ARCHITECTURES AND ALGORITHMS FOR RELIABLE 5G NETWORK DESIGN

Presentata da: Bahare Masood Khorsandi

Coordinatore Dottorato: **Supervisore:**
Prof.ssa Ing. Alessandra Costanzo   Prof.ssa Ing. Carla Raffaelli

Abstract

The fifth generation of mobile technology (5G) is positioned to address the demands and business contexts of 2020 and beyond. It is expected to enable a fully mobile and connected society and to empower economic transformations in countless ways coming from novel services. Applications like intelligent transportation systems, smart manufacturing, virtual and augmented reality, e-Health services, etc require massive Machine Type Communications (mMTC), enhanced Mobile Broadband (eMBB), ultra Reliable Low Latency Communications (uRLLC) to be supported by single infrastructure. This enhanced performance is expected to be provided along with the capability to control a highly heterogeneous environment, and among others, ensure security and privacy.

As the amount of data traffic on mobile networks continues to grow, network operators are meeting the demands by adopting Cloud/Centralized Radio Access Network architectures (C-RAN). This new approach to network architecture has two clear advantages. The first is a significant reduction in both CApital EXpenditure (CAPEX) and OPerational EXpenditure (OPEX) for operators. The second is improved user experience through less interference. Maintaining this network architecture will require high capacity and low latency links to transport data. Given the strict requirements of these links, commonly referred to as "fronthaul" links, dedicated fiber connections are usually required. Hence, the minimization of network cost and energy consumption has become a necessity for mobile network operators.

Xhaul (front/mid/backhaul), defined as the common flexible transport solution for future 5G networks, aims at integrating the fronthaul and backhaul networks with all their wired and wireless technologies in a common packet-based transport network under an SDN (Software-Defined Network)-based and NFV (Network Function Virtualization)-enabled common control. This solution will hence enable a flexible and software-defined reconfiguration of all networking elements in a multi-tenant and service-oriented unified management framework.

This Ph.D. thesis investigates the resilient and cost-efficient design of both C-RAN and Xhaul architectures. Minimization of network resources as well as reuse of already deployed infrastructure, either based on fiber, wavelength, bandwidth or Processing Units (PU), is investigated and shown to be effective to reduce the overall cost. Moreover, the design of a survivable network against a single node (Baseband Unit hotel (BBU), Centralized/Distributed Unit (CU/DU)) or link failure proposed. The novel function location algorithm, which adopts dynamic function chaining in relation to the evolution of the traffic estimation also proposed and showed remarkable improvement in terms of bandwidth saving and multiplexing gain with respect to conventional C-RAN. Finally, the adoption of Ethernet-based fronthaul and the introduction of hybrid switches is pursued to further decrease network cost by increasing optical resource usage.

Dedication

I would like to dedicate this Ph.D. dissertation to four beloved people who mean so much to me. They were always by my side throughout this journey:

First and foremost, to my Mom and Dad whose words of encouragement and push for tenacity ring in my ears and love for me knew no bounds.

Next to my lovely brother, Mobin, who always believe in me and never let me give up my dreams.

Last but not least, I am dedicating this work to my love and my best friend, Max, who has been a constant source of inspiration and support.

Acknowledgements

The process of earning the highest academic degree and writing a dissertation is long and difficult and it is certainly not done single handily. First and foremost, I would like to express my sincere gratitude to my supervisor Professor Carla Raffaelli for the continuous support of my Ph.D. journey, research and contribution in the field, for her patience, motivation, and immense knowledge. Her guidance helped me throughout these years in research and writing this thesis.

Besides my supervisor, I would like to thank Professor Didier Colle and Dr. Wouter Tavernier and IDLab of Ghent University who welcomed me warmly and providing me this opportunity to work with their significant research group.

I would also like to express my gratitude towards Professor Lena Wosinska, Professor Paolo Monti, Dr. Matteo Fiorani and Dr. Carlos Natalino da Silva who were collaborating closely and supporting in each step of the way, for their insightful comments and encouragement which incensed me to widen my research from various perspectives.

Finally, none of this would have been possible without the support, concern, and strength of my colleagues in the NetLab group of the University of Bologna.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Fifth Generation Mobile Network (5G)

In the past years, the wireless industry unifies around Long-Term Evolution (LTE). Now it is evolving deployments to a single technology, enabling an ecosystem larger than ever before [1]. Already more than a quarter of all global mobile subscribers are using LTE and it is expected that by 2021 this will increase to more than half [2]. While LTE deployments continue to expand and grow across the world, certain regions such as Korea, Japan, China, and the U.S. have nearly reached or exceeded 90 percent penetration of LTE [3]. This pushed up the focus in the mobile industry towards 5th Generation (5G) mobile technology, standards development, demos, and trials.

Over the next few years, 5G is expected to reinvent entire industries with new use cases, business models, and organizations that will emerge in response to shifting technology and business landscapes. The growth of 5G wireless technologies is necessitating approaches that include new architectures. Today, emerging 5G markets including AR (Augmented Reality)/VR (Virtual Reality), V2X (Vehicle-to-Everything), transportation, manufacturing, health, and education are being tooled with applications that operate in a time-sensitive fashion, requiring a range of data bandwidth, varying degrees of cell densification and spectrum operating range [4]. Unlike previous generations, 5G platforms are relying on strong distributed cloud foundations of network and compute transformation that will lead operators to new market growth.

There continue to be growing demands for higher throughput and more data capacity, particularly for video, to provide better broadband services. But data demand is just one of the drivers for 5G. In addition, 5G is targeted to address new vertical markets including massive Machine Type Communications (mMTC) [5], ultra Reliable Low Latency Communications (uRLLC) [6] and enhanced Mobile Broadband (eMBB) [7]. Bellow, a summary of the characterization of each of these technologies is given:

- **massive Machine Type Communications (mMTC)** are characterized by fully automatic data generation, exchange, processing, among intelligent machines, with or without the low intervention of humans a.k.a machine-centric instead of human-centric. With the rapid penetration of embedded devices, mMTC is becoming the dominant communication paradigm for a wide range

of emerging smart services including healthcare, manufacturing, utilities, consumer goods, and transportation. Industry analysts predict that 50 billion devices will be connected to mobile networks worldwide by 2020 [8]. While mobile phone devices communicating among humans will still exist, machine-type devices sending bits of information to other machines, servers, clouds, or humans will account for a much larger proportion.

- **ultra Reliable Low Latency Communications (uRLLC)** are a new service category to accommodate emerging services and applications having stringent latency and reliability requirements. At its core, uRLLC mandates a departure from expected utility-based network design approaches, in which relying on average quantities (e.g., average throughput, average delay and average response time) is no longer an option but a necessity. uRLLC focused applications require an End-to-End (E2E) delivery of data with reliability, security, and minimum latency. Such requirements have driven the 3rd Generation Partnership Project (3GPP) [9] to set the desired Quality of Service (QoS) requirements such as an air interface latency of 1 ms and 99.999 percent system reliability for uRLLC.

- **enhanced Mobile Broadband (eMBB)** mainly aiming to fulfill users' demand for an increasingly digital lifestyle and focusing upon facilities that implicate high requirements for bandwidth. eMBB focuses on supporting the ever-increasing end-user data rate and system capacity. To fulfill this demand, eMBB introduces two major technology enhancements. (I) A shift of frequency spectrum to *cmWave* and *mmWave* range to achieve much higher bandwidth allocations and (II) advanced antenna array that includes tens or even hundreds of TX/RX antenna elements to enable massive Multiple Input Multiple Output MIMO and beamforming [10].

In conclusion, the capabilities of 5G will extend far beyond those of the current LTE networks, therefore new technologies and architectures are needed. Those new technologies and architectures have to tackle the above mentioned strict requirements and design in a way that be able to optimize cost and energy efficiency.

1.2 Network Architecture Evolution Toward 5G

Evolved Packet System (EPS) of LTE refers to the logical architecture composed of the Radio Access Network (RAN), called the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and the Evolved Packet Core (EPC) [11]. The objective of this logical architecture is to enable a flat IP-based network and provide a standardized set of network elements and network interfaces. These elements and interfaces enable operators to integrate equipment and implementations from different vendors into a single system while ensuring interoperability. To support diverse services such as e-Health, the Internet of Things (IoT), and V2X in future mobile networks, we see a need for enhancing the EPS toward a flexible mobile network accommodating novel architectural principles while maintaining backward compatibility.

In a traditional Distributed Radio Access Network (D-RAN), the Base Station (BS) comprises two modules, (I) the Remote Radio Unit (RRU) for transmis-

sion and reception of radio signals, Digital-to-Analog/Analog-to-Digital Conversion (DAC/ADC) of the baseband signals, frequency conversion, and power amplification, and (II) the Baseband Unit (BBU) performing the digital processing functions of layer 1, 2 and 3 [12]. Every BS hosts its "local BBU" and has a dedicated housing facility, which is not shared with other BSs. Hence, in D-RAN, power consumption, as well as investment and maintenance costs, increase linearly with the number of BSs. Given the rapid traffic growth envisioned for the next years, simply increasing BSs density in D-RAN does not represent a scalable solution. A novel network architecture, called Centralized-RAN (C-RAN), has been proposed as a more scalable alternative to D-RAN in terms of both power and cost-efficiency [13]. The main idea of C-RAN is that multiple BBUs are placed in a single physical location (BBU hotel), which is connected to several RRUs through a high capacity fronthaul network. Thanks to this centralization, the baseband resources in the BBU hotel can also be virtualized and shared among several BSs, and a significant reduction in the overall computational resources can be achieved due to multiplexing gain. BBU centralization also allows to share of maintenance costs and power consumption among several BSs and promotes the utilization of advanced interference cancellation techniques such as the Coordinated Multi-point (CoMP).

Despite C-RAN appealing design aspects, one key obstacle in its adoption is the excessive capacity requirements on the fronthaul links to provide BBU and RRU connections. Shifting all baseband processing to the remote BBU hotel implies the adoption of a high number of optical channels with strict latency constraints. To relax the excessive fronthaul requirements, the concept of C-RAN is being revisited, and more flexible distribution of baseband functionalities between the RRU and BBU hotel is considered. Rather than offloading all baseband processing to a single entity like the BBU hotel, it is possible to divide it into several blocks throughout the network which leads to a significant reduction of the bandwidth needed on the transport links [14]. This concept is known as "functional split" and was firstly introduced in the new architecture design for the 5G access network named "Xhaul" or "cross-haul". Next Generation Fronthaul Interface (NGFI) is defined as the fronthaul interface between BBU and RRU for the next generation of radio network infrastructure [15]. NGFI redefines the baseband processing split through the positioning of baseband function stack components between BBU and RRU. Design methodologies to apply functional split in the 5G network in order to exploit this potential still need investigation. In particular, the bandwidth available on the fronthaul links should be efficiently used and dynamically allocated to service slices.

1.3   Survivable Network Design

By having the rapid growth of mobile user's demands, wireless mobile networks become a part of everyday life. Therefore, the interruption or failure of the service for even a short period may have fatal consequences in terms of QoS and user satisfaction. In this context, how to prevent service failure and minimizing the failure time if occurred becomes a critical issue. Hence, "resiliency" is one of the main requirements for mobile networks, which is the ability to provide and maintain five-nines QoS in the face of various faults and challenges to normal operation [16]. Network survivability and resiliency is a well-established research area for Wavelength Division Multiplexing (WDM) optical networks [17]. However, these works mainly

focus on the path and link protection/restoration in mesh WDM networks. The most common network survivability techniques include 1+1 Automatic Protection Switching (1+1 APS), Demand-wise Shared Protection (DSP), Shared Backup Path Protection (SBPP) [18].

In the context of radio access network architectures whether it is C-RAN or Xhaul, an important aspect to deal with is the entities that are centralized to execute the sole level of baseband processing which makes the network vulnerable to failures. The failure of a single active node may cause severe service outages, calling for efficient and reliable design. Furthermore, having a failure in any part of the transport network specifically on links might cause service outage for a large area with a significant number of users. Considering, the cost of providing fronthaul links, careful planning is necessary especially when a large number of BSs have to be deployed.

1.4   Contribution and Outline of the Thesis

This research aims to investigate the various optimization methodologies for accomplishing a survivable 5G radio access network. This work contains the two main architecture designs, namely C-RAN and Xhaul. An in-depth introductory for both cases exist over the technological features and implementation principles from a networking standpoint. Then, different analyses based on the survivability of network design and their benefits over current aggregation infrastructure are also presented. In the end, this work is extended by a comprehensive investigation based on statistical multiplexing for packet-based fronthaul over the traffic aggregation with different priorities.

