A Realistic Open-Data-based Cost Model for Wireless Backhaul Networks in Rural Areas

Gabriele Gemmi
Università Ca' Foscari Venezia
Universitat Politècnica de Catalunya
gabriele.gemmi@unive.it

Llorenç Cerdà-Alabern
Universitat Politècnica de Catalunya
llorenc.cerda@upc.edu

Leonardo Maccari
Università Ca' Foscari Venezia
leonardo.maccari@unive.it

***Abstract*—Broadband Internet provision is an increasing demand in many rural areas and wireless internet service providers have emerged as an opportunity to fill this need. However, this type of operator typically consists of a small business with little resources, and difficulty to plan and assess a reliable and economincally sustainable infrastructure. In this paper, we try to bring some aid to this challenging problem by describing a reliable mesh-based backhaul design, together with a detailed CapEx/OpEx economic assessment. We apply our model using real data from ten Italian rural municipalities. Our numerical results show that having clusters of 200 subscribers, a reliable backhaul could be deployed with a monthly subscription and price per Mb/s extremely competitive compared to existing market offers.**

***Index Terms*—WISP, WBN, Digital Divide, Wireless Backhaul, Resiliency, Economic Modeling**

I. Introduction

The need for broadband Internet access in rural areas is even stronger than in urban areas, due to the general lack of services, and this has been exacerbated by the COVID pandemic with the increase of remote working. However, the sparsity and the low density of population reduce the margins for telecommunications operators, slowing down the deployment of optical fiber connections in these areas and encouraged the creation of new grassroots operators targeted to fill this gap in a do-it-yourself fashion. These initiatives take advantage of low-cost wireless equipment to build network access, while a combination of optical fiber/wireless links has been used for the network infrastructure. In some cases, non-profit wireless community networks have been created and deployed directly by their users [1], while in other cases Wireless Internet Service Providers (WISPs) — typically small businesses with reduced margin profits and the aim to provide this important service to their community [2] — provided connectivity.

In this paper, we investigate the deployment of networks by WISPs in rural areas, focusing on both the economical and technological aspects, since these facets are intimately related. We set 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 Operative Expenditure (OpEx). Both aspects need to be considered since some features may affect with opposite signs the CapEx and the OpEx. Using open data provided by public administrations we also estimate the demand for connectivity in rural areas, which provides realistic constraints for network design, and the cost of the available connectivity options in these areas.

We focus on a model of a Wireless Backhaul Network (WBN) because the main cost of wired connectivity in a rural area is due to the 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 that point, we deploy aerial fiber to a set of gateway nodes, while a wireless backbone made of relay nodes will bring connectivity close to the users, that connect wirelessly to any relay node. We propose a methodology to choose gateway locations and to plan the network topology. We apply our backhaul 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 is the first paper proposing:

- a detailed technical and economic model for network deployment in rural areas with a mixed wireless/wired approach,
- a methodology for a reliable backhaul topology design and node placement using geographical data,
- the application of our model in a real-case situation based on open data, and available connectivity options.

For the sake of reproducibility and extension, all our data and source code are freely accessible. [1]

II. Related Work

Since low-cost wireless equipment was made available at the beginning of the 2000s, wireless community networks [3] and WISPs [4] 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 [5] and [2] 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

[1] https://github.com/UniVe-NeDS-Lab/ODCM

 DOI 10.23919/CNSM55787.2022.9964562

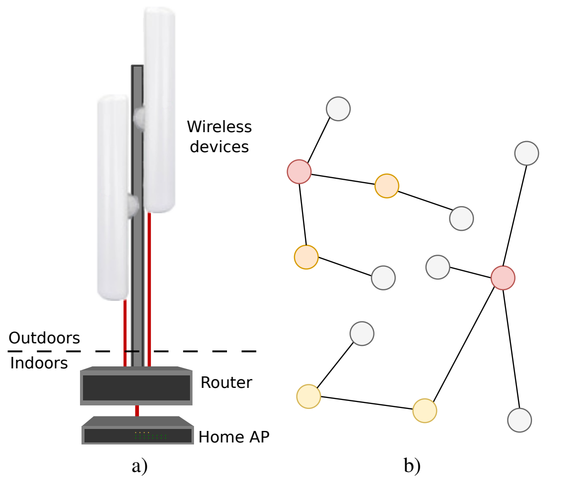


Fig. 1: a) Graphical depiction of a relay node with two radios and one router. b) Example topology made of two gateways (red), four relay nodes (yellow), and seven leaf nodes (white).

