscapes, the surge in demand for extremely high capacity is pushing the shift towards ultra-dense, high-frequency Radio Access Networks (RANs). Conversely, in rural regions, the extremely low population density hinders the widespread adoption of broadband technology.

In both cases, the deployment costs of fiber backhaul networks pose a significant barrier to the expansion of these technologies. However, Wireless Backhaul Networks (WBNs) present a promising solution. They enable network operators to leverage wireless technologies to expand their backhaul network.

Throughout the preceding chapters, we've demonstrated how to optimize such infrastructures during both the planning and operational stages, utilizing a range of technologies. Foremost among these are Geographic Information System (GIS) technologies. With access to detailed morphological and topological datasets, GIS allows for a comprehensive environmental analysis, pinpointing the best methods to set up and link network elements.

Furthermore, by employing economic models, we've illustrated how these optimization techniques can substantially reduce deployment costs. This is achievable during the network's setup phase through strategic planning and during its operational phase by minimizing recurring expenses, such as energy usage and maintenance costs.

By making this thesis, its accompanying code, and data available under open licenses, we aim to empower the research community and network operators. Our hope is that they will leverage this work to enhance internet access universally, ensuring everyone, everywhere can benefit.

# BIBLIOGRAPHY


[1] G. Gemmi, M. Polese, T. Melodia, *et al.*, "Optimal and dynamic integrated access and backhaul topology adaptation for next-generation cellular networks," in *Submitted: 2024 IEEE International Conference on Computer Communications (INFOCOM)*, Core 2022 A*, 2024 (cit. on p. 103).

[2] G. Gemmi, M. Segata, and L. Maccari, "Estimating coverage and capacity of high frequency mobile networks in ultradense urban areas," *Submitted: Computer Communications*, 2024, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 38).

[3] G. Gemmi, L. Cerdà-Alabern, L. Navarro, *et al.*, "Toward smart community networks," *IEEE Network*, vol. 37, no. 2, pp. 128–134, 2023, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 145).

[4] G. Gemmi, L. Cerdà-Alabern, and L. Maccari, "A robust and cost-effective wireless backhaul design for rural areas," *Submitted: IEEE Transactions on Network and Service Management*, 2023, JCR 2022 Q2, SJR 2022 Q1 (cit. on p. 146).

[5] G. Gemmi, M. Elkael, M. Polese, *et al.*, "Joint routing and energy optimization for integrated access and backhaul with open ran," in *Accepted: 2023 IEEE Global Communications Conference (GLOBECOM)*, Core 2023 B, 2023 (cit. on p. 103).

[6] E. Moro, G. Gemmi, M. Polese, *et al.*, "Toward open integrated access and backhaul with O-RAN," in *21st Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2023 (cit. on pp. 103, 117, 131).

[7] G. Gemmi, M. Segata, and L. Maccari, "Vehicles or pedestrians: On the gNB placement in ultradense urban areas," in *2023 18th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2023, pp. 9–12 (cit. on p. 38).

[8] G. Gemmi, R. L. Cigno, and L. Maccari, "On the properties of next generation wireless backhaul," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 1, pp. 166–177, 2023, JCR 2022 Q1, SJR 2022 Q1 (cit. on p. 38).

[9] G. Gemmi, L. Cerdà-Alabern, and L. Maccari, "A realistic open-data-based cost model for wireless backhaul networks in rural areas," in *18th International Conference on Network and Service Management (CNSM)*, Core 2023 B, 2022, pp. 55–63 (cit. on p. 146).

[10] 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, JCR 2022 Q2, SJR 2022 Q1 (cit. on pp. 37, 118, 126, 161).

[11] G. Gemmi, R. L. Cigno, and L. Maccari, "WIP: Analysis of feasible topologies for backhaul mesh networks," in *2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Core 2021 A/B, 2021, pp. 240–243 (cit. on p. 38).

[12] "Enhanced Throughput for Operation in License-exempt Bands above 45 GHz," *IEEE Std. 802.11ay-2021*, pp. 1–768, 2021. DOI: 10.1109/IEEESTD.2021.9502046 (cit. on p. 1).

[13] 3GPP, "NG-RAN; Architecture Description, Version 17.0.0," 3GPP TS 38.401 (V17.0.0), Tech. Rep., 2021 (cit. on pp. 1, 9).

[14] I. F. Akyildiz, C. Han, Z. Hu, *et al.*, "Terahertz band communication: An old problem revisited and research directions for the next decade," *IEEE Trans. on Communications*, vol. 70, no. 6, pp. 4250–4285, Jun. 2022 (cit. on p. 1).

[15] C. Madapatha, B. Makki, C. Fang, *et al.*, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1374–1389, 2020 (cit. on p. 1).

[16] Small Cell Forum, "Hyperdense HetNets: Definition, drivers and barriers," Tech. Rep., 2017. [Online]. Available: https://scf.io/en/documents/180_Hyperdense_HetNets_Definition_drivers_and_barriers.php (cit. on pp. 1, 41).

[17] 3GPP, "Release 15 Description; Summary of Rel-15 Work Items," 3GPP TR 21.915 (V15.0.0), Tech. Rep., 2019 (cit. on p. 6).

[18] 3GPP, *NR; Backhaul Adaptation Protocol (BAP) specification*, TS 38.340 V17.5.0, 2023 (cit. on p. 8).