This thesis is divided into seven chapters, including the current one, and they are structured as follows:

- In **Chapter 2**, the focus is on presenting an overview of the C-RAN architecture, advantages, and challenges of its implementation. A classification of the various architectural solutions for an antenna is based on the network architecture is also presented. The final part of this chapter is dedicated to the introduction of the C-RAN transport protocol its constraints and specifications.

- **Chapter 3** is dedicated to the first category of reliable design for C-RAN. In specific, the centralized model which is based on the global orchestration hypothesis. Moreover, different approaches for implementing a survivable access network based on different constraints and requirements shown in detail.

- In **Chapter 4**, the second category of reliable design for C-RAN has been demonstrated. The distributed approach is presented which is based on Machine Learning (ML) method. We also present the case study for vehicular networks in which a distributed method can be efficiently implemented.

- In **Chapter 5**, a comprehensive introduction of the new RAN architecture for 5G is provided. The network divisions and terminology is explained in detail. The new transport network protocol is also introduced. Since the big part of new technology is the baseband functional split, in this chapter a demonstration of the different options and their constraints are also presented.

- **Chapter 6** is the continuation of the introduction started in Chapter 4 with the addition of the new methodology of baseband functional chaining which is applied to new 5G RAN and can be adapted to the variation of user traffic. Furthermore, an investigation of the survivability aspect of this new methodology is also presented. In this context, two protection approaches are introduced: I) Dedicated Path Protection (DPP) II) Shared Path Protection (SPP).

- In **Chapter 7**, an architecture capable of multiplexing fronthaul and backhaul traffic together on the same optical resources is provided, to increase transport resources usage. The main focus of this chapter is to demonstrate the benefits of the statistical multiplexing gain for a packet-based fronthaul.

- **Chapter 8** provides conclusions about supplied and open issues with corresponding future work.

Chapter 2

Cloud/Centralized Radio Access Network Architecture (C-RAN)

2.1 Introduction

The data traffic explosion generated by an increasing number of connected devices, e.g., smartphones and tablets, requires an evolution of the current Radio Access Networks (RANs) architecture and technologies. Advances already available for today's RANs include larger frequency bandwidths, mechanisms for increased spectral efficiency, e.g., Orthogonal Frequency Division Multiplexing (OFDM), and Multiple Input Multiple Output (MIMO) transmission/reception systems. These improvements are the basis of 4G standards like Long Term Evolution (LTE) and LTE Advanced. Several other technological advances are under investigation and will play a key role in future 5G networks.

However, the need for radical changes in the current network architecture is necessary to face the revolution in the data traffic volumes. Centralized RAN (C-RAN) is introduced as one of those novel 5G paradigms which evolve the mobile network architecture. C-RAN introduces encouraging savings in the network total cost and energy consumption. Despite the attractive advantages, C-RAN also comes with its challenges in the fronthaul transport network.

In this chapter, the technology evolution toward C-RAN is illustrated in detail. Section 2.2 is dedicated to the evolution path starting from radio access design for LTE and Distributed Radio Access Network (D-RAN) to the need for more complex design required for 5G and ending up on centralized design. In Section 2.3 the detail of how Base Station (BS) evolved until the centralization era described. At the end of this chapter, there is one dedicated section, Section 2.4 for the transport protocol which introduced specifically for C-RAN. Its advantages and strict requirements also shown in detail. To conclude, in Section 2.5, the main advantages and challenges of C-RAN are then detailed described.

2.2 Radio Access Network Evolution Towards C-RAN

A typical mobile network in shown in figure 2.1. It is divided into three parts: Radio Access Network (RAN), backhaul network and core network, also known as Evolved
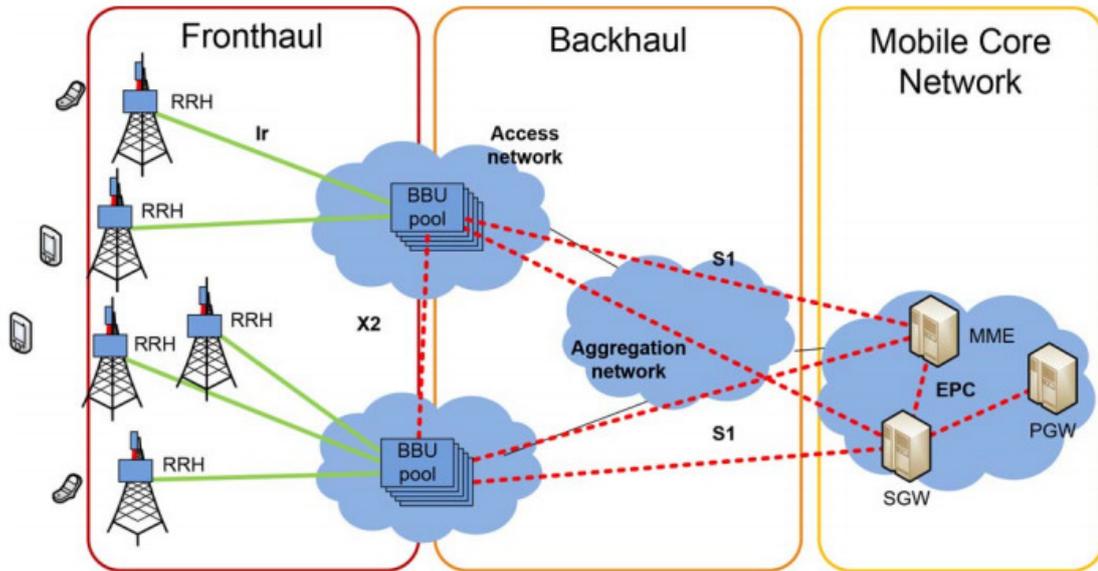
Figure 2.1: Network architecture.

Packet Core (EPC). The RAN includes all and solely the systems performing radio-access related functions, i.e., directly managing radio transmission and reception towards/from mobile devices. The backhaul network performs traffic aggregation and transport between the RAN and the core network. For this reason, its architecture and implementation can be almost agnostic concerning radio access and core architectures. Finally, the core network is in charge of all remaining non radio access related functions and acts as a gateway towards all other mobile and fixed networks, i.e., towards the Internet [19].

The RAN is in charge of exchanging data with the end-users, through Base Stations (BSs). Each BS performs radio access functions, i.e., it manages the transfer of user and control data towards (downlink) and from (uplink) several users simultaneously, using the physical layer and multiple access protocols, according to the so-called radio, or air, interface. The processing equipment of a BS is made up of two parts: a Baseband Unit (BBU), sometimes referred to as a Digital Unit (DU), and a Remote Radio Unit (RRU), also referred to as Remote Radio Head (RRH) or simply Radio Unit (RU). An example of a traditional BS is depicted in figure 2.2. The figure contains the section which illustrated different parts of the base station such as Power Amplifier (PA) and Radio Frequency (RF) which are responsible for the radio processing part and baseband, transport, control and synchronization for the baseband processing part.

The LTE RAN uses a flat architecture with a single type of node, the eNodeB. The eNodeB is responsible for all radio related functions in one or several cells. It is important to note that an eNodeB is a logical node and not a physical implementation [20]. One common implementation of an eNodeB is a three-sector site, where a base station is handling transmissions in three cells, although other implementations can be found as well, such as one baseband processing unit to which several RRUs are connected. One example of the latter is a large number of indoor cells, or several cells along a highway, belonging to the same eNodeB. Thus, a base station is a possible implementation of, but not the same as, an eNodeB. As can
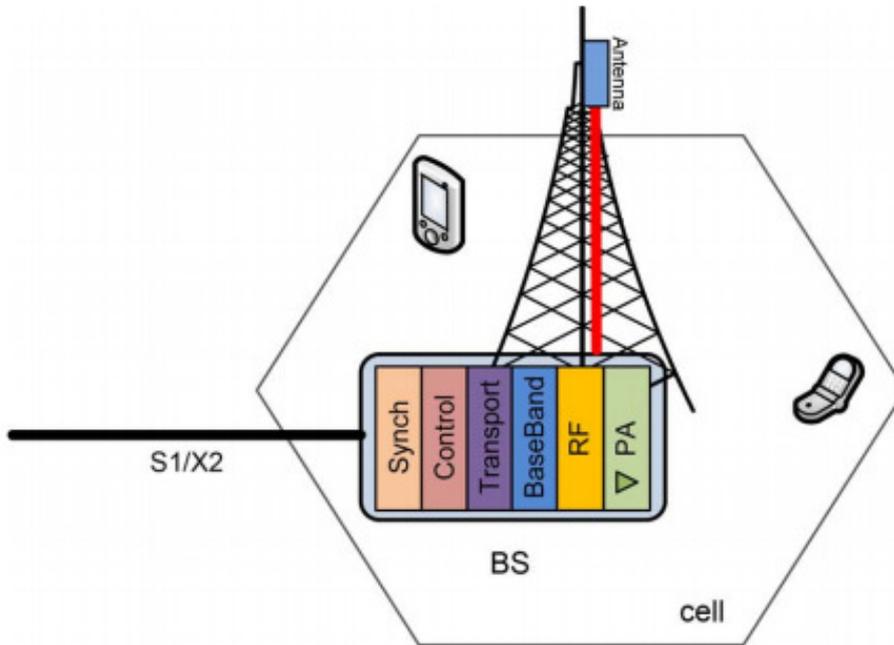
Figure 2.2: Traditional base station.

be seen in figure 2.3, the eNodeB is connected to the EPC or core network using the $S1$ interface, more specifically to the Serving Gateway (S-GW) utilizing the $S1$ user-plane part, $S1 - u$, and to the Mobility Management Entity (MME) employing the $S1$ control-plane part, $S1 - c$. One eNodeB can be connected to multiple MMEs/S-GWs for load sharing and redundancy. The $X2$ interface, connecting eNodeBs, is mainly used to support active-mode mobility. This interface may also be used for multi-cell Radio Resource Management (RRM) functions such as Inter-Cell Interference Coordination (ICIC). The $X2$ interface is also used to support lossless mobility between neighboring cells employing packet forwarding.

Mobile data transmission volume is continuously rising. It is forecasted to grow 13-fold from 2012 until 2017 according to Cisco [21]. Therefore, to satisfy growing user demands, mobile network operators have to increase network capacity. As spectral efficiency for the LTE standard is approaching the *Shannon limit*, the most prominent way to increase network capacity is by either adding more cells, creating a complex structure of Heterogeneous and Small cell Networks (HetSNets) [22] or by implementing techniques such as multiuser Multiple Input Multiple Output (MIMO) [23] as well as massive MIMO [24], where numerous antennas simultaneously serve a number of users in the same time-frequency resource. However, this results in growing inter-cell interference levels and high costs.

C-RAN is a novel mobile network architecture, which has the potential to answer the previously mentioned challenges. The concept was first proposed in [25]. In C-RAN, baseband processing is centralized in colocations known as BBU hotels and shared among sites. This means that it can adapt to non-uniform traffic and utilizes the resources, i.e., base stations, more efficiently. Due to the fact that fewer BBUs are needed in C-RAN compared to the traditional architecture, C-RAN has also the potential to decrease the cost of network operation, because power and energy consumption is reduced compared to the traditional RAN architecture. New BBU
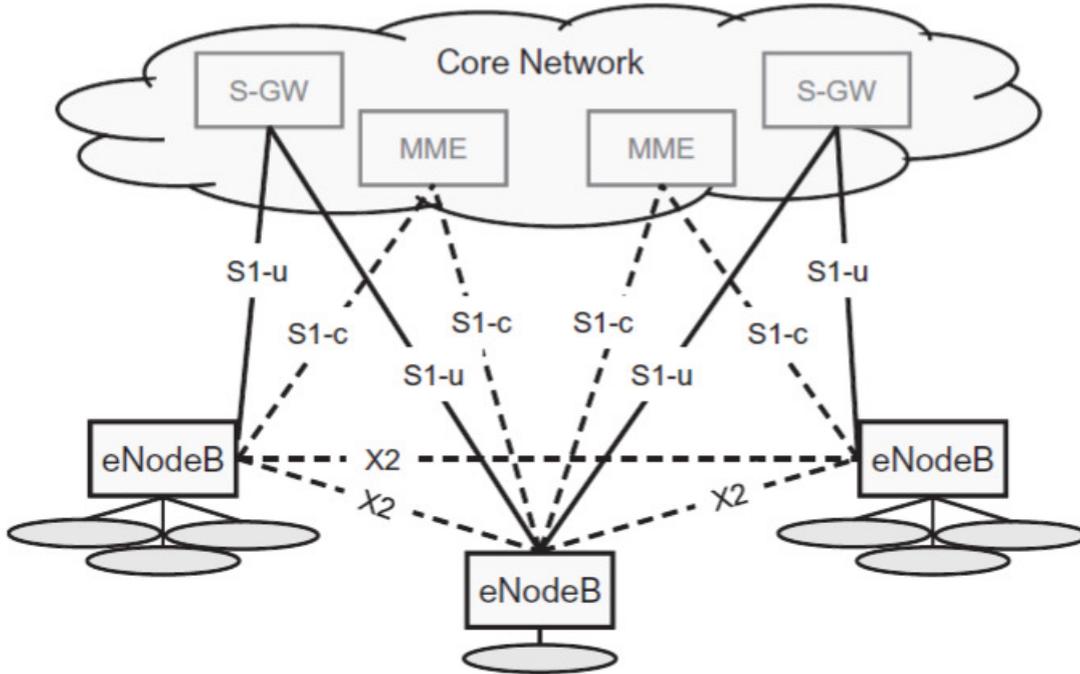
Figure 2.3: LTE architecture.

can be added and upgraded easily, thereby improving scalability and easing network maintenance. BBU hotel can be shared by different network operators, allowing them to rent RAN as a cloud service. As BBUs from many sites are co-located in one hotel, they can interact with lower delays. Methods for implementing load balancing between the cells are also facilitated. Furthermore, network performance is improved, e.g., by reducing delay during intra BBU hotel handover [26].