related to economy and planning similar to ours can be found. In [6] the authors investigate the economic feasibility of the growth of a wireless community network. In [7] a WISP backhaul optimization model is formulated in order to minimize the energy consumption. Examples of real use cases include [8] which describes the WISP planning in a rural region of Northern California; [9] where it is discussed the implementation of a WISP in mountainous areas in Pakistan; and [10] where there is an economic study of a WISP in a specific district in the state of Kerala in India. These studies, however, differ from ours in the simplicity of their economic model, which does not consider user demand, detailed CapEx/OpEx aspects, or the network backhaul design, which does not include network reliability planning.

In the context of 5G/6G cellular networks there is a large number of works dealing with economic and topology planning models [11]. E.g. in [12] an optimal network planning and cost assessment tool is developed for 5G networks. In [13] 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 the user performance in dense areas, but no specific provision for under served rural areas [14]. Some effort is this direction is being made on 6G [15] but of course it is far from being a viable present solution.

III. Reliable Wireless Backhaul Design

We briefly describe the design of a router node depicted in Fig. 1a, 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 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 [6], [16], but our model could be easily modified to use different technologies.

When there are multiple devices, they are connected through a router that takes care of the delivery of packets with some standard routing protocol, and in the user house, 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 $\mathcal{V}$ of physical locations (buildings), the realistic way we choose $\mathcal{V}$ is explained in Sect. IV-A. For the time being, we can assume $\mathcal{V}$ is given. Exploiting recent ray-tracing advances, 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. We do not consider non-line-of-sight links for performance and reliability reasons, and as the link length is limited and the frequency is centered around 5GHz, we also do not consider the obstruction of the Fresnel zone, that could be alleviated using poles of different heights depending on the link, this is left for future refinements of the model.

For scalability reasons, we can not connect every node to a single gateway, so the network must be a clustered network with one gateway per cluster. A WBN is thus made of three types of nodes, graphically depicted in Fig. 1b. 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, that is used to create a wireless backhaul that brings the connectivity close to the users. Relay nodes route the user traffic, and we call $\mathcal{R}$ their set:

$$\mathcal{R} = \{n \in \mathcal{V} \mid \delta(n) > 1 \wedge n \notin \mathcal{B}\} \quad (1)$$

where $\delta(n)$ is the degree of node n in the 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 $\mathcal{L}$:

$$\mathcal{L} = \{n \in \mathcal{V} \mid \delta(n) = 1 \wedge n \notin \mathcal{B}\} \quad (2)$$

Note that we distinguish between relays and leaf nodes based on their degree, not on their functional role. Both kinds of nodes can provide connectivity to the households of the buildings where they are placed.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the visibility graph. The Reliable Wireless Backhaul Design Problem (RWBDP) consists in finding a set $\mathcal{B} \subset \mathcal{V}$ of $k = |\mathcal{B}|$ gateway nodes and a set of edges $\mathcal{E}' \subseteq \mathcal{E}$ that interconnect all the nodes $v \in \mathcal{V}$ to a specific gateway $b_i \in \mathcal{B}$ with a multi-hop path, providing a configurable resilience to failures.

Ideally, we could formulate an optimization problem that contains all the constraints and find an optimal solution. However, this approach can scale up to tens of nodes (like in previous works [13]) while in our setting we have thousands of nodes and tens of thousands of potential edges. The goal of our paper is not to simplify the problem to make it tractable with some theoretical formulation that allows to reach some global optimum, or to introduce new approximation algorithms to marginally improve the algorithmic state of the art, but to provide a readily usable solution for network planning and engineering in arbitrarily large graphs, 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 three steps: First, we partition $\mathcal{G}$ into k clusters; then we find the best gateway b_i for each cluster; and finally, we elaborate a strategy to select a $\mathcal{E}_i^* \subset \mathcal{E}$ in order to connect each node in the cluster to its gateway b_i. In the following subsections, we will go through the details of each step of the process.

A. 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 $\mathcal{V}_1, \dots \mathcal{V}_k$ based on the maximization of some topological metric. In our specific case each cluster corresponds to a gateway, and the cost of the gateway is a large part of the CapEx due to the fiber connection, which leads us to the choice to minimize the number of gateways and thus, saturate their wireless capacity, that is the network bottleneck.