[19] M. Polese, L. Bonati, S. D'Oro, *et al.*, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, 2023 (cit. on pp. 9, 110, 111, 117).

[20] A. Kelkar and C. Dick, "Nvidia aerial gpu hosted ai-on-5g," in *2021 IEEE 4th 5G World Forum (5GWF)*, 2021, pp. 64–69. DOI: 10.1109/5GWF52925.2021.00019 (cit. on p. 10).

[21] Intel, *Flexran lte and 5g nr fec software development kit modules*, 2019. [Online]. Available: https://www.intel.com/content/www/us/en/developer/articles/technical/flexran-lte-and-5g-nr-fec-software-development-kit-modules.html (cit. on p. 10).

[22] A. Filali, B. Nour, S. Cherkaoui, *et al.*, "Communication and computation o-ran resource slicing for urllc services using deep reinforcement learning," *IEEE Communications Standards Magazine*, vol. 7, no. 1, pp. 66–73, 2023. DOI: 10.1109/MCOMSTD.0002.2100078 (cit. on p. 10).

[23] L. Bonati, M. Polese, S. D'Oro, *et al.*, "Neutran: An open ran neutral host architecture for zero-touch ran and spectrum sharing," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2023. DOI: 10.1109/TMC.2023.3311728 (cit. on pp. 10, 104).

[24] T. S. Rappaport, *Wireless communications: Principles and practice, 2/E*. Pearson Education India, 2010 (cit. on p. 15).

[25] G. Mao, B. D. Anderson, and B. Fidan, "Path loss exponent estimation for wireless sensor network localization," *Computer Networks*, vol. 51, no. 10, pp. 2467–2483, 2007 (cit. on p. 15).

[26] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report (TR) 38.901, Jun. 2018 (cit. on pp. 16, 126, 137).

[27] ITU, "Guidelines for evaluation of radio interface technologies for IMT-2020," Tech. Rep. M.2412-0, 2017. [Online]. Available: https://www.itu.int/pub/R-REP-M.2412 (cit. on pp. 16, 17, 42, 117, 120, 122).

[28] P. Kyösti *et al.*, "WINNER II channel models," Eur. Commission, Brussel, Belgium, IST-WINNER Tech. Rep. D1.1.2, 2007 (cit. on p. 17).

[29] European Telecommunications Standards Institute, "Study on channel model for frequencies from 0.5 to 100 GHz," European Telecommunications Standards Institute, Sophia Antipolis, France, TR 138 901 V14.0.0, May 2017 (cit. on pp. 17, 69).

[30] A. Al-Hourani, "On the probability of line-of-sight in urban environments," *IEEE Wireless Communications Letters*, vol. 9, no. 8, 2020 (cit. on pp. 17, 118).

[31] H. Ling, R.-C. Chou, and S.-W. Lee, "Shooting and bouncing rays: Calculating the rcs of an arbitrarily shaped cavity," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 2, pp. 194–205, 1989. DOI: 10.1109/8.18706 (cit. on pp. 18, 120).

[32] J. Hoydis, F. A. Aoudia, S. Cammerer, *et al.*, "Sionna rt: Differentiable ray tracing for radio propagation modeling," *arXiv preprint arXiv:2303.11103*, 2023 (cit. on p. 19).

[33] M. Lecci, P. Testolina, M. Polese, *et al.*, "Accuracy versus complexity for mmwave ray-tracing: A full stack perspective," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7826–7841, 2021, ISSN: 1558-2248. DOI: 10.1109/TWC.2021.3088349 (cit. on pp. 19, 121).

[34] J. Pierce, "An introduction to information theory: Symbols, signals & noise. new york: Dover," 1980 (cit. on p. 20).

[35] *Mikrotik wireless devices' prices and datasheets*, Accessed: 2022-04-28. [Online]. Available: https://mikrotik.com/products/group/wireless-systems (cit. on pp. 21, 168, 171).

[36] "22.5 Parameters for VHT-MCSs," *IEEE Std. 802.11ac-2013*, pp. 323–339, 2013. DOI: 10.1109/IEEESTD.2000.92296 (cit. on p. 21).

[37] F. Kaltenberger, A. P. Silva, A. Gosain, *et al.*, "Openairinterface: Democratizing innovation in the 5g era," *Computer Networks*, vol. 176, p. 107 284, 2020, ISSN: 1389-1286 (cit. on pp. 21, 108).

[38] "OpenAirInterface SNR-MCS tables." (), [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/-/tree/develop/openair1/SIMULATION/NR_PHY/BLER_SIMULATIONS/AWGN (visited on 07/31/2023) (cit. on p. 21).

[39] 3GPP, *NR; User Equipment (UE) radio access capabilities*, TS 38.306 V17.5.0, 2023 (cit. on p. 21).

[40] T. Pötsch, S. Yousaf, B. Raghavan, *et al.*, "Zyxt: A network planning tool for rural wireless isps," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 2018, pp. 1–11 (cit. on pp. 29, 34).

[41] V. Gadiraju, H.-C. Wu, C. Busch, *et al.*, "Wireless sensor or access-point deployment using coverage-area maximization over visibility graph," in *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2021, pp. 1–7. DOI: 10.1109/BMSB53066.2021.9547093 (cit. on pp. 29, 34).

[42] K. Kvarfordt, "Planning for Closure of the Logan City/Cache County Landfill and Surrounding Landscape," *Master's Thesis*, 2010 (cit. on pp. 29, 33).