On the other hand, a C-RAN introduces strict capacity and latency requirements on the transport network [27], which derive from the transmission of digital In-phase and Quadrature (IQ) data streams between RRUs and BBU hotels, i.e., the fronthaul traffic. In this context, a Passive Optical Network (PON) can play an important role. PON is a fiber-optic network utilizing a point-to-multipoint topology and optical splitters to deliver data from a single transmission point to multiple user endpoints.

In contrast to an active optical network, electrical power is only required at the send and receive points, making a PON inherently efficient from an operating cost standpoint. PONs are used to simultaneously transmit signals in both the upstream and downstream directions to and from the user endpoints. Thanks to its ability to provide high capacity and low latency connections between RRUs and BBU hotels.

However, the deployment of C-RANs with an optical WDM transport might result in high deployment cost, if the network is not properly designed. C-RAN architecture is targeted by mobile network operators, as envisioned by China Mobile Research Institute, IBM, Alcatel-Lucent, Huawei, ZTE, Nokia Siemens Networks, Intel and Texas Instruments. Moreover, C-RAN is seen as a typical realization of a mobile network supporting soft and green technologies in the 5G mobile network in the year 2020 horizon [28].

2.3   Base Station Evolution

The RAN is the direct interface to mobile devices (UE) via radio links established towards BS. Each BS manages the transfer of users and controls data towards (downlink) and from (uplink) several UEs simultaneously, utilizing physical-layer and multiple-access protocols, according to the so-called radio, or air, interface. Some higher-layer radio access functions (e.g., radio resource control) can be either performed by other network nodes (e.g., Base Station Controllers (BSC), or Radio Network Controllers, (RNC)) that manage several BSs, or directly embedded into the BSs themselves.

Each BS manages UEs belonging to a specific coverage area, denoted as "cell", and the RAN also coordinates the procedures for user mobility, i.e., allowing UEs to move across adjacent cells (handovers), without losing data connection. BSs are placed into premises denoted as "cell sites", whose geographic coordinates are influenced by many different factors, most notably coverage, capacity planning and infrastructural/costs constraints [29]. To save costs, a consolidated practice is implementing more than one BS into a single cell site, thus dividing the coverage area into up to three cells, denoted also as "sectors". A typical cell site consists of a tower, on top of which there are installed BS directional antennas (at least one per sector), and a cabinet, or shelter, where the remaining BS equipment is installed. The cabinet also hosts collateral systems that do not perform network functions but ensure proper BS working. They typically consist of power supplying (AC/DC converters, backup batteries) and cooling systems (fans, air conditioning).

2.3.1   Distributed Traditional Base Station

This architecture is shown in figure 2.4. In a traditional Distributed Radio Access Network (D-RAN), BS comprises two modules, RRU, and BBU which hosts its "local BBU" and has a dedicated housing facility, which is not shared with other BSs. Hence, in D-RAN, power consumption, as well as investment and maintenance costs, increase linearly with the number of BSs. Given the rapid traffic growth envisioned for the next years, simply increasing BSs density in D-RAN does not represent a scalable solution. The RRU is connected to the antenna through coaxial cable. In general, this architecture experience high power loss in the coaxial cable depending on the distance between the antenna and the cell cabinet. This type of architecture was employed in 1G and 2G mobile networks.

2.3.2   Base Station with Remote Radio Unit (RRU)

This architecture is shown in figure 2.5. In this architecture, the BBU remains in the cell cabinet while the RRU is placed beside the antenna. The main advantage of this solution is that the RRUs can be placed on rooftops to reduce air conditioning energy consumption. The BBUs can be placed in a more convenient site with lower rental and maintenance costs. The Common Public Radio Interface (CPRI) [30] protocol is used as a radio interface protocol for In-phase/ Quadrature (IQ) data transmission between RRU and BBU. CPRI requires a very high data bit rate and very low latency. Each RRU is statistically assigned to one BBU. This architecture is first deployed in 3G networks.
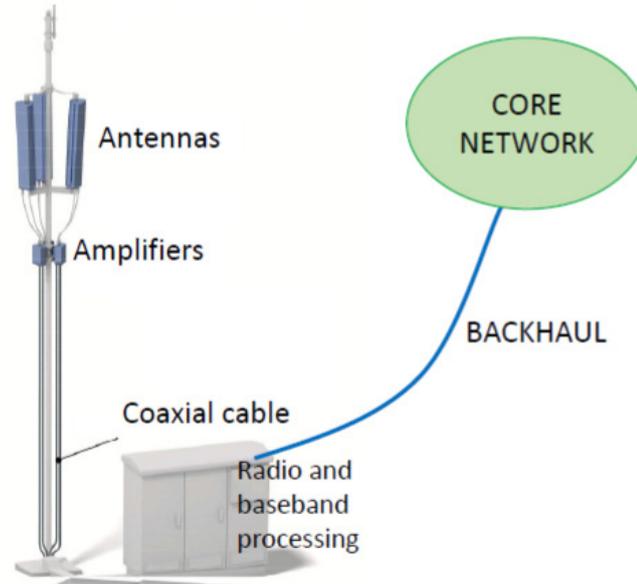
Figure 2.4: Distributed base station.

2.3.3 Centralized Base Station

This architecture is shown in figure 2.6. In a C-RAN architecture, the BBUs are not only separated from the RRUs, but they are located in a centralized unit, the BBU hotel, capable to host several BBUs. This way the housing facility expenses and energy consumption can be considerably reduced. Moreover, a centralized unit provides a common communication channel between the BBUs. This can be exploited to perform coordinated processing. A further step is taken by implementing a virtualized BBU hotel consisting of General Purpose Processors (GPP) for baseband processing. GPPs can dynamically be assigned to different RRUs. This allows performing load balancing and efficient resource utilization. The term C-RAN stands at the same time for centralized, clean, cooperative and cloud RAN.

2.3.4 Virtualization in C-RAN

Virtualization technology facilitates the logical isolation of resources while the physical resources are shared in a dynamic and scalable way. As it is shown in figure 2.7 those resources include network, computing or storage resources. From those resources, network virtualization is critical in C-RAN and its deployment architectures. Network virtualization consists of multiple nodes and links that are deployed on the same physical machine. Thus, such technology enables flexible control mechanisms, efficient resources, low cost, and diverse applications [31].

In the context of C-RAN, network virtualization is done at the BBU hotel level. Each BBU is a virtual node while the communication between them is the virtual link. The hotel operates on the one physical machine sharing Centralized Processing Unit (CPU), memory and network resources between multiple BBUs. RRUs connect the BBU hotel which distributes them over the BBUs in its virtual machine. Such technology comes with many advantages including reducing the cost, minimizing the time required for BBU communication, and most importantly scalability. Adding or removing of BBUs becomes easier as those BBUs are virtual machines which are
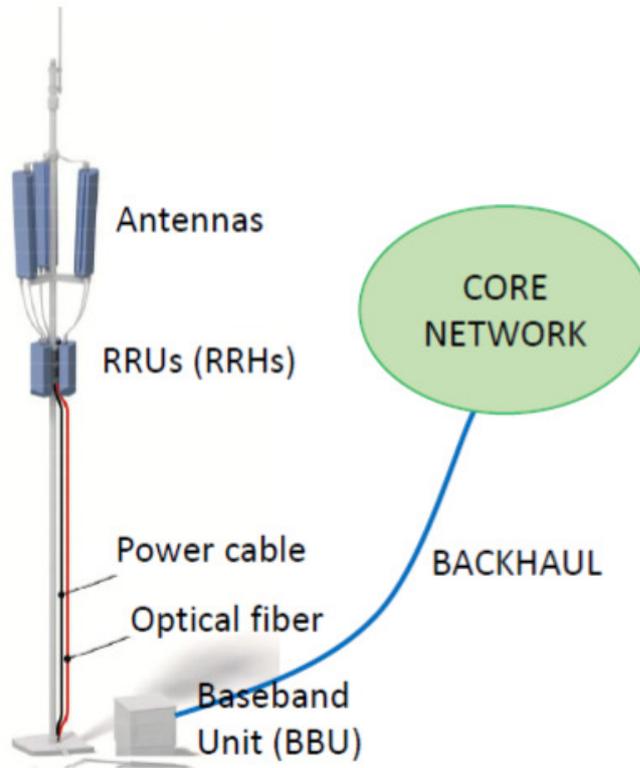
Figure 2.5: Base station with RRU.

much easier to turn off and up than physical machines.

2.4 C-RAN Transport Protocols

Despite the C-RAN advantages, it is not becoming as popular as one would have expected, due to the extremely large capacity required to transport data from antenna sites to BBU hotels (also known as fronthaul links). When a large number of antennas are employed, a bit rate over fronthaul links dramatically increases, requiring high capacity connections (e.g., fiber cables), limiting the dissemination of this architectural solution.

In October 2002, Nokia, NEC, LG, and Samsung launched the Open Base Station Standard Initiative (OBSAI) initiative [32]. The organization was tasked with standardizing the architecture of wireless base stations, internal interfaces, control modules, transmission modes, baseband, and radio frequencies. However, the possibilities of OBSAI were limited to the fact that standards were developed in the mainstream of base stations of only one manufacturer – Nokia. OBSAI divides the base stations into four main modules: a transmission, processing, radiofrequency and control module. The first provides external standard network interfaces, including Internet Protocol (IP) and Asynchronous Transfer Mode (ATM); in the second, base frequency signals are processed; the third receives transmits and amplifies RF signals and converts them from digital to analog; the fourth provides system synchronization, management, and configuration of base station equipment, monitors the operating status of other modules, and generates reports to the network element management system.
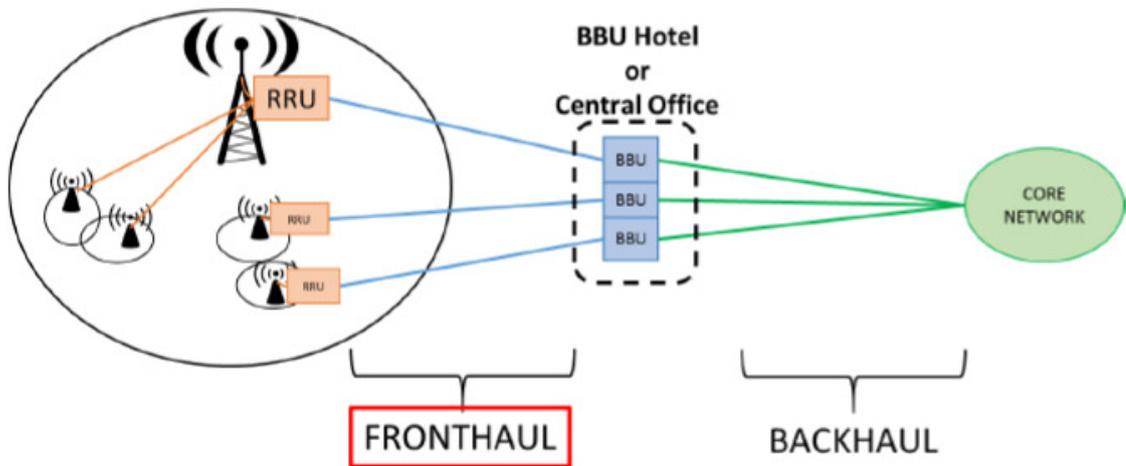
Figure 2.6: Centralized base stations.
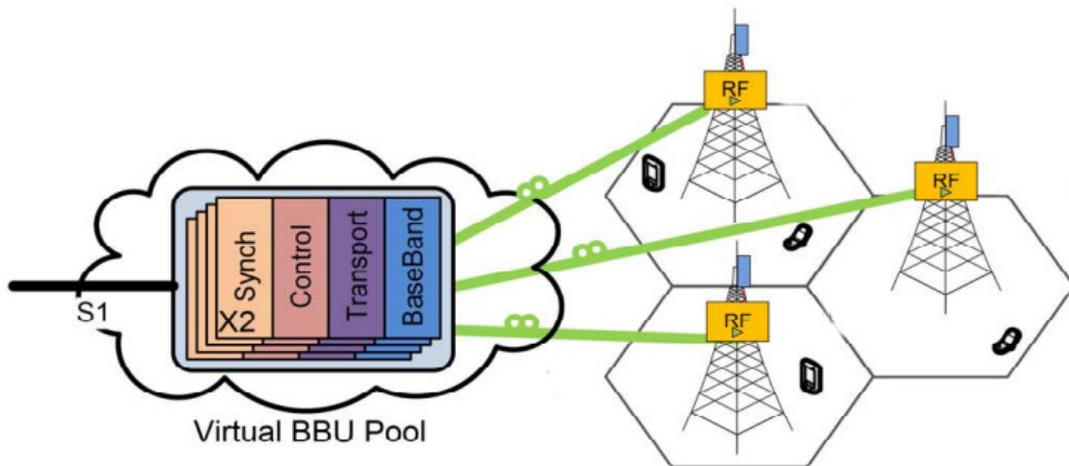
Figure 2.7: Virtualized C-RAN.