Each node in the network can serve more than household (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 households corresponding to 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) $\sum_{n \in \mathcal{V}_i} s(n) = \frac{1}{k} \sum_{n \in \mathcal{V}} s(n)$
4) the number of edges with end-points in two different subsets is minimized

The well know *METIS* graph partitioner [17] is one of the few algorithms that enables us to have clusters of same (or at least similar) size. It uses the Kernighan–Lin (KL) algorithm [18] 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:

$$\mathcal{E}_i = \{(v_j, v_k) \in \mathcal{E} \mid v_j, v_k \in \mathcal{V}_i\} \tag{3}$$

B. Gateway Selection

Given a certain $\mathcal{V}_i$ we need to identify a node that is suitable for being the gateway of the cluster. Centrality is a concept used in graph theory to identify nodes that have *central* properties in a graph, with respect to some metric [19].

Since the link-length influences the end-to-end delay and the so-called time-tot-the-first-byte, the optimal gateway should be the node that has the minimum average distance to all the others. This is expressed by the closeness centrality metric [19], defined as the reciprocal of the sum of the length (in hops) of the shortest path between that node and all the other ones. More formally, given a graph $\mathcal{G}$, a node $n \in \mathcal{V}$, the closeness centrality can be defined as:

$$C(n, \mathcal{G}) = \frac{1}{\sum_{t \in \mathcal{G}} d(n, t)} \tag{4}$$

where $d(n, t)$ represents the length of the shortest path between n and t. We then select the gateway $b_i \in \mathcal{V}_i$ as $b_i = \arg\max_{n \in \mathcal{G}_i} C(n, \mathcal{G}_i)$.

Computing the centrality of nodes requires computing the shortest path between any couple of nodes, so it is polynomial with the size of $\mathcal{V}$ and it takes negligible time with thousands of nodes.

C. 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, economical and technical constraints lead us to minimize the number of edges chosen. We want the tree with the minimal number of edges that guarantees the minimal shortest path to the gateway, that is a Shortest Path Tree (SPT) that can be computed with classical Dijkstra's algorithm in polynomial time. We use the link length as the weight for the shortest path, as it is computationally easier than to handle.

At the end of this phase 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 an SPT over $\mathcal{V}_i$ rooted at the gateway b_i.

D. Graph Augmentation

The SPT illustrated in the previous section produces a backhaul network that is not resilient to failures, in fact, the failure of a single link close to the gateway could disconnect large portions of a cluster. 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 proved NP-Hard. For this reason, a heuristic is needed to solve it [20]. 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 edges. 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 relays and gateway of the cluster) to make it 2-edge-connected.

IV. Economical Modelling

The process of network design and cost estimation can be split into three different steps. In the first one, we take open data describing the building and population distribution in a certain rural municipality. From these data we build a feasible demand, that is a percentage of buildings that will be part of the network. The position of the buildings is key to determining if they are reachable via wireless links, as we assume line-of-sight communications. The second step is the design of the network topology as we already described. Finally, once the network topology has been produced we can estimate the cost of setting up and operating the network (the CapEx and OpEx, respectively). This section describes the first and third steps, that are an original methodological contribution of this paper. The final goal is to compute a monthly cost per subscriber, given by:

$$S_c = \frac{1}{60\, n_s}\,(C + 5 \cdot O) \tag{5}$$

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 amortizement 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.

A. Demand Model

In order to build a realistic backhaul network we first need to define $\mathcal{V}$ selecting a subset of the buildings in the area of analysis where the subscribers are. 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 OpenStreetMap (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. We refer to the ratio between n_s and the total number of households as the subscriber ratio, which is a configuration parameter. The details of this process are in Appendix A.

Symbol	Value	Description	Source
d_l	300€	Deployment cost of a leaf node	$*$
a_l	100€	Cost for a leaf radio	[21]
d_r	1000€	Deployment cost of a relay node	$*$
a_r	200€	Cost for a single 120° PtMP radio	[21]
r_r	500€	Cost for a relay router	[22]
d_{gw}	10000€	Deployment cost of a gateway node	$*$
r_{gw}	5000€	Cost of a gateway router	[22]
cap_f	6000€	Cost to deploy aerial fiber (per km)	[23]

$*$ Values obtained by interviews.

TABLE I: CapEx costs

B. 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 β. 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. I 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, power over Ethernet switches, and redundancy when needed. The costs are extracted from data sheets and works in the literature, and interviews with members of the guifi.net community network based in Catalonia, which offers connectivity in rural areas.