[43] M. W. Lake and P. E. Woodman, "Visibility studies in archaeology: A review and case study," *Environment and Planning B: Planning and Design*, vol. 30, no. 5, pp. 689–707, 2003 (cit. on pp. 29, 33).

[44] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems Journal*, vol. 4, no. 1, pp. 25–30, 1965 (cit. on p. 30).

[45] A. Osterman, L. Benedičič, and P. Ritoša, "An IO-efficient parallel implementation of an R2 viewshed algorithm for large terrain maps on a CUDA GPU," *Int. Jou. of Geographical Information Science*, vol. 28, no. 11, pp. 2304–2327, 2014 (cit. on pp. 30, 31).

[46] S. K. Lam, A. Pitrou, and S. Seibert, "Numba: A LLVM-based Python JIT compiler," in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, New York, NY, USA, 2015, pp. 1–6 (cit. on p. 31).

[47] W. R. Franklin and C. K. Ray, "Higher isn't necessarily better: Visibility algorithms and experiments," in *Advances in GIS Research: Sixth International Symposium on Spatial Data Handling*, Taylor & Francis, 1994, pp. 751–770 (cit. on pp. 31, 88, 89).

[48] M. Van Kreveld *et al.*, "Variations on sweep algorithms: Efficient computation of extended viewsheds and class intervals," 1996 (cit. on p. 31).

[49] T. Lozano-Pérez and M. A. Wesley, "An algorithm for planning collision-free paths among polyhedral obstacles," *Communications of the ACM*, vol. 22, no. 10, pp. 560–570, 1979 (cit. on p. 33).

[50] L. Lacasa, B. Luque, F. Ballesteros, *et al.*, "From time series to complex networks: The visibility graph," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008 (cit. on p. 33).

[51] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—i," *Mathematical programming*, vol. 14, no. 1, pp. 265–294, 1978 (cit. on pp. 34, 43).

[52] P. Slavík, "A tight analysis of the greedy algorithm for set cover," in *Symposium on Theory of Computing (STOC)*, Association for Computing Machinery (ACM), 1996, pp. 435–441 (cit. on p. 34).

[53] Y. Ran, Y. Shi, and Z. Zhang, "Local ratio method on partial set multi-cover," en, *Journal of Combinatorial Optimization*, vol. 34, no. 1, pp. 302–313, Jul. 2017 (cit. on p. 35).

[54] Y. Shi, Y. Ran, Z. Zhang, *et al.*, "Approximation algorithm for the partial set multi-cover problem," *Journal of Global Optimization*, vol. 75, no. 4, pp. 1133–1146, Dec. 2019 (cit. on p. 35).

[55] Y. Ran, Y. Shi, C. Tang, *et al.*, "A primal-dual algorithm for the minimum partial set multi-cover problem," *Journal of Combinatorial Optimization*, vol. 39, no. 3, pp. 725–746, Apr. 2020 (cit. on pp. 35, 45).

[56] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of operations research*, vol. 4, no. 3, pp. 233–235, 1979 (cit. on p. 35).

[57] M. Wright, "Optimization methods for base station placement in wireless applications," in *IEEE Vehicular Technology Conference*, vol. 1, 1998, pp. 387–391 (cit. on p. 39).

[58] M. M. Ahamed and S. Faruque, "5G Network Coverage Planning and Analysis of the Deployment Challenges," *MDPI Sensors*, vol. 21, no. 19, 6698(1–26), 2021 (cit. on pp. 39, 41).

[59] C. K. Anjinappa, F. Erden, and I. Güvenç, "Base Station and Passive Reflectors Placement for Urban mmWave Networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3525–3539, 2021 (cit. on p. 39).

[60] Y. Zhang, L. Dai, and E. W. M. Wong, "Optimal BS Deployment and User Association for 5G Millimeter Wave Communication Networks," *IEEE Trans. on Wireless Communications*, vol. 20, no. 5, pp. 2776–2791, 2021 (cit. on p. 39).

[61] B. B. Haile, E. Mutafungwa, and J. Hämäläinen, "A Data-Driven Multiobjective Optimization Framework for Hyperdense 5G Network Planning," *IEEE Access*, vol. 8, pp. 169 423–169 443, 2020 (cit. on p. 39).

[62] N. Adam, C. Tapparello, W. Heinzelman, *et al.*, "Placement Optimization of Multiple UAV Base Stations," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–7. DOI: 10.1109/WCNC49053.2021.9417488 (cit. on p. 39).

[63] P. Jacquet, D. Popescu, and B. Mans, "Connecting flying backhauls of drones to enhance vehicular networks with fixed 5g nr infrastructure," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops*, 2020, pp. 472–477. DOI: 10.1109/INFOCOMWKSHPS50562.2020.9162670 (cit. on p. 39).

[64] C. K. Anjinappa, F. Erden, and I. Guvenc, "Base Station and Passive Reflectors Placement for Urban mmWave Networks," *IEEE Transactions on Vehicular Technology (TVT)*, vol. 70, no. 4, pp. 3525–3539, Apr. 2021, ISSN: 1939-9359. DOI: 10.1109/tvt.2021.3065221 (cit. on p. 39).

[65] F. Devoti and I. Filippini, "Planning mm-wave access networks under obstacle blockages: A reliability-aware approach," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2203–2214, 2020. DOI: 10.1109/TNET.2020.3006926 (cit. on p. 39).