In June 2003, Ericsson, Huawei, NEC, Nortel Networks and Siemens created an alternative organization for the development of the Common Public Radio Interface (CPRI), it also began developing universal standards for key internal interfaces with an emphasis on interfaces between the base frequency band and radio frequencies. It included base station suppliers, which only increased its importance. NEC moved to the CPRI camp less than a year after its foundation; At the moment, more than 100 manufacturers have joined the organization supporting CPRI. According to the CPRI standards, the base station consists of two parts: a BaseBand Unit (BBU), or a Radio Equipment Control (REC) unit, and a Radio Frequency Unit (RRU) or Radio Access Equipment (RAE). The main distinguishing feature of CPRI interfaces is the separation between the base frequency band and the radio frequency band. CPRI standardizes interfaces between the BBU and RRU, ensuring compatibility of equipment from different manufacturers.

Finally, in May 2010, the European Telecommunications Standards Institute (ETSI) has initiated a new Industry Specification Group (ISG) called Open Radio Interface (ORI) [33]. ORI's goal is to develop an interface specification envisioning

Table 2.1: CPRI bit rate requirements for different antenna configurations.

CPRI Option	Bit rate [Mbps]
1	614.4
2	1228.8
3	2457.6
4	3072.0
5	4915.2
6	6144.0
7	9830.4
8	10137.6
9	12165.12
10	24330.24

interoperability between elements of BSs of cellular mobile network equipment; release four is currently close to approval. The interface defined by the ORI ISG is built on top of the CPRI with the removal of some options and the addition of other functions to reach full interoperability.

Currently, CPRI is, by far, the most adopted specification for fronthaul interface implementation. However, some parts are left vendor-proprietary, thus interoperability of equipment from different vendors is not possible. The main difference between CPRI and OBSAI on one side and ORI on the other side is that the first two groups are composed only by equipment makers, whereas ORI members include also several network operators. In spite of a few differences between CPRI, OBSAI, and ORI, some key common aspects are the following: All BSs are split into two parts connected with the fronthaul interface. The fronthaul most adapted physical layer is an optical fiber.

As mentioned, the requirements of C-RAN can be extremely high, requiring dedicated high speed and low latency connections. CPRI, sets fixed bit rates, depending on the antenna configuration. Its bit rate can be calculated as follows:

$$R_{CPRI} = N_s \times N_{antenna} \times R_s \times 2 \times N_{res} \times O_{cw} \times O_{lc} \qquad (2.1)$$

where N_s and $N_{antenna}$ are the number of sectors and the number of MIMO elements per sector, R_s and N_{res} are the sampling rate and number of bits per sample, O_{cw} and O_{lc} represent the overhead introduced by CPRI control words and line coding overhead. As an example, let's consider an antenna with 3 sectors, 4 MIMO elements, a single 20 [MHz] channel with a sampling rate of 30.72 [MHz], 15 [bits] per sample, $O_{cw} = 16/15$ and $O_{lc} = 66/64$ [byte]. The resulting CPRI rate is $R_{CPRI} = 12165.12$ [Mbps], which corresponds to CPRI option 9, as reported in table 2.1 CPRI also imposes extremely low jitter requirements ($+/-0.002$ [ppm]) to retrieve correctly the clock in the BBU. In addition to this, the Hybrid Automatic Repeat Request (HARQ) mechanism of LTE must be performed by the BBU within a computational time of 3 [ms], leaving around 200 [μs] to transport the data to/from BS site. To relax these requirements, new baseband splits have been investigated recently, like the new eCPRI protocol and different options proposed by 3GPP. With the new splits, some of the functions are left at the BS site, in the RRU, while others are centralized in the BBU, depending on the selected split.

2.5 Advantages and Challenges

The unique design of C-RAN enables it to have several advantages over traditional cellular networks in which BBUs are distributed. A centralized BBU comes with many advantages and some drawbacks which will be discussed in the following:

- The first advantage is the capacity enhancement arising from the fact that C-RAN allows implementing scheduling techniques for interference reduction. The BBUs are provided with a low latency communication channel through which they can jointly contribute to interference reduction. CoMP techniques for interference reduction have been proposed. Strict synchronization and low latency requirements must be satisfied.

- Another advantage is energy savings can be reached by reducing the number of facilities. Energy consumption from air conditioning and power supplies is reduced because of sharing among several BBUs in the hotel. Moreover, a fewer number of BBUs is needed compared to a traditional D-RAN. The virtualization process in BBU hotels allows us to selectively turn off unneeded BBUs without compromising a 24/7 service commitment.

- Besides, the waste of processing resources can be solved through a virtualized hotel solution with load balancing and resource sharing. Load balancing allows overloaded BBUs to migrate the traffic to under-loaded units. Resource sharing allows the overall capacity required in the hotel to be smaller than the sum of the single capacities of the base stations due to the enabled multiplexing gain, therefore the number of BBUs can be reduced.

- Also it is worth mentioning the possibility of implementing various advanced technologies that require high processing and cannot be implemented in traditional networks. As BBUs can be located in powerful data centers and have efficient information exchange, they can do the extensive computation that cannot be done in current networks. As a result, joint processing and cooperative radio sharing technologies will become possible with C-RAN architecture.

Some main challenges need to be addressed to be able to implement C-RAN architecture. Challenges are mainly related to fronthaul traffic requirements, which are introduced to be exchanged between RRUs and BBU hotels.

- One of the main problems is that the bit rates for the traffic transported on the fronthaul links do not scale with the varying traffic load condition of the cell, resulting in fully non-elastic traffic. Given that the one BBU hotel is connected to more than one cell site, the amount of data carried on such fronthaul links will be very huge.

- The other drawback is latency and jitter requirements must be strictly supported, in addition to high bandwidth and cost-efficiency. Stringent timing conditions for some physical layer procedures between BS and UEs are specified by Radio Access Technology (RAT) standards. Most of them explicitly pose bounds on the latency due to the internal processing of radio frames by the BS. In BBU hotelling the BS functions are spread between BBUs and

RRUs, potentially located very far apart from each other, therefore the "fronthaul latency", i.e., the delay contribution due to the transport of fronthaul signals along with the RAN infrastructure, has a relevant impact on the total latency budget inside the BS.

Chapter 3

Design Methodologies for Reliable C-RAN - Centralized Approach

3.1 Introduction

Network survivability and resiliency is a well-established research area for Wavelength Division Multiplexing (WDM) optical networks [34], which are usually adopted to deploy the fronthaul segment and support the high capacities required by Common Public Radio Interface (CPRI). However, the following works mainly focus on the path and link protection/restoration in mesh WDM networks. Several protection schemes have been already proposed for the backhaul part of the 5G networks [35] but no in-depth investigation has been done on the fronthaul part. In Centralized Radio Access Network (C-RAN), all the baseband processing functions are centralized in one or few locations and a failure might have a significant impact on the performance of the network, causing service outage for a large number of users. In particular, the development of a reliable C-RAN to meet capacity and delay requirements for a large number of cells is one of the major challenges.

In the literature, cost and energy-efficient strategies have been proposed to address cost issues while maximizing resource usage in C-RAN [36]. However, all these works do not account for network reliability, which is also one of the key requirements for 5G. Studies concerning optical network resiliency against attacks can be found in [37] while survivability against disasters is discussed in [38]. Reliability of optical devices and related failure studies are conducted in [39]. All these studies are equally applicable also to C-RAN even though they were not though for this architecture. However, they do not account for failures in BBU hotels.

In this chapter, the concept of resiliency for C-RAN will be investigated in two main categories. We first demonstrate the "centralized approach" which needed the supervision of the Software-Defined Network (SDN) controller. The results of this approach in different use cases also shown. The second category is investigating the survivable C-RAN with the Machine Learning (ML) approach where no global information is provided a priori.

3.2 Centralized Method - SDN approach

Software-Defined Networking (SDN) and Network Function Virtualization (NFV) are two promising technologies that are expected to increase the efficiency of 5G networks and enhance the flexibility of network configuration and management [40]. In SDN, a centralized SDN controller, which handles network management operations, is decoupled from the data plane and enables network programmability. Applications and services running on top of 5G networks will take full advantage of the underlying network programmability, communicating with the SDN in a way that optimizes the resource allocation and utilization in a centralized way.

NFV is another cornerstone technology of 5G, which is employed to build an agile and programmable virtualized infrastructure [41]. NFV can provide the infrastructure virtualization which enables virtualizing the 5G RAN. This approach, which is known as Virtualized RAN (vRAN), is recognized as a very promising area of innovation in the 5G ecosystem, resulting in cost reductions and scalability benefits for 5G deployments. Specifically, it allows developing a C-RAN based 5G architecture with low-cost servers [42].

Having an entity such SDN which has the global view of all underlying infrastructures and every single node and link statues at any time gives the strong potential for optimal decision-making algorithms for designing a survivable RAN.

3.2.1 Cost Formulation

The potential flexibility and efficiency offered by the C-RAN architecture need properly defined algorithms to assign the required functionalities to network servers concerning evolving network needs. One aspect that is of primary importance is represented by BBU hotel reliability and protection. Protection of BBU hotel functionalities in case of failure needs to be properly designed. Centralization of BBU functionalities, meaning that several BBUs are physically located in the same node, requires, among the main concerns, that the cost of building a suitable structure to provide cooling and energy system for BBU hotels is maintained low. Moreover, to provide resilience, extra ports or BBU hotels need to be added to the total cost of the resilient network. One of the challenges for network designers is to ensure enough reliability while maintaining both the cost and energy consumption as low as possible. In the solution proposed here, each RRU is provided with a primary lightpath to the main BBU hotel serving as a Digital Unit (DU) and a backup lightpath to a second backup hotel which is activated in case of failure. The switching operation from the primary to the backup hotel is performed through proper signaling by the SDN controller in case of failure.

The survivable fronthaul design problem addressed in this chapter is defined as follows:

- **Given** the physical topology of the WDM mesh transport network, the number of RRUs connected to each transport node, the cost of opening and connecting to a new BBU hotel.

- **Find** the minimum number of BBU hotels, BBU hotel ports, and wavelengths to have full coverage and resilience for all RRUs.
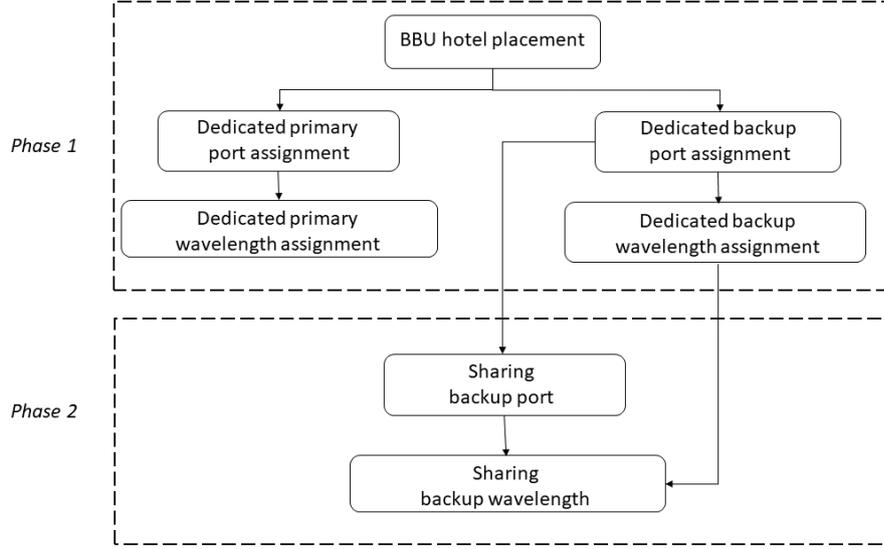
Figure 3.1: Two-phase organization of the resilient design process.