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}:

$$C_{gw}(b) = d_{gw} + r_{gw} + a_r \cdot n_a(b) \tag{6}$$

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):

$$C_r(r) = d_r + r_r + a_r \cdot n_a(r) \tag{7}$$

The cost of a leaf node is given simply by the sum of the cost for 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, we sum the length (in km) of every arch of the tree and we multiply it by the cost of a km of areal fiber cap_f. Eq. (8) 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 fiber backhaul:

$$C = \sum_{b \in \mathcal{B}} C_{gw}(b) + \sum_{r \in \mathcal{R}} C_r(r) + (d_l + a_l)|\mathcal{L}| + C_f(\mathcal{B}) \tag{8}$$

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 on 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:

$$n_a(r) = \max\left(\left\lceil \frac{2 \cdot sp(r) \cdot c_{sub}}{c_{ch}} \right\rceil, \left\lceil \frac{\phi(r)}{\beta} \right\rceil\right) \tag{9}$$

Through Eq. (9) 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 devices with a narrow beamwidth, in order to reduce overall interference. Note however that this is an approximation of the minimum number of radios given the total capacity needed at the relay, as our planning does not include the orientation of the devices. 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. (9), with the only difference that the multiplier 2 is removed, as the gateway does not relay on wireless devices but on fiber. We call $n_a = \sum_v n_a(v)$ the total number of wireless devices in the network.

C. Estimating the OpEx

As shown in Eq. (10), the yearly OpEx is made of three different parts. The first (O_w) is the cost of leasing the needed wholesale capacity at the closest Internet Exchange Point (IXP) (see Waites et al. [24] for a description on the role of IXPs in rural connectivity), the second is the cost of the transit from the closest PoP of some operator, to the IXP (O_t), the third is the cost of maintenance of the backhaul (O_m):

$$O = O_w + O_t + O_m \tag{10}$$

The basic costs we consider for the calculation are reported in Tab. II. 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. II) with a minimum unit of 1 Gb/s.

$$O_w = \left\lceil \frac{c_{sub}}{1000} \cdot n_s \right\rceil \cdot op_w \tag{11}$$

Symbol	Value	Description	Source
m_u	$200€/h$	Unplanned maintenance cost	$*$
m_p	$50€/h$	Planned maintenance cost	$*$
mttf_r	$22.8y$	Mean time to failure of a router	[25] [22]
mttf_a	$11.4y$	Mean time to failure of a radio	[21]
mttr_r	$2h$	Mean time to repair of a router	[25]
mttr_a	$4h$	Mean time to repair of a radio	$*$
op_w	1680€/y	1Gb/s wholesale at the local IXP	[26]
op_t^{10}	31200€/y	10Gb/s of transport to local IXP	[27]
op_t^{100}	55200€/y	100Gb/s of transport to local IXP	[27]

$*$ Values obtained by interviews.

TABLE II: OpEx costs

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:

$$O_t = \begin{cases} op_t^{10} & \text{if } \frac{c_{sub}}{1000} \cdot n_s < 10\ Gb/s \\ op_t^{100} & \text{otherwise} \end{cases} \tag{12}$$

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. (13) and is made of four terms:

$$\begin{aligned} O_m = {} & \frac{|\mathcal{B}|}{\text{mttf}_r}(\text{mttr}_r \cdot m_u + r_{gw}) + \frac{|\mathcal{R}|}{\text{mttf}_r}(\text{mttr}_r \cdot m_u + r_r) \\ & + \sum_{r \in \mathcal{R}} n_a(r) \frac{1}{\text{mttf}_a}(\text{mttr}_a \cdot m_u + a_r) + \frac{|\mathcal{L}|}{\text{mttf}_a}(\text{mttr}_a \cdot m_p + a_l) \end{aligned} \tag{13}$$

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 relay 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. (13) for relays and m_p for the leaf nodes.