[66] P. Fiore, E. Moro, I. Filippini, *et al.*, "Boosting 5G mm-Wave IAB Reliability with Reconfigurable Intelligent Surfaces," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 758–763. DOI: 10.1109/WCNC51071.2022.9771934 (cit. on p. 39).

[67] S. Yang, F. Dai, M. Cardei, *et al.*, "On multiple point coverage in wireless sensor networks," in *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference*, 2005 (cit. on p. 39).

[68] M. Farsi, M. A. Elhosseini, M. Badawy, *et al.*, "Deployment techniques in wireless sensor networks, coverage and connectivity: A survey," *IEEE Access*, vol. 7, pp. 28 940–28 954, 2019 (cit. on p. 39).

[69] V. P. Munishwar and N. B. Abu-Ghazaleh, "Coverage algorithms for visual sensor networks," *ACM Transactions on Sensor Networks*, vol. 9, no. 4, 2013 (cit. on p. 39).

[70] S. M. B. Malek, M. M. Sadik, and A. Rahman, "On balanced k-coverage in visual sensor networks," *Elsevier Jou. of Network and Computer Applications*, vol. 72, pp. 72–86, Sep. 2016 (cit. on pp. 39, 45, 47).

[71] F. Yaghoubi, M. Furdek, A. Rostami, *et al.*, "Consistency-Aware Weather Disruption-Tolerant Routing in SDN-Based Wireless Mesh Networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 2, pp. 582–595, 2018 (cit. on pp. 40, 77).

[72] O. Taghizadeh, P. Sirvi, S. Narasimha, *et al.*, "Environment-Aware Minimum-Cost Wireless Backhaul Network Planning With Full-Duplex Links," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2582–2593, 2019 (cit. on p. 40).

[73] Y. Li, A. Cai, G. Qiao, *et al.*, "Multi-Objective Topology Planning for Microwave-Based Wireless Backhaul Networks," *IEEE Access*, vol. 4, pp. 5742–5754, 2016 (cit. on p. 40).

[74] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, *et al.*, "Hybrid Radio/Free-Space Optical Design for Next Generation Backhaul Systems," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2563–2577, 2016 (cit. on p. 40).

[75] M. Polese, M. Giordani, A. Roy, *et al.*, "Distributed path selection strategies for integrated access and backhaul at mmwaves," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, 2018 (cit. on pp. 40, 90, 100, 102).

[76] M. Polese, M. Giordani, T. Zugno, *et al.*, "Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges," *IEEE Communications Magazine*, vol. 58, no. 3, 2020, ISSN: 1558-1896 (cit. on pp. 40, 90, 100, 109).

[77] B. Zhai, M. Yu, A. Tang, *et al.*, "Mesh Architecture for Efficient Integrated Access and Backhaul Networking," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2020 (cit. on p. 40).

[78] M. Simsek, O. Orhan, M. Nassar, *et al.*, "IAB Topology Design: A Graph Embedding and Deep Reinforcement Learning Approach," *IEEE Communications Letters*, vol. 25, no. 2, pp. 489–493, 2021 (cit. on p. 40).

[79] H. Heijmans and C Ronse, "The algebraic basis of mathematical morphology I. Dilations and erosions," *Computer Vision, Graphics, and Image Processing*, vol. 50, no. 3, pp. 245–295, 1990 (cit. on p. 42).

[80] D. S. Hochbaum, "Approximating covering and packing problems: Set cover, vertex cover, independent set, and related problems," in *Approximation algorithms for NP-hard problems*, 1996, pp. 94–143 (cit. on pp. 43, 45).

[81] C. Lund and M. Yannakakis, "On the hardness of approximating minimization problems," *J. ACM*, vol. 41, no. 5, pp. 960–981, 1994 (cit. on p. 43).

[82] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems," DEC Research Report TR-301, Tech. Rep., Sep. 1984 (cit. on p. 47).

[83] E. J. Oughton and T. Russell, "The importance of spatio-temporal infrastructure assessment: Evidence for 5G from the Oxford–Cambridge Arc," en, *Computers, Environment and Urban Systems*, vol. 83, Jun. 2020. (visited on 09/30/2021) (cit. on pp. 48, 52, 53).

[84] K. Haneda and et al., "5G 3GPP-Like Channel Models for Outdoor Urban Microcellular and Macrocellular Environments," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–7 (cit. on pp. 52, 63).

[85] L. Codecá, R. Frank, S. Faye, *et al.*, "Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52–63, Apr. 2017, ISSN: 1939-1390. DOI: 10.1109/mits.2017.2666585 (cit. on pp. 62, 65).

[86] P. A. Lopez, E. Wiessner, M. Behrisch, *et al.*, "Microscopic Traffic Simulation using SUMO," in *21st IEEE International Conference on Intelligent Transportation Systems (ITSC 2018)*, Maui, HI: IEEE, Nov. 2018, pp. 2575–2582, ISBN: 978-1-7281-0323-5. DOI: 10.1109/itsc.2018.8569938 (cit. on p. 65).

[87] G. Gemmi, M. Segata, and L. Maccari, "Vehicles or pedestrians: On the gnb placement in ultradense urban areas," in *IEEE/IFIP Wireless On-demand Network systems and Services Conference*, 2023 (cit. on pp. 66, 71).