The proposed methodology is organized into two phases as shown in figure 3.1 and described in details in the following:

Phase 1 focuses on BBU hotel placement concerning resilience. The placement requires that each RRU is connected to two separate BBU hotels, one for primary use and the other for backup in case of failure. Since the network has to provide the service for all RRUs at all the times, each RRU will have in this phase a dedicated port both in the primary and in the backup BBU hotels. In addition to resilience for BBU hotel ports, single link failure is also considered in this study. So each RRU will be provided with two lightpaths toward its primary and backup BBU hotels. In case of failure in any segment of the primary lightpath, then the affected RRU can transmit its data using the backup lightpath under SDN control.

Phase 2 has the task of sharing the resources identified in *Phase 1* to increase the overall utilization and save resources. The basic sharing policy is that two or more RRUs can share the same backup port if and only if they have different primary ports located in two different BBU hotels. The reason is that, if two RRUs have their primary ports in the same BBU hotel and failure happens in that hotel, then both RRUs will shift their loads to the same backup port. The same policy is adopted for sharing backup wavelengths. Two RRUs can share the same backup wavelength if and only if they are using two different primary lightpaths. So in case of failure in any part of the primary lightpath, RRUs can use the backup one without conflicting with others.

To calculate the cost, the formulas with the different contributions are here introduced and all the notation will be presented in the table 3.1.

The number of BBU hotels needed to provide both full coverage and resilience is calculated using the following formula:

$$C_B = \sum_{i=1}^{n} B_i \tag{3.1}$$

Table 3.1: Notations used in the formulas.

N	Set of transport nodes, $\|N\| = n$
L	Set of optical links, $\|L\| = l$
B	Set of active BBU hotels, $\|B\| = b$
h	Distance in hops between each pair of BBU hotel and RRU.
C_B	Total number of BBU hotels.
B_i	1 if node $i \in N$ host a BBU hotel, 0 otherwise.
C_P	Average number of ports.
PP_{ij}	1 if BBU hotel $i \in B$ is assigned a primary port to the RRUs connected to node $j \in N$, 0 otherwise.
BP_{ij}	1 if BBU hotel $i \in B$ is assigned a backup port to the RRUs connected to node $j \in N$, 0 otherwise.
C_W	Average number of wavelengths.
PW_{ij}	1 if link $i \in L$ contains a primary wavelength assigned to the RRUs connected to node $j \in N$, 0 otherwise.
BW_{ij}	1 if link $i \in L$ contains a backup wavelength assigned to the RRUs connected to node $j \in N$, 0 otherwise.

where B_i is a boolean variable equal to 1 when the node is set as a BBU hotel, that is it hosts BBU functionalities related to possible multiple RRUs.

A further cost parameter is represented by the number of ports needed to support primary and backup functionalities. A value averaged over the total number of nodes, namely C_P, is calculated by the following formula, which considers the total number of primary ports in addition to shared backup ports resulting after *Phase 2*:

$$C_P = \frac{\sum_{i=1}^{b} \sum_{j=1}^{n} PP_{ij} + \sum_{i=1}^{b} \sum_{j=1}^{n} BP_{ij}}{n} \quad (3.2)$$

Finally, the average number of wavelengths needed to support BBU hotel reliability is calculated as C_W:

$$C_W = \frac{\sum_{i=1}^{l} \sum_{j=1}^{n} PW_{ij} + \sum_{i=1}^{l} \sum_{j=1}^{n} BW_{ij}}{l} \quad (3.3)$$

Two different sets of algorithms are described in this chapter to implement *Phase 1* of designing a survivable C-RAN, namely BBU hotel location. The first is the Fixed Distance Algorithm (FDA) which has the constraint of a maximum distance between BBU hotels and RRUs. As a consequence, the placement solution performed for *Phase 1* will guarantee that each RRU will find both primary and backup BBU hotels within a given distance and, consequently, with a possible bounded delay. As a drawback, the solution is expected to be characterized by a quite large number of BBU hotels to ensure protection.

The second is the Variable Distance Algorithm (VDA) which is based on the Facility Location Problem (FLP) [43]. These algorithms are applied to networking contexts to find the optimal location for network functions, given a set of possible nodes, under cost constraints. The benefit of this approach is that the overall cost of deploying resilient BBU hotel placement is minimum even though no guarantee

is given to RRU to find either a primary or a backup BBU hotel within a given distance. The objective of the VDA algorithm is to minimize the total cost of the deployment.

3.3 Fixed Distance Algorithm (FDA)

These algorithms performing BBU hotel placement under the assumption of a maximum distance between BBU hotels and RRUs. As a consequence, the placement solution performed in *Phase 1* will guarantee that each RRU will find both primary and backup BBU hotels within a certain distance and, consequently, with a possible bounded delay. As a drawback, the solution is expected to be characterized by a quite large number of BBU hotels to ensure protection.

The assignment procedure can start by assigning the Primary BBU Hotel (P) or the Backup BBU Hotel (B) first, and then by proceeding further. How this choice impacts on resulting costs will be evaluated. Besides, the starting nodes has also an impact on the total number of BBU hotels, depending on network topology. Two extreme different option will be considered, namely *Max-D* when the algorithm starts from the node with the highest nodal degree, and *Min-D* when instead it starts from the node with the lowest nodal degree. As a consequence of all possible combinations, we will have for *Phase 1* the options indicated as *Min-D-P*, *Max-D-P*, *Min-D-B*, *Max-D-B*. In the following FDA *Phase 1* placement algorithms will be described.

3.3.1 Primary BBU Hotel First (P)

The objective of this methodology, in addition, to minimize the total number of BBU hotels, is to prioritize the connectivity between each RRU and its primary BBU hotel. Figure 3.2 shows the procedure to find the best placement for BBU hotels so that all RRUs have access to their primary ports, and then by applying the resilient placement, either by associating backup functionalities to already connected BBU hotels or by adding extra BBU hotels. A node is first chosen based on *Max-D* or *Min-D* policy. Once a new node i is selected to be a candidate host for a BBU hotel, the strategy then checks all nodes j that are within certain distance h from node i and that can be reached using a transparent lightpath.

P or *B* methodology can be adopted as shown in boxes (a) or (b) in figure 3.2 in dashed line. This part is the only difference in the assignment algorithm between *P* and *B* algorithms. In case of the *P* approach, if a primary port (and wavelength) for RRUs connected to node j has not been assigned yet, the BBU hotel i is mapped as node j primary BBU hotel, otherwise, in case of node j has already a primary port, node i is set as its backup BBU hotel and backup port (and wavelength) is assigned according. The opposite happens for the *B* approach. Once all nodes j within the h hop distance are checked, node i is no longer considered and another node in the set of possible locations for BBU hotels is chosen. These operations are repeated until all RRUs are assigned primary and backup BBU hotels.
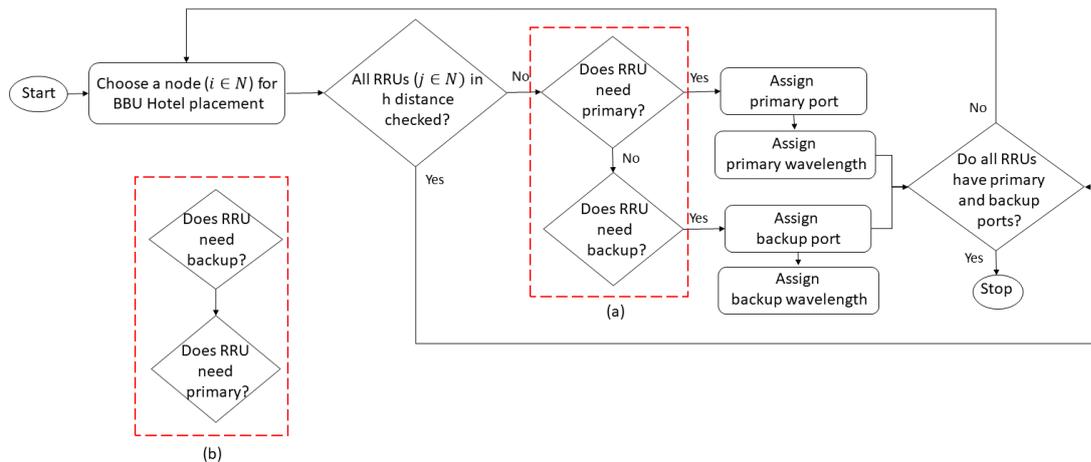
Figure 3.2: Fixed Distance Algorithm [FDA], the case of Primary BBU Hotel First (P).

3.3.2   Backup BBU Hotel First (B)

This methodology is investigated to maximize the sharing of backup BBU hotel ports. The flow chart is shown again in figure 3.2 including box (b) instead of (a). Again the starting point impacts results, so both $Max - D - B$ and $Min - D - B$ will be considered.

With B algorithms a transport node i is chosen based on the *Max-D* or *Min-D* policy. Then this node is checked as backup BBU hotel first for all nodes j connected within the distance h. If node j has already been assigned a backup BBU hotel then node i will be connected to j as its primary BBU hotel. These operations are repeated until all RRUs are assigned a backup and a primary BBU hotel.

3.3.3   Fixed Distance Algorithm Results

This section presents the performance analysis of the survivable BBU hotel placement strategies. The results are obtained using a Java-based simulator. The reference topology of the optical transport network considered for the performance assessment is presented in figure 3.3 [44]. It consists of a metro/aggregation network with 38 nodes and 59 bidirectional fiber links, all with the same length ($N1$). To evaluate the effect of network size and topology on the results, different configurations are considered based on reduced-size versions of the original one with 20 ($N2$) and 14 ($N3$) nodes, as indicated in the figure by dashed and dashed/dotted lines, respectively. Each node in the transport network is assumed to serve the same upstream traffic represented by 10 RRUs connected to it, each one requiring two transparent lightpaths, i.e., one connecting the RRU to the primary and one connecting the RRU to the backup BBU hotel.

Figure 3.4 shows the comparison of the total number of BBU hotels (C_B) as a function of the distance (h) in hops for the different approaches applied in FD algorithms, for $N1$, $N2$, and $N3$ networks. As can be seen in each set of results, the *Min-D* approach requires a higher number of BBU hotels in comparison to the *Max-D* one. All trends are decreasing by relaxing the distance constraint, that is by
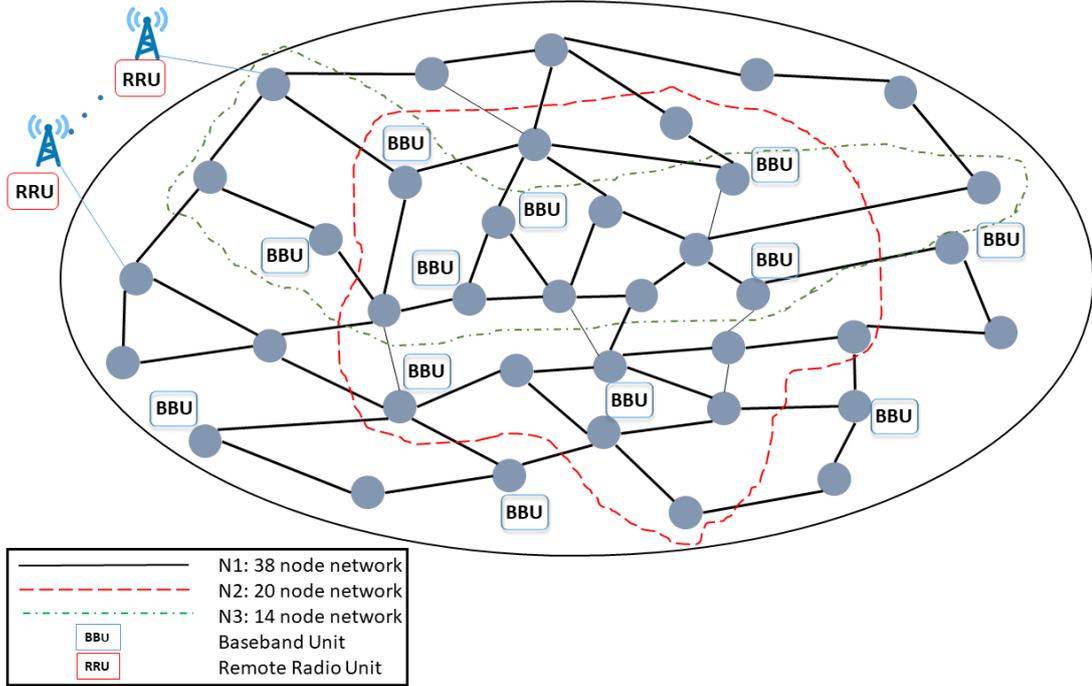
Figure 3.3: N1, N2 and N3 fronthaul network topologies used in the evaluations.

increasing h. In the case of the $N3$ network, the behavior for both Min-D and Max-D approaches is the same after 4 hops. The reason is the small size of the topology and the fact that after 4 hops the algorithm can find in any case the best solution which is two BBU hotels with both methods. In the $N2$ network, small variations between the Min-D and Max-D approaches are present but they are always very close to each other so that these variations can be related to the effect of the topology.