V. Experiments and Results

As already mentioned in Sect. I 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

Symbol	Value	Description
β	120°	Beamwidth of relay antenna
c_{ch}	360 Mb/s	Channel capacity at MCS8
c_e	0.84	802.11ac MAC efficiency [28]
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 III: Parameter used in the Experiments

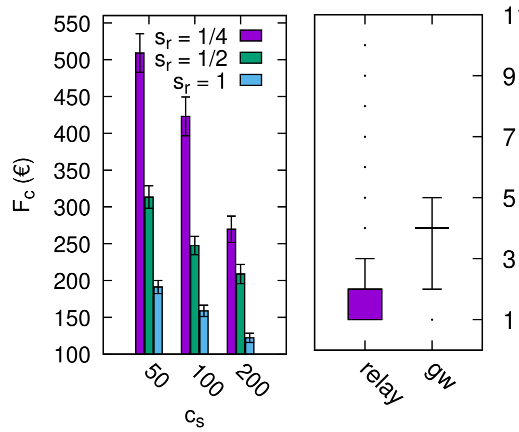
municipalities have on average an area of 83 km^2, 1558 households, and 3110 buildings. We assume the use of 802.11ac, as is the newest Wi-Fi standard that is widely supported by commercial hardware. However, any other kind of wireless standards, such as 802.11ay or 802.11ax could be used by modifying the configuration parameters, which are reported in Tab. III. Among them we mention the transmission frequency f, the transmission bandwidth s_{ch} and the channel capacity c_{ch} we used in Eq. (9). We consider two kinds of devices: a sectorial antenna with beamwidth β, 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 (500 runs per configuration). Images report 95% confidence intervals as error bars.

A. CapEx

We first show the cost of 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:

$$F_c = \frac{1}{n_s} C_f(\mathcal{B}) \quad (14)$$

Fig. 2a 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 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. 2b 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. 2b 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



a) b)

Fig. 2: a) Cost of fiber divided by number of subscribers; b) Boxplot of the number of devices per gateway with $c_s = 200$.

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 show results for this cluster size, without excluding that in future works we can better model the channel allocation and increase this number.

Fig. 3 shows the overall CapEx per subscriber divided by each cost type considered in Sect. IV-B:

$$C_c = \frac{C}{n_s} \quad (15)$$

We observe a similar decreasing trend due to the reduced (relative) cost of the 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.

B. Capacity for Subscribers

To estimate the performance provided to the subscribers, we first need to compute the maximum capacity of each link. Given two nodes we compute the received power in dBm using the Friis free space path loss equation with the correct power and antenna gain (note our links are deterministically in line-of-sight and we can use higher poles to avoid partial obstruction of the Fresnel zone). We then use a function $bw()$ that maps the received power to the negotiable bit-rate as per the data sheets of the devices [21]. They range between $bw(-72\,\text{dBm}) = 400\,\text{Mb}/s$ for the coding scheme MCS9 [2]

[2]We consider a $s_{ch} = 40\,Mhz$ channel width, 2 MIMO streams and guard interval 400 ns, as per device specifications. The devices considered are: mANTBox 19s and LHG XL5 ac.

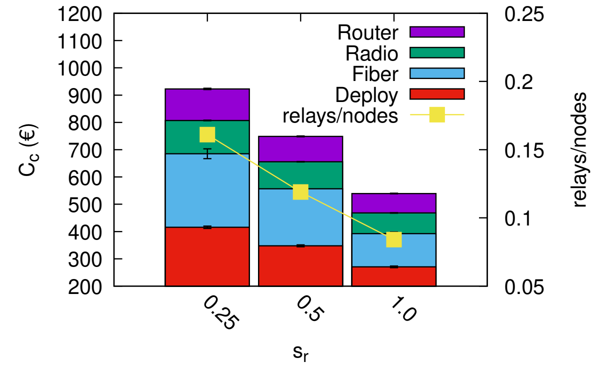
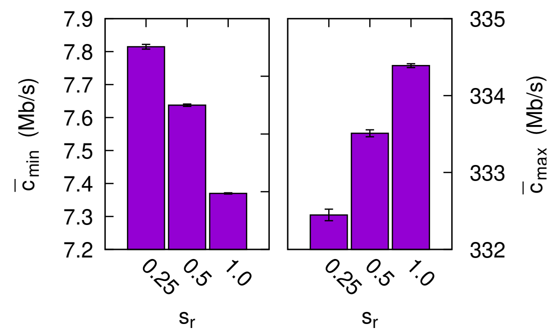


Fig. 3: Overall CapEx when $c_s = 200$ (left axis) and fraction of relays in the network (right axis).



a) b)

Fig. 4: a) Effective average minimum capacity for subscriber; b) Effective maximum capacity.

and $bw(-96\,\text{dBm}) = 30\,\text{Mb}/s$ for the coding scheme MCS0, and we multiply this value for a MAC efficiency parameter $c_e = 0.84$ [28]. 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 capacity channel $c_{ch} = 360$ Mb/s made in Eq. (9). We then compute the average minimum and maximum bit-rate per subscriber ($\bar{c}_{min}$ and $\bar{c}_{max}$, respectively). Both depend on the bottleneck on the path from the gateway to the subscriber, see Appendix B for the details on how we derive them.