[88] F. Yaghoubi, M. Furdek, A. Rostami, *et al.*, "Reliable topology design of wireless networks under correlated failures," in *International Conference on Communications (ICC)*, IEEE, 2018 (cit. on pp. 77, 80, 146, 162).

[89] W. Ellens, F. Spieksma, V. P., *et al.*, "Effective graph resistance," *Linear Algebra and its Applications*, vol. 435, no. 10, pp. 2491–2506, 2011 (cit. on pp. 78, 100).

[90] A. Osterman and P. Ritosa, "Radio Propagation Calculation: A Technique Using 3D Fresnel Zones for Decimeter Radio Waves on Lidar Data," *IEEE Antennas and Propagation Magazine*, vol. 61, no. 6, pp. 31–43, 2019 (cit. on p. 85).

[91] J.-A. A. d'Auriac, N. Cohen, H. El Mafthoui, *et al.*, "Connected tropical subgraphs in vertex-colored graphs," *Discrete Mathematics and Theoretical Computer Science*, vol. 17, no. 3, pp. 327–348, 2016 (cit. on p. 87).

[92] L. Maccari and R. Lo Cigno, "A week in the life of three large wireless community networks," *Ad Hoc Networks*, vol. 24, Part B, 2015 (cit. on pp. 89, 149).

[93] P. V. Trinh, A. T. Pham, A. Carrasco-Casado, *et al.*, "Quantum key distribution over fso: Current development and future perspectives," in *Progress in Electromagnetics Research Symposium*, Tokyo, Japan, 2018 (cit. on p. 93).

[94] L. Bonati, S. D'Oro, M. Polese, *et al.*, "Intelligence and learning in o-RAN for data-driven NextG cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021. DOI: 10.1109/mcom.101.2001120. [Online]. Available: https://doi.org/10.1109%2Fmcom.101.2001120 (cit. on p. 104).

[95] D. Johnson, D. Maas, and J. Van Der Merwe, "Nexran: Closed-loop ran slicing in powder -a top-to-bottom open-source open-ran use case," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization*, ser. WiNTECH '21, New Orleans, LA, USA: Association for Computing Machinery, 2021, 17–23, ISBN: 9781450387033 (cit. on p. 104).

[96] M. Polese, L. Bonati, S. D'Oro, *et al.*, "Colo-ran: Developing machine learning-based xapps for open ran closed-loop control on programmable experimental platforms," *IEEE Transactions on Mobile Computing*, 2022 (cit. on p. 104).

[97] L. Baldesi, F. Restuccia, and T. Melodia, "Charm: Nextg spectrum sharing through data-driven real-time o-ran dynamic control," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 240–249. DOI: 10.1109/INFOCOM48880.2022.9796985 (cit. on p. 104).

[98] M. Polese, V. Ariyarathna, P. Sen, *et al.*, "Dynamic spectrum sharing between active and passive users above 100 ghz," *Communications Engineering*, vol. 1, no. 1, p. 6, 2022 (cit. on p. 104).

[99] E. Amaldi, A. Capone, M. Cesana, *et al.*, "Optimization models and methods for planning wireless mesh networks," *Computer Networks*, vol. 52, no. 11, pp. 2159–2171, 2008, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2008.02.020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128608000935 (cit. on p. 105).

[100] A. Capone and G. Carello, "Scheduling optimization in wireless mesh networks with power control and rate adaptation," in *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, vol. 1, 2006, pp. 138–147. DOI: 10.1109/SAHCN.2006.288418 (cit. on p. 105).

[101] L. Maccari and R. L. Cigno, "Pop-routing: Centrality-based tuning of control messages for faster route convergence," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9. DOI: 10.1109/INFOCOM.2016.7524407 (cit. on p. 105).

[102] B. Zhang and I. Filippini, "Mobility-aware resource allocation for mmwave iab networks via multi-agent rl," in *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, 2021, pp. 17–26. DOI: 10.1109/MASS52906.2021.00012 (cit. on p. 105).

[103] M. Pagin, T. Zugno, M. Polese, *et al.*, "Resource Management for 5G NR Integrated Access and Backhaul: A Semi-Centralized Approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, Feb. 2022 (cit. on p. 105).

[104] M. Polese, M. Giordani, A. Roy, *et al.*, "Distributed path selection strategies for integrated access and backhaul at mmwaves," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7. DOI: 10.1109/GLOCOM.2018.8647977 (cit. on p. 105).

[105] B. Gahtan, R. Cohen, A. M. Bronstein, *et al.*, "Deep reinforcement learning for scheduling and power allocation in a 5g urban mesh," *arXiv preprint arXiv:2210.01423*, 2022 (cit. on p. 105).

[106] R. Mundlamuri, O. Esrafilian, R. Gangula, *et al.*, "Integrated access and backhaul in 5g with aerial distributed unit using openairinterface," *arXiv preprint arXiv:2305.05983*, 2023 (cit. on p. 105).

[107] E. Moro, M. Polese, I. Filippini, *et al.*, "IABEST: An Integrated Access and Backhaul 5G Testbed for Large-Scale Experimentation," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, ser. MobiCom '22, Sydney, NSW, Australia: Association for Computing Machinery, 2022, 772–774, ISBN: 9781450391818. DOI: 10.1145/3495243.3558750. [Online]. Available: https://doi.org/10.1145/3495243.3558750 (cit. on pp. 106, 111).