Besides, to minimize the number of BBU hotels in the network, the other goal of the design procedure is to maximize the sharing of the BBU hotel ports among RRUs for protection. Figure 3.5, 3.6 and 3.7 report the comparison between the average number of ports per node (C_P(as a function of the distance h in hops in different network topologies, namely $N1$, $N2$, and $N3$. The average number of ports per node is calculated by formula 3.2, which is the sum of the total number of primary ports plus the total number of shared backup ports, that are assumed to serve multiple RRUs for protection purpose.

Figure 3.5 shows the results for the $N1$ network, with the best results achieved when Backup First (B) and Max-D approaches are applied. The reason is that by assigning backup BBU hotels first the possibility of better sharing is allowed, especially when using the Max-D node first which allows covering more RRUs with fewer BBU hotels. Following the same reasoning, the worst result is obtained with combined Primary First (P) and Min-D approaches.

Figure 3.6 shows the average number of ports per node (C_P(as a function of the distance constraint h for the $N2$ network. The best approach is achieved by combining Max-D and Backup First (B). All trends are increasing at h. The reason behind this is the opposite of figure 3.4. When the distance constraint is more relaxed, the number of BBU hotels needed for protection is decreasing and the chance of sharing BBU hotel ports decreases too. That is why the total number of ports increases. The extreme case is for the primary approach after 6 hops: no
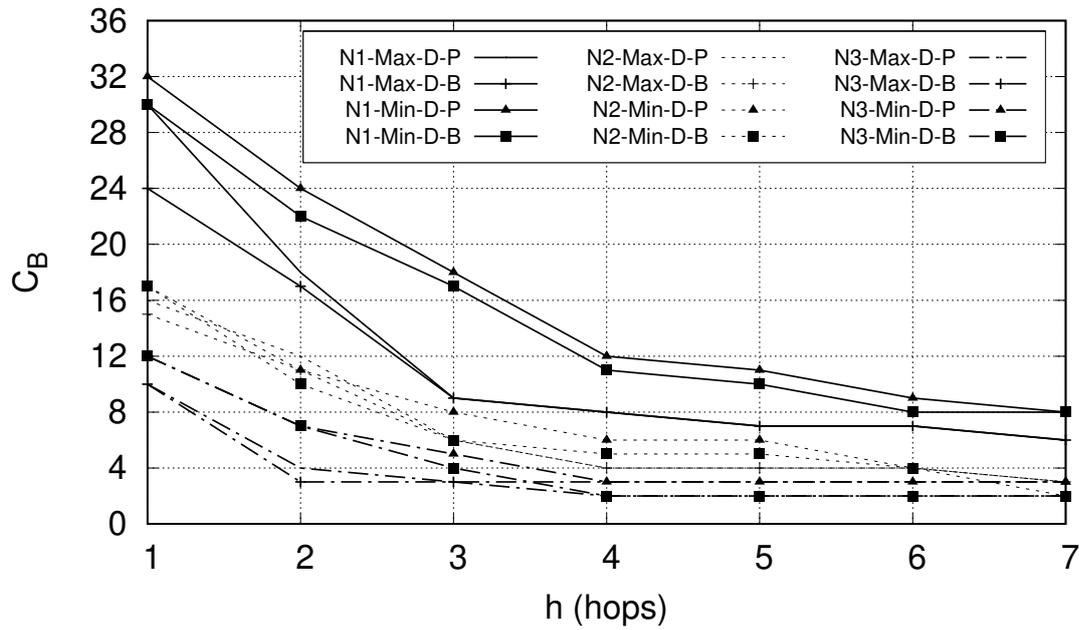
Figure 3.4: The number of BBU hotels (C_B(as a function of the maximum distance between an RRU and a BBU hotel h in $N1$, $N2$ and $N3$ network topologies, comparing *Min-D* and *Max-D* combined with P and B techniques.
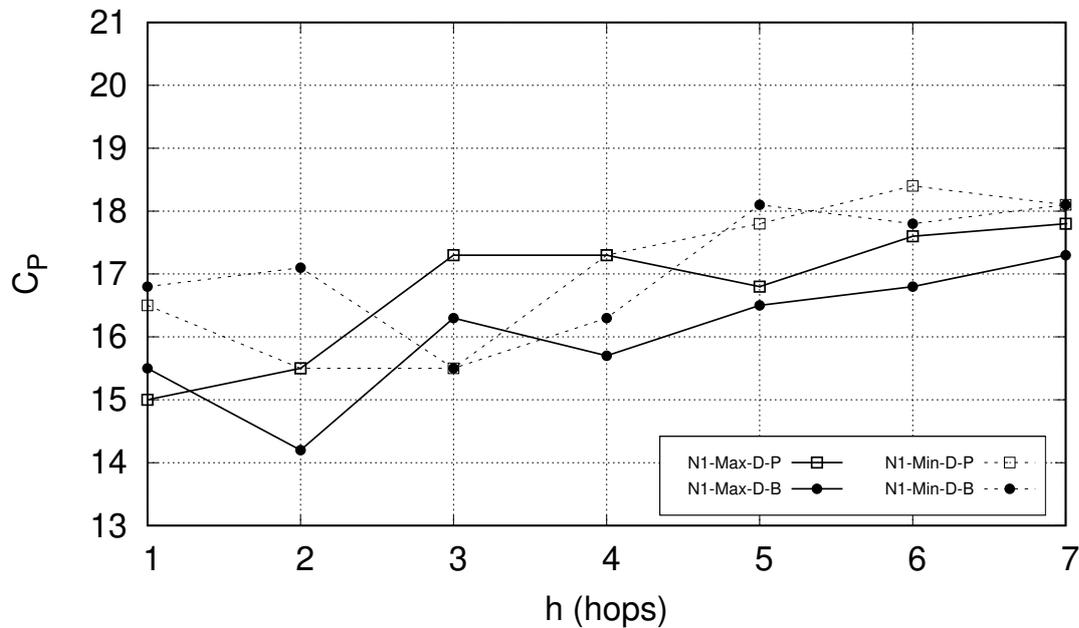
Figure 3.5: The total number of ports, averaged to the number of nodes (C_p), as a function of the maximum distance between an RRU and a BBU hotel h, comparing *Min-D* and *Max-D* combined with P and B techniques in $N1$ network.
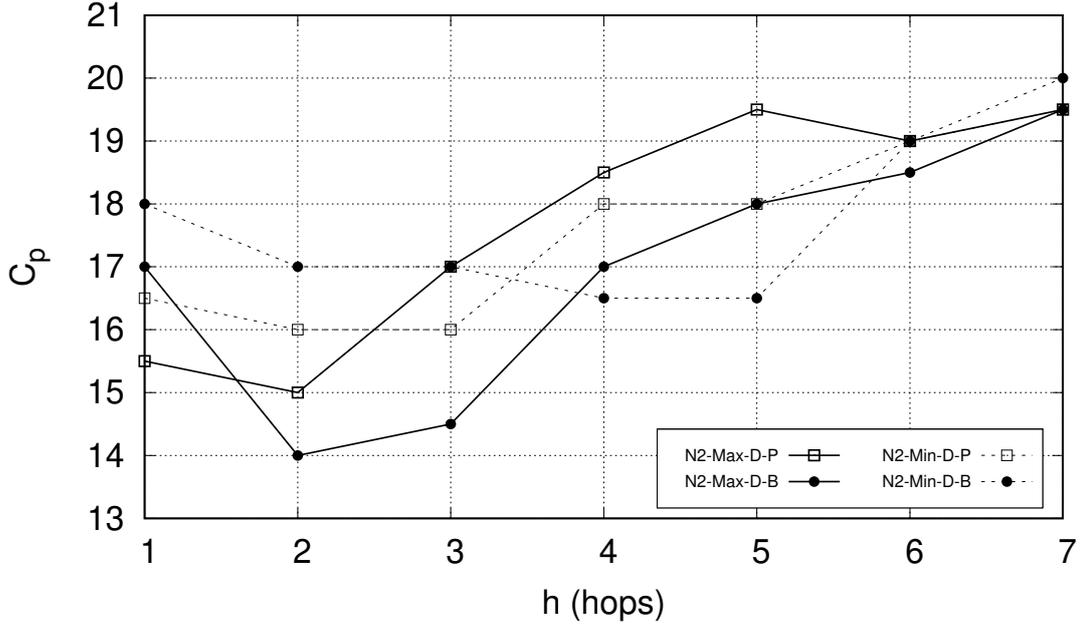
Figure 3.6: The total number of ports averaged to the number of nodes (C_p), as a function of the maximum distance between an RRU and a BBU Hotel h, comparing *Min-D* and *Max-D* combined with P and B techniques in $N2$ network.

sharing technique is effective here and consequently, the number of backup ports is the same as the primary ports.

The same comparison for the $N3$ network leads to slightly different results which are reported in figure 3.7. All the trends are increasing by relaxing the distance constraint. In the 3 initial hops, all the trends varying very close to each other and the best results belong to *Min-D* with the combination of the Backup First (B) approach. After 3 hops in the case of Primary First (P) and 4 hops in case of Backup First (B), no better-sharing results can be achieved due to the small size of the network.

To represent the effect of the size of the topology on the FD algorithms, in figure 3.8, the best approach in the largest and smallest network topologies are taken into account. The best sharing technique results are obtained with *Max-D-B* for the largest size of the network, namely the $N1$ network. This shows that the sharing algorithm is more effective as the size of the network increases.