Fig. 4a shows $\bar{c}_{min}$, the value is slightly greater than the expected one (c_{sub}) as Eq. (9) often overprovisions the number of the radios due to geographical constraints or the *upper part* operator ($\lceil\,\rceil$). Fig. 4b 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.

C. Comparison with other ISPs

Fig. 5 shows S_c (as in Eq. (5)) which corresponds to an ideal monthly recurring cost for subscribers including OpEx and CapEx amortized over 60 months (5 years). Tab. IV shows the comparison of WBN for two values of s_c with the monthly price of xDSL offers and Starlink satellite connectivity. Some

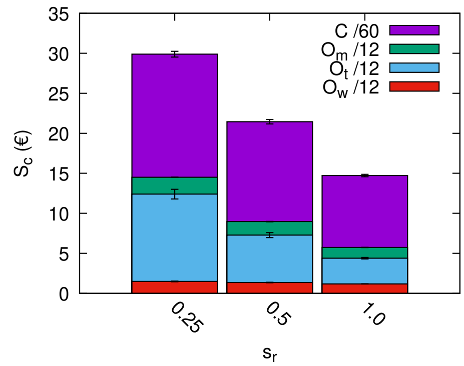


Fig. 5: Lower bound on the subscription price assuming the CapEx get amortized on 60 months.

	$WBN_{0.25}$	$WBN_{0.5}$	xDSL	SAT
Upfront cost (€/subscriber)	922.79	748.81	480	719
Recurring cost (€/subscriber)	14.51	8.97	29	99
Monthly cost over 5y	29.89	21.45	34	110
Unconnected Households	8%	9%	18%	-
Max Speed (Mb/s)	332.45	333.51	12.61	250
Min Speed (Mb/s)	7.81	7.64	7.2	-
Monthly Max (€/Mb/s)	0.09	0.06	2.69	0.44
Monthly Min (€/Mb/s)	3.82	2.81	4.72	-

TABLE IV: Comparison with commercial offers.

details are important to notice: with 0.25 penetration the cost is 12% below the lowest price, and with 0.5 penetration it is 37% lower. 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 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. Here the comparison is harder to make, because our model roughly respects a contention ratio of 50 between maximum and minimum capacity, while we do not know the contention ratio for 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 of wired links and the allocated capacity for the whole municipality) 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 with our 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.

VI. Conclusions

This paper introduced a model for the design of a WBN that takes into account the economical and the technical constraints, based on real data. With our approach, we could evaluate the feasibility of a multi-hop mesh network to provide connectivity in rural areas, and showed that it is highly competitive with existing offers. The model is fully configurable and reusable (we share all the code and the generated data) and fills a gap between technical research, which often makes unrealistic assumptions when using synthetically generated data, and real-world requirements that are accessible only by operators and are rarely considered in scientific works.

ACKNOWLEDGMENT

This work has received funding through the Spanish grant PID2019-106774RB-C21 and Generalitat de Catalunya through 2017-SGR-990.

REFERENCES

[1] P. Micholia, M. Karaliopoulos *et al.*, "Community networks and sustainability: a survey of perceptions, practices, and proposed solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3581–3606, 2018.

[2] 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

[3] 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.

[4] S. Hasan, Y. Ben-David *et al.*, "The challenges of scaling WISPs," in *Annual symposium on computing for development (DEV)*, 2015.

[5] 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

[6] L. Maccari, G. Gemmi *et al.*, "Towards scalable community networks topologies," *Ad Hoc Networks*, vol. 94, 2019.

[7] D. Cameron, A. Valera, and W. K. Seah, "Elasticwisp: Energy-proportional wisp networks," in *Network Operations and Management Symposium (NOMS)*. IEEE/IFIP, 2020.

[8] T. Pötsch, P. Schmitt *et al.*, "Helping the lone operator in the vast frontier," in *Workshop on Hot Topics in Networks (HotNets)*. ACM, 2016.