[108] L. Bonati, M. Polese, S. D'Oro, *et al.*, "OpenRAN Gym: AI/ML Development, Data Collection, and Testing for O-RAN on PAWR Platforms," *Computer Networks*, vol. 220, pp. 1–11, 2023 (cit. on p. 106).

[109] L. Bonati, P. Johari, M. Polese, *et al.*, "Colosseum: Large-Scale Wireless Experimentation Through Hardware-in-the-Loop Network Emulation," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2021, pp. 105–113. DOI: 10.1109/DySPAN53946.2021.9677430 (cit. on pp. 106, 112, 117, 119).

[110] O.-R. WG3, "O-RAN Near-Real-time RAN Intelligent Controller E2 Service Model (E2SM) KPM 2.0," Open-RAN Alliance, Technical Specification (TS), Jul. 2021, Version 2.0.0 (cit. on p. 108).

[111] "F1 interface - Data plane status." (2021), [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/-/wikis/f1-interface (cit. on p. 109).

[112] O-RAN Working Group 10, *O-RAN Operations and Maintenance Interface Specification 8.0*, O-RAN.WG10.O1-Interface.0-v08.00 Technical Specification, 2022 (cit. on p. 110).

[113] D. Villa, M. Tehrani-Moayyed, P. Johari, *et al.*, "CaST: A Toolchain for Creating and Characterizing Realistic Wireless Network Emulation Scenarios," in *Proceedings of the 16th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH '22, Sydney, NSW, Australia: Association for

Computing Machinery, 2022, ISBN: 9781450395274. DOI: 10.1145/3556564.3558243 (cit. on p. 114).

[114] M. Cudak, A. Ghosh, A. Ghosh, *et al.*, "Integrated Access and Backhaul: A Key Enabler for 5G Millimeter-Wave Deployments," *IEEE Communications Magazine*, vol. 59, no. 4, pp. 88–94, 2021. DOI: 10.1109/MCOM.001.2000690 (cit. on p. 115).

[115] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE communications surveys & tutorials*, vol. 18, no. 2, pp. 949–973, 2015 (cit. on p. 115).

[116] P. Fiore, E. Moro, I. Filippini, *et al.*, "Boosting 5G mm-Wave IAB Reliability with Reconfigurable Intelligent Surfaces," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022. DOI: 10.1109/WCNC51071.2022.9771934 (cit. on p. 115).

[117] A. Bavelas, "Communication patterns in task-oriented groups," *The journal of the acoustical society of America*, vol. 22, no. 6, pp. 725–730, 1950 (cit. on p. 116).

[118] S. Y. Jun, C. Lai, D. Caudill, *et al.*, "Quasi-deterministic channel propagation model for 60 ghz urban wi-fi access from light poles," *IEEE Antennas and Wireless Propagation Letters*, vol. 21, no. 8, 2022 (cit. on p. 118).

[119] C. Villien, N. Deparis, V. Mannoni, *et al.*, "Prediction of toa-based localization accuracy using crlb and 3d buildings with field trial validation," in *Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit)*, 2022 (cit. on p. 118).

[120] "Open Street Map 3D buildings." (), [Online]. Available: https://osmbuildings.org/ (cit. on p. 118).

[121] O-RAN Working Group 2, *O-RAN Non-RT RIC Architecture 1.0*, O-RAN.WG2.Non-RT-RIC-ARCH-TS-v01.00 Technical Specification, 2021 (cit. on p. 119).

[122] C. Madapatha, B. Makki, H. Guo, *et al.*, "Constrained Deployment Optimization in Integrated Access and Backhaul Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6. DOI: 10.1109/WCNC55385.2023.10119093 (cit. on p. 120).

[123] J.-H. Kwon, B. Lim, and Y.-C. Ko, "Resource Allocation and System Design of Out-Band Based Integrated Access and Backhaul Network at mmWave Band," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6503–6517, 2022, ISSN: 1939-9359. DOI: 10.1109/TVT.2022.3164751 (cit. on p. 131).

[124] N. Piovesan, D. López-Pérez, A. De Domenico, *et al.*, "Machine learning and analytical power consumption models for 5G Base Stations," *IEEE Communications Magazine*, no. 10, pp. 56–62, 2022 (cit. on p. 134).

[125] 3GPP, "Study on integrated access and backhaul," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.874, Jan. 2019, Version 16.0.0 (cit. on p. 134).

[126] 3GPP, "Study on Scenarios and Requirements for Next Generation Access Technologies," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, Apr. 2022, Version 17.0.0 (cit. on pp. 137, 138).

[127] A. Baiocchi, L. Chiaraviglio, F. Cuomo, *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. DOI: 10.1109/TGCN.2017.2686598 (cit. on p. 138).

[128] G. Barlacchi, M. De Nadai, R. Larcher, *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 (cit. on p. 138).

[129] P. A. Frangoudis, G. C. Polyzos, and V. P. Kemerlis, "Wireless community networks: An alternative approach for nomadic broadband network access," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 206–213, 2011 (cit. on p. 146).

[130] S. Hasan, Y. Ben-David, M. Bittman, *et al.*, "The challenges of scaling WISPs," in *Annual symposium on computing for development (DEV)*, 2015 (cit. on p. 146).