The further set of results is related to the number of wavelengths needed to support the network configuration with primary and backup BBU hotels, after the assignment. The average number of wavelengths per link (C_W(is reported in figure 3.9 as a function of distance (h) in hops. All the trends are increasing by relaxing the distance constraint. The reason is that, by increasing the distance, the number of BBU hotels decreases and therefore each RRU needs more wavelengths connect to the primary and backup BBU hotels. In general, all four different approaches show the same behavior with few hops allowed (one or two hops). Afterward, the gap between them increases and the best behavior is shown by the *Max-D*. So another important conclusion is that the *Max-D* approach provides, at the same time, the least number of BBU hotels and the least amount of wavelengths, compared to other approaches.
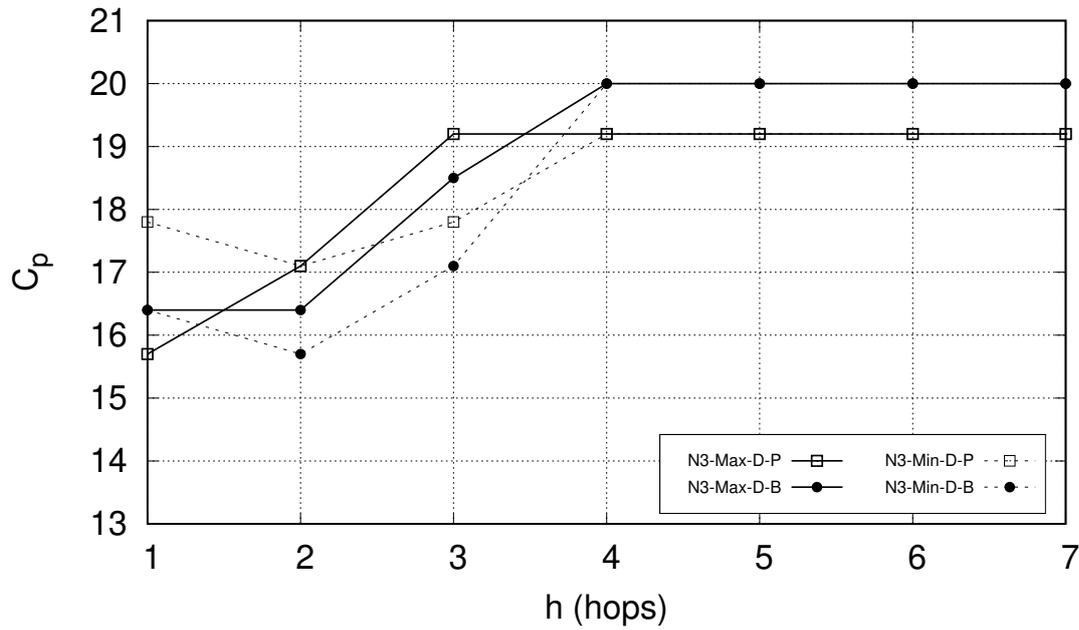
Figure 3.7: The total number of ports, averaged to the number of nodes C_p, as a function of the maximum distance between an RRU and a BBU hotel h, comparing *Min-D* and *Max-D* combined with P and B techniques in $N3$ network.
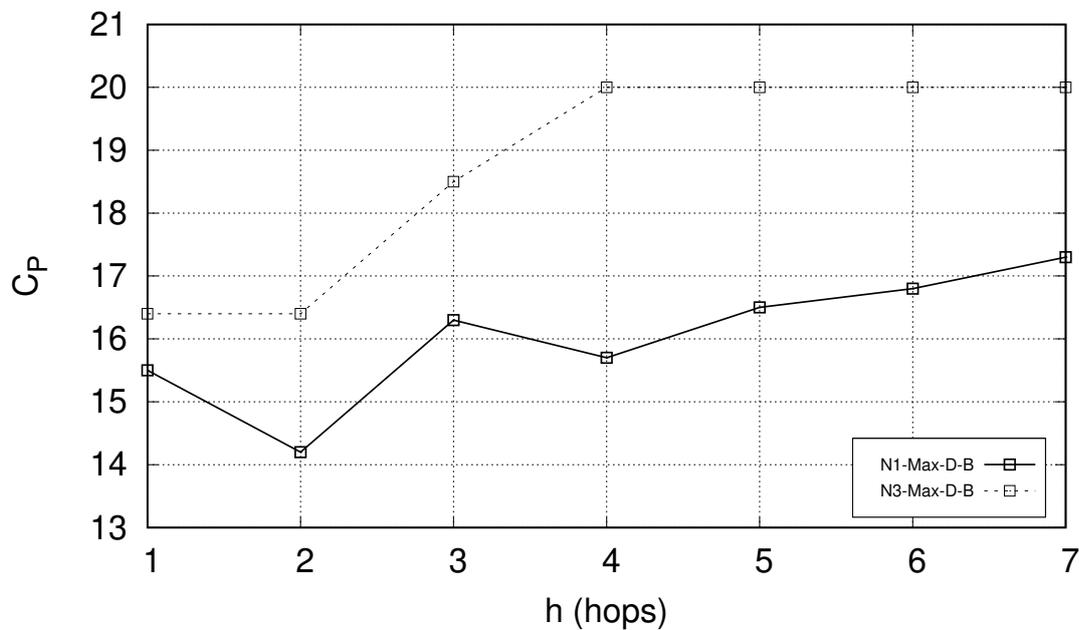
Figure 3.8: The total number of ports, averaged to the number of nodes (C_P), as a function of the maximum distance between a RRU and a BBU hotel (h), comparing *Max-D-B* technique in the largest $N1$ and the smallest $N3$ size of the networks.
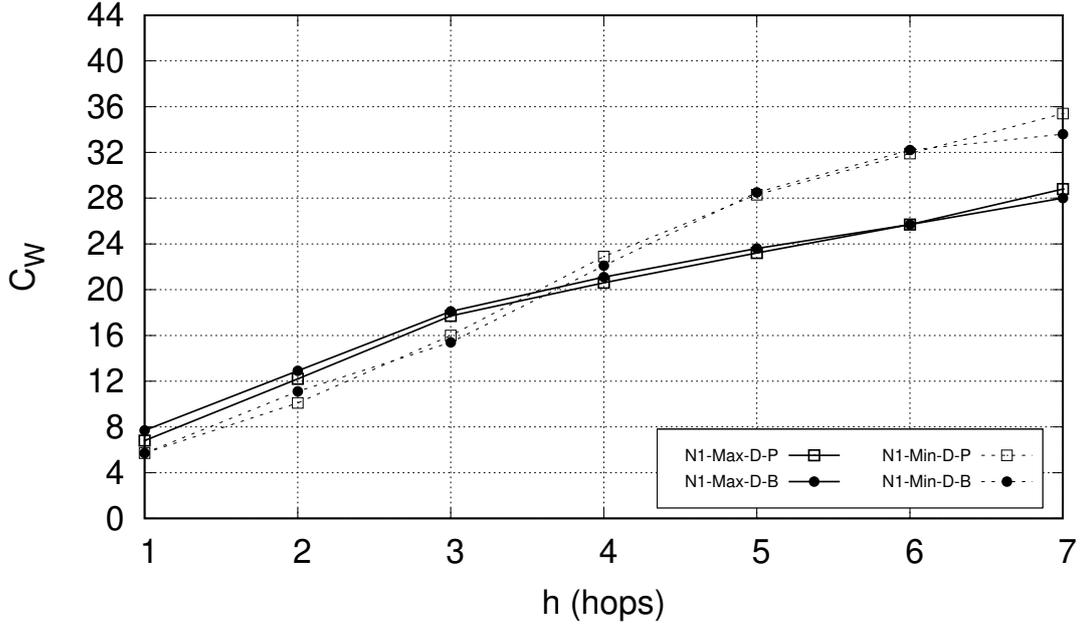
Figure 3.9: The total number of wavelengths, averaged to the number of links (C_W), as a function of the maximum distance between an RRU and a BBU hotel (h), comparing *Min-D* and *Max-D* combined with P and B techniques in $N1$ network.

Figure 3.10 shows the comparison of the average number of wavelengths (C_W) as a function of distance (h) in hops for the $N2$ network. As for figure 3.10, all the approaches have very close values when the number of allowed hops is low. As the distance constraint (h) is increasingly relaxed, the gap between graphs increases, having *Max-D* and *Min-D* curves almost the same behavior, either the primary or the backup approach is adopted. The least number of wavelengths is given by the Primary Fist (P) and *Max-D* approach combined and, in second place with small differences, by the Backup First (B) and *Max-D* approach. The reason is the same as for ports due to the smaller size of topology and also to the effect of the topology itself.

Figure 3.11 shows the same comparisons of figure 3.10 but this time for the smallest size of networks ($N3$). As can be expected, after 3 hops for all approaches there will be no improvement in sharing wavelengths. An explanation for this behavior could be the steady number of BBU hotels in figure 3.4 in a higher number of hops. So the algorithm shows no further improvement also for sharing the number of wavelengths. The best approach which could obtain the most sharing in terms of the number of wavelengths is the *Max-D* approach. Both Primary First (P) and Backup First (B) showing almost the same values after 4 hops. As a conclusion, the algorithm seems to perform better when the network is larger.

To complete the evaluation of the different approaches in relation to the size of the topology, the results for the *Max-D* technique, are represented in the same figure. Figure 3.12 shows the comparison of the average number of wavelengths (C_W) as a function of distance (h) in hops for all three networks ($N1$, $N2$, and $N3$). For 4 initial hops, all the trends are very similar and very close to each other. After 4 hops the difference increases as a consequence of the relaxing of the distance constraint. Having a smaller size network leads to fewer needs for wavelengths.
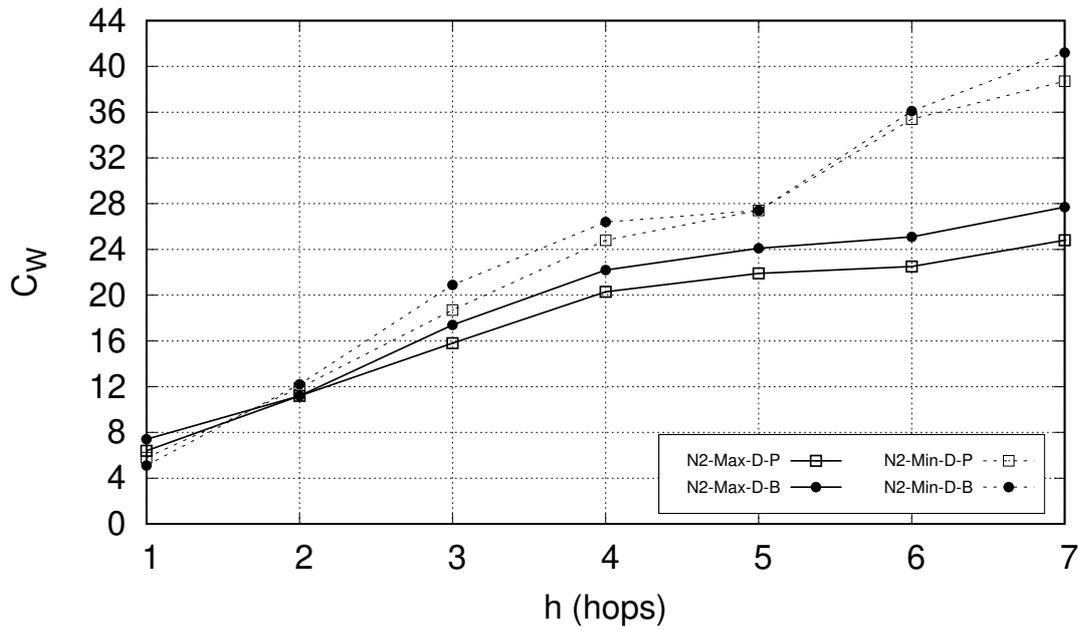
Figure 3.10: The total number of wavelengths, averaged to the number of links (C_W), as a function of the maximum distance between an RRU and a BBU hotel (h), comparing *Min-D* and *Max-D* combined with P and B techniques in $N2$ network.
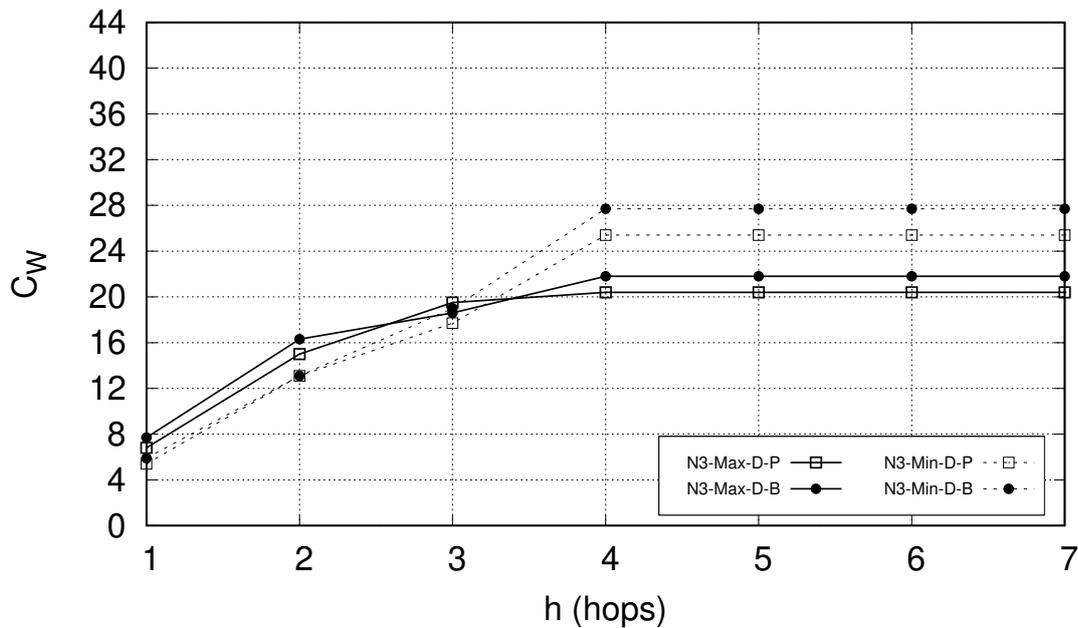
Figure 3.11: The total number of wavelengths, averaged to the number of links (C_W), as a function of the maximum distance between an RRU and a BBU hotel (h), comparing *Min-D* and *Max-D* combined with P and B techniques in $N3$ network.
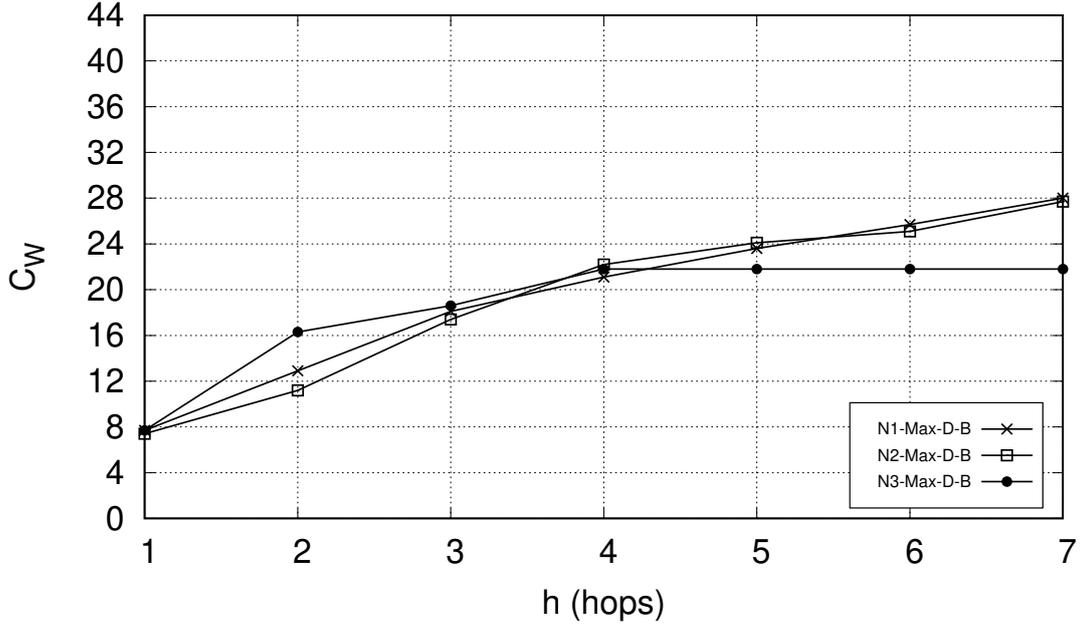
Figure 3.12: The total number of wavelengths, averaged to the number of links (C_W), as a function of the maximum distance between an RRU and a BBU hotel (h), comparing *Max-D-B* technique in $N1$, $N2$, and $N3$ networks.