[9] 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.

[10] S. Mubaraq, J. Hwang *et al.*, "Economic analysis of networking technologies for rural developing regions," in *Workshop on Internet Economics (WIE)*, vol. 10. ACM, 2005.

[11] 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*, vol. 32, no. 3, p. e4155, 2021.

[12] E. J. Oughton, K. Katsaros *et al.*, "An open-source techno-economic assessment framework for 5G deployment," *IEEE Access*, vol. 7, pp. 155 930–155 940, 2019.

[13] F. Yaghoubi, M. Furdek *et al.*, "Reliable topology design of wireless networks under correlated failures," in *International Conference on Communications (ICC)*. IEEE, 2018.

[14] L. Maccari, M. Karaliopoulos *et al.*, "5g and the internet of everyone: Motivation, enablers, and research agenda," in *2018 European Conference on Networks and Communications (EuCNC)*, 2018.

[15] A. Chaoub, M. Giordani *et al.*, "6g for bridging the digital divide: Wireless connectivity to remote areas," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 160–168, 2022.

[16] 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.

[17] 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.

[18] 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.

[19] N. Mark, *Networks: an introduction*. Oxford University Press New York, 2010.

[20] S. Khuller and R. Thurimella, "Approximation Algorithms for Graph Augmentation," *Journal of Algorithms*, vol. 14, no. 2, pp. 214–225, 1993.

[21] "Mikrotik wireless devices' prices and datasheets," accessed: 2022-04-28. [Online]. Available: https://mikrotik.com/products/group/wireless-systems

[22] "Mikrotik routers' prices and datasheets," accessed: 2022-04-28. [Online]. Available: https://mikrotik.com/products/group/ethernet-routers

[23] I. Chlamtac, A. Gumaste, and C. A. Szabó, Eds., *Broadband Services: Business Models and Technologies for Community Networks*. John Wiley & Sons, Ltd, 2005.

[24] W. Waites, J. Sweet *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.

[25] S. Verbrugge, D. Colle *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.

[26] L. Cerdà-Alabern, R. Baig, and L. Navarro, "On the Guifi.net community network economics," *Computer Networks*, vol. 168, 2020.

[27] "Xarxa oberta fiber transport service," accessed: 2022-05-03. [Online]. Available: https://web.archive.org/web/20220503174738/https://xarxaoberta.cat/services/transparent

[28] E. H. Ong, J. Kneckt *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.

APPENDIX

A. Determining w_i

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[3] 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.

We call V_i^j the volume of the building p_i in the section c_j:

$$V_i^j = \frac{\cap(c_j, p_i)}{|p_i|} \times v_i \tag{16}$$

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. (16) 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:

$$H_i^j = \frac{V_i^j}{\sum_{p_k \in \mathcal{P}} V_k^j} \times h_j. \tag{17}$$

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

[3]See the LiDAR datasets released by the Italian Ministry of Environment https://www.mite.gov.it

of having a household in a given building b_i in the municipality under analysis.

$$w_i = \frac{1}{\sum_{c_k \in \mathcal{C}} h_k} \sum_{p_j \in \mathcal{C}} H_i^j \tag{18}$$

B. Capacity Computation

Let $n_0, n_1, n_2 \ldots n_t$ be a sequence of nodes on the shortest path from a gateway ($g = n_0$) to a subscriber (s hosted on 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 $n_i \to n_j$ (based on the $bw()$ function introduced earlier). Then we derive the average bit-rate per neighbor of node n:

$$\hat{r}(n) = \frac{c_e}{\delta(n)} \sum_{n_j \in \Delta(n)} r(n, n_j) \tag{19}$$

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:

$$r_{min}(n) = \frac{\hat{r}(n) \cdot n_a(n)}{2 \cdot sp(n)} \tag{20}$$

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. (20) is modified removing the 2 at the denominator. The minimum capacity for subscriber s is the bottleneck on the path from the gateway:

$$c_{\min}(s) = \min_{n_i \in \{n_0 \ldots n_t\}} r_{min}(n_i) \tag{21}$$

and $\bar{c}_{min}$ is the average on all subscribers. To compute $c_{max}(s)$ we use:

$$c_{\max}(s) = \min_{n_i \in \{n_1 \ldots n_t\}} r(n_{i-1}, n_i) \tag{22}$$

where $\bar{c}_{\max}(s)$ is the average on all subscribers.