[131] A. Neumann, "Cooperation in open, decentralized, and heterogeneous computer networks," Ph.D. dissertation, Universitat Politècnica de Catalunya (UPC), 2017. [Online]. Available: http://upcommons.upc.edu/handle/2117/114450 (cit. on p. 146).

[132] Y. Ben David, "Connecting the last billion," Ph.D. dissertation, UC Berkeley, 2015. [Online]. Available: https://digitalassets.lib.berkeley.edu/etd/ucb/text/BenDavid_berkeley_0028E_15758.pdf (cit. on p. 146).

[133] L. Maccari, G. Gemmi, R. Lo Cigno, *et al.*, "Towards scalable community networks topologies," *Ad Hoc Networks*, vol. 94, 2019 (cit. on pp. 146, 160).

[134] D. Cameron, A. Valera, and W. K. Seah, "Elasticwisp: Energy-proportional wisp networks," in *Network Operations and Management Symposium (NOMS)*, IEEE/IFIP, 2020 (cit. on p. 146).

[135] T. Pötsch, P. Schmitt, J. Chen, *et al.*, "Helping the lone operator in the vast frontier," in *Workshop on Hot Topics in Networks (HotNets)*, ACM, 2016 (cit. on p. 146).

[136] A. Hameed, A. N. Mian, and J. Qadir, "Low-cost sustainable wireless internet service for rural areas," *Wireless Networks*, vol. 24, no. 5, pp. 1439–1450, 2018 (cit. on p. 146).

[137] S. Mubaraq, J. Hwang, D. Filippini, *et al.*, "Economic analysis of networking technologies for rural developing regions," in *Workshop on Internet Economics (WIE)*, ACM, vol. 10, 2005 (cit. on p. 146).

[138] T. Sharma, A. Chehri, and P. Fortier, "Review of optical and wireless backhaul networks and emerging trends of next generation 5G and 6G technologies," *Transactions on Emerging Telecommunications Technologies*, no. 3, e4155, 2021 (cit. on p. 146).

[139] E. J. Oughton, K. Katsaros, F. Entezami, *et al.*, "An open-source techno-economic assessment framework for 5G deployment," *IEEE Access*, vol. 7, pp. 155 930–155 940, 2019 (cit. on pp. 146, 171).

[140] L. Maccari, M. Karaliopoulos, I. Koutsopoulos, *et al.*, "5g and the internet of everyone: Motivation, enablers, and research agenda," in *2018 European Conference on Networks and Communications (EuCNC)*, 2018 (cit. on p. 146).

[141] A. Chaoub, M. Giordani, B. Lall, *et al.*, "6g for bridging the digital divide: Wireless connectivity to remote areas," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 160–168, 2022 (cit. on p. 146).

[142] M. C. Barela, M. T. Cunanan, P. Martinez, *et al.*, "Community cellular networks in the philippines: Challenges and opportunities towards sustainability," in *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*, 2019, pp. 24–29 (cit. on p. 146).

[143] N. Keleher, M. C. Barela, J. Blumenstock, *et al.*, "Connecting isolated communities: Quantitative evidence on the adoption of community cellular networks in the philippines," in *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, ser. ICTD '20, Guayaquil, Ecuador: Association for Computing Machinery, 2020, ISBN: 9781450387620 (cit. on p. 146).

[144] M. C. Barela, M. T. Cunanan, P. Martinez, *et al.*, "Community cellular networks in the philippines: Challenges and opportunities towards sustainability," in *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*, 2019, pp. 24–29 (cit. on p. 146).

[145] S. Hasan, M. C. Barela, M. Johnson, *et al.*, "Scaling community cellular networks with community cellular manager," in *Proceedings of the 16th USENIX Conference on Networked Systems Design and Implementation*, ser. NSDI'19, Boston, MA, USA: USENIX Association, 2019, 735–750, ISBN: 9781931971492 (cit. on p. 147).

[146] P. Hu, "Closing the management gap for satellite-integrated community networks: A hierarchical approach to self-maintenance," *IEEE Communications Magazine*, vol. 59, no. 12, 2021 (cit. on p. 147).

[147] J. Bicket, D. Aguayo, S. Biswas, *et al.*, "Architecture and evaluation of an unplanned 802.11b mesh network," in *International Conference on Mobile Computing and Networking (MobiCom)*, 2005 (cit. on p. 148).

[148] P. A. Frangoudis, G. C. Polyzos, and V. P. Kemerlis, "Wireless community networks: An alternative approach for nomadic broadband network access," *IEEE Communications Magazine*, vol. 49, no. 5, 2011 (cit. on p. 149).

[149] B. Braem, C. Blondia, C. Barz, *et al.*, "A Case for Research with and on Community Networks," *SIGCOMM Computer Communications Review*, vol. 43, no. 3, 2013 (cit. on p. 149).

[150] L. Maccari, "Detecting and mitigating points of failure in community networks: A graph-based approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, 2019 (cit. on pp. 149, 150, 159).

[151] Y. Zhang, C. Ji, N. Qiao, *et al.*, "Distributed Pricing and Bandwidth Allocation in Crowdsourced Wireless Community Networks," *To appear in IEEE Transactions on Mobile Computing*, 2022. DOI: 10.1109/TMC.2022.3174000 (cit. on pp. 149, 158).

[152] A. Pilichos, M. Karaliopoulos, and I. Koutsopoulos, "From community networks to community data: The applea farming mobile app," in *European Conference on Networks and Communications (EUCNC)*, 2018 (cit. on p. 156).