As mentioned before, minimizing all network resources at the same time is not possible. They result to be in an opposite relation with each other meaning optimizing one leads to overusing the other. Figure 3.13 shows the relation between the total number of BBU hotels (C_B) needed and the average number of wavelengths (C_W) in the $N1$ network. The plot compares the different methodologies previously explained. Besides, the values regarding the number of needed BBU hotels and the average number of wavelengths in case of no protection are shown under the name of *Max-D-WP-BBU* and *Max-D-WP-Wave*. Figure 3.13 is the clear evidence of the fact that by relaxing the distance constraint, there will be fewer BBU hotels and more wavelengths are needed to cover all the RRUs in the network. This can be seen also in the case of no protection.

Figure 3.14 shows the comparison between the total number of BBU hotels (C_B) and the average number of wavelengths (C_W) as a function of the distance constraint (h) in $N2$ network. As in the $N1$ network, the opposite relationship between the number of BBU hotels and the number of wavelengths can be seen. The results without protection are also plotted as a reference. *Max-D-WP-BBU* has much fewer values for BBU hotels at the few first hops in comparison with the other scheme. By relaxing the distance constraint, all trends become closer to each other. This shows that the protection schemes not only can provide full protection but also do so at limited additional cost.

Comparison between the total number of BBU hotels (C_B) and the average number of wavelengths (C_W) as a function of h constraint for $N3$ network can be seen in figure 3.15. As the figure reported, as long as the number of BBU hotels decreases due to relaxed constraints on distance, the average number of wavelengths increases respectively. As can be expected the WP approaches having the least values both for the number of BBU hotels and the average number of wavelengths
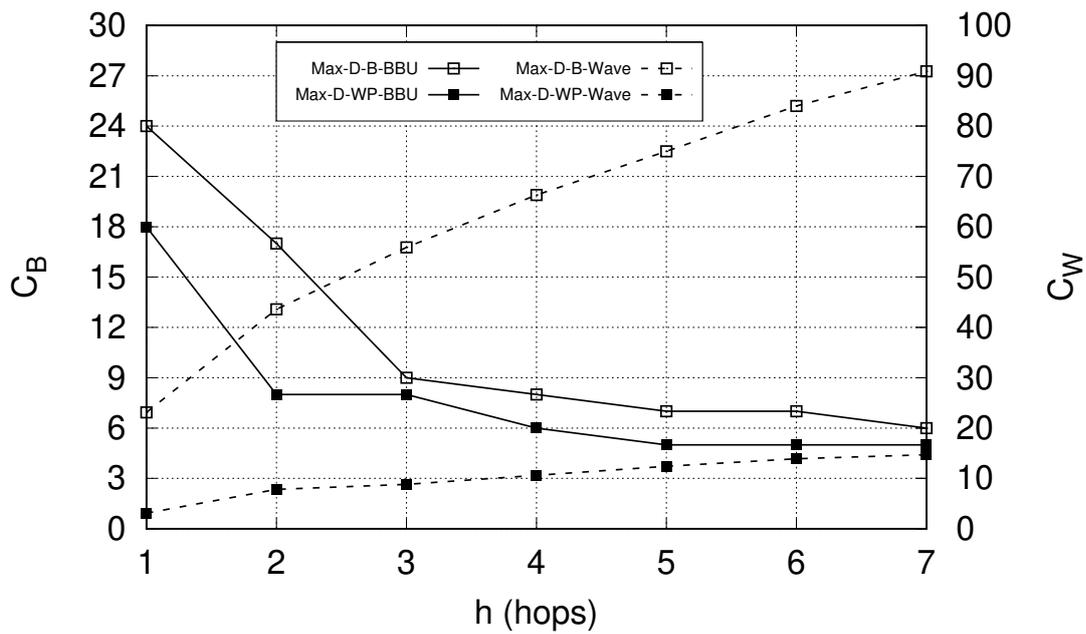
Figure 3.13: The total number of BBU hotels (C_B) and the average number of wavelengths per link (C_W) both as functions of the distance constraint (h), comparing *Max-D* technique with respect to the case of without protection (WP) in the $N1$ network.
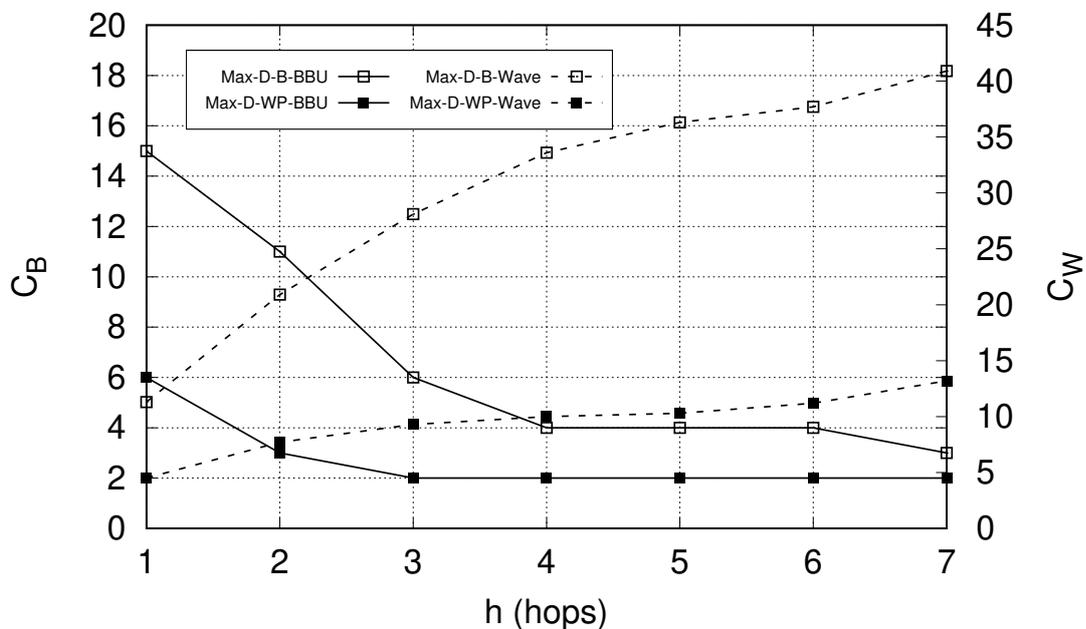
Figure 3.14: The total number of BBU hotels (C_B) and the average number of wavelengths per link (C_W) both as functions of the distance constraint (h), comparing *Max-D* technique with respect to the case of without protection (WP) in the $N2$ network.
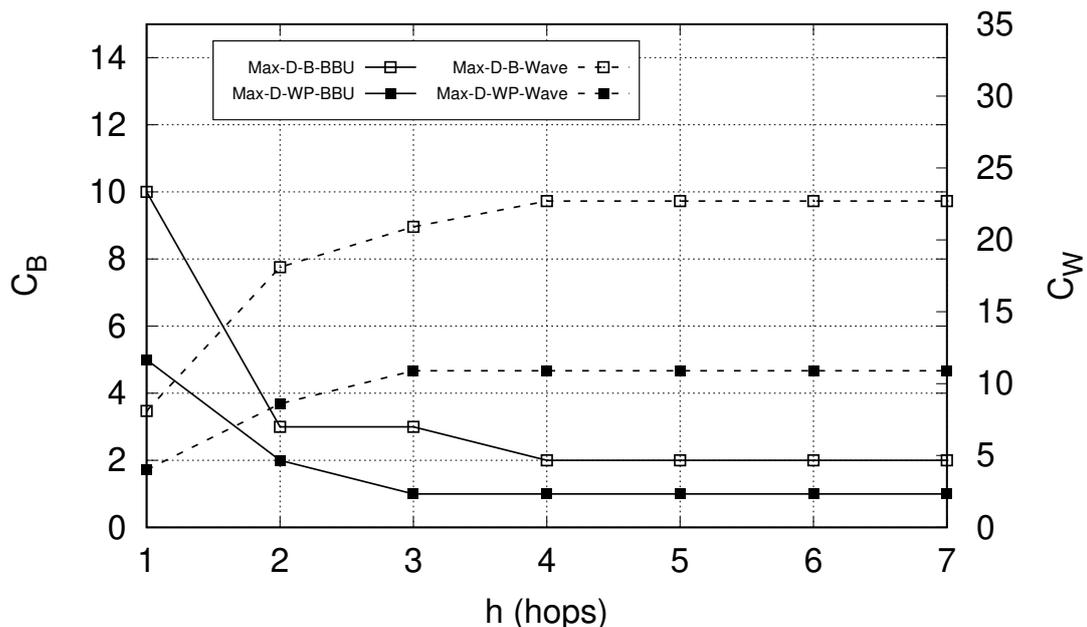
Figure 3.15: The total number of BBU hotels (C_B) and the average number of wavelengths per link (C_W) both as functions of the distance constraint (h), comparing *Max-D* technique with respect to the case of without protection WP in the $N3$ network.

but by relaxing the distance all trends come closer.

3.4 Variable Distance Algorithm (VDA)

Variable Distance Algorithm (VDA) is based on the Facility (or Node) Location Problem (FLP) which is applied to networking contexts to find the optimal location for network functions, given a set of possible nodes, under cost constraints. The algorithm for node location reported in [45] is extended here to propose the VDA algorithm by considering also the location of backup functions, in addition to primary functions, while choosing the BBU hotels within the set of transport nodes in the fronthaul network. The benefit of this approach is that the overall cost of deploying resilient BBU hotel placement is minimum even though no guarantee is given to RRU to find either a primary or a backup BBU hotel within a given distance.

In this part, the classical FLP presented in [46] and [47] is extended by introducing the concept of resiliency against single BBU hotel failure. Different design methodologies for survivable C-RAN architectures based on heuristic and an Integer Linear Programming (ILP) are proposed. The main objective of the study is to find the optimal placement for the BBU hotels to have protected service for RRUs while minimizing the total distance between RRUs and BBUs. The minimization of backup BBUs and the related deployment are also discussed.

In this part, due to the change in the algorithm, the previous formulas have been modified. All the new notations are explained in the table 3.2. The activation cost of BBU hotels needed to provide full coverage and resiliency of the target area is calculated using the following formula:

$$C_B = \sum_{i=1}^{s} B_i \beta_i \tag{3.4}$$

where B_i is a boolean variable equal to 1 when the node is set as a BBU hotel, that is when it hosts BBU functionalities related to one or more RRUs. β_i is a parameter associated with the activation cost for a BBU hotel in node i.

To provide reliability against single BBU hotel failure, it is sufficient to ensure that each RRU is connected to two BBU ports placed in different BBU hotels, one in the primary and one in the backup hotel. The overall distance between BBU hotels and RRUs