[153] M. Selimi, A. M. Khan, E. Dimogerontakis, *et al.*, "Cloud services in the Guifi.net community network," *Computer Networks*, vol. 93, 2015 (cit. on p. 156).

[154] E. Dimogerontakis, L. Navarro, M. Selimi, *et al.*, "Contract networking for crowdsourced connectivity," in *2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, 2020 (cit. on pp. 156–158).

[155] Prieto-Egido, Ignacio and Simó-Reigadas, Javier and Castro-Barbero, Eva and Tacas, River Quispe, "Expanding Rural Community Networks Through Partnerships with Key Actors," in *International Conference on Human-Computer Interaction*, 2022 (cit. on pp. 157, 158).

[156] L. Cerdà-Alabern, R. Baig, and L. Navarro, "On the guifi.net community network economics," *Computer Networks*, vol. 168, 2020, ISSN: 1389-1286 (cit. on p. 158).

[157] A. N. Mian, S. Waqas Haider Shah, S. Manzoor, *et al.*, "A value-added IoT service for cellular networks using federated learning," *Computer Networks*, vol. 213, 2022, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2022.109094 (cit. on p. 158).

[158] F. Freitag, P. Vilchez, L. Wei, *et al.*, "Poster: Testbed in wireless city mesh network with application to federated learning experiments," in *ACM International Conference on the Internet of Things (IoT)*, 2021 (cit. on p. 158).

[159] L. Maccari and R. Lo Cigno, "A week in the life of three large wireless community networks," *Ad Hoc Networks*, vol. 24, Part B, pp. 175–190, 2015 (cit. on p. 160).

[160] G. Gemmi, R. Lo Cigno, and L. Maccari, "On the properties of next generation wireless backhaul," *IEEE Transactions on Network Science and Engineering*, 2022 (cit. on p. 161).

[161] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998 (cit. on p. 163).

[162] B. W. Kernighan and S. Lin, "An efficient heuristic procedure for partitioning graphs," *The Bell System Technical Journal*, vol. 49, no. 2, pp. 291–307, 1970 (cit. on p. 163).

[163] N. Mark, *Networks: an introduction*. Oxford University Press New York, 2010 (cit. on p. 163).

[164] D. Villeneuve and G. Desaulniers, "The shortest path problem with forbidden paths," *European Journal of Operational Research*, vol. 165, no. 1, pp. 97–107, 2005, ISSN: 0377-2217 (cit. on p. 164).

[165] L. Di Puglia Pugliese and F. Guerriero, "Shortest path problem with forbidden paths: The elementary version," *European Journal of Operational Research*, vol. 227, no. 2, pp. 254–267, 2013, ISSN: 0377-2217 (cit. on p. 165).

[166] M. Elkael, A. Araldo, H. Castel-Taleb, *et al.*, "An exact algorithm to solve multiobjective, multi-constrained shortest path problems with forbidden paths," 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4413972 (cit. on p. 165).

[167] S. Khuller and R. Thurimella, "Approximation Algorithms for Graph Augmentation," *Journal of Algorithms*, vol. 14, no. 2, pp. 214–225, 1993 (cit. on p. 166).

[168] *Mikrotik routers' prices and datasheets*, Accessed: 2022-04-28. [Online]. Available: https://mikrotik.com/products/group/ethernet-routers (cit. on pp. 168, 171).

[169] I. Chlamtac, A. Gumaste, and C. A. Szabó, Eds., *Broadband Services: Business Models and Technologies for Community Networks*. John Wiley & Sons, Ltd, 2005. (visited on 04/13/2022) (cit. on p. 168).

[170] L. Cerdà-Alabern, "On the Guifi.net community network economics," *Computer Networks*, vol. 168, no. 4, pp. 3581–3606, 2018. (visited on 05/14/2020) (cit. on pp. 168, 171).

[171] W. Waites, J. Sweet, R. Baig, *et al.*, "RemIX: A Distributed Internet Exchange for Remote and Rural Networks," in *Workshop on Global Access to the Internet for All (GAIA)*, ACM, 2016 (cit. on p. 170).

[172] S. Verbrugge, D. Colle, M. Pickavet, *et al.*, "Methodology and input availability parameters for calculating OpEx and CapEx costs for realistic network scenarios," *Journal of Optical Networking*, vol. 5, no. 6, pp. 509–520, 2006 (cit. on p. 171).

[173] *Xarxa oberta fiber transport service*, Accessed: 2022-05-03. [Online]. Available: https://web.archive.org/web/20220503174738/https://xarxaoberta.cat/services/transparent (cit. on p. 171).

[174] E. H. Ong, J. Kneckt, O. Alanen, *et al.*, "Ieee 802.11 ac: Enhancements for very high throughput wlans," in *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, IEEE, 2011 (cit. on pp. 172, 176).

[175] *Trasparenza tariffaria delle offerte di linea fissa*, Accessed: 2022-08-29. [Online]. Available: https://web.archive.org/web/20220829180318/https://www.tim.it/assistenza/trasparenza-tariffaria/trasparenza-tariffaria-delle-offerte-di-linea-fissa (cit. on p. 179).

[176] *Order starlink*, Accessed: 2022-06-03. [Online]. Available: https://www.starlink.com/ (cit. on p. 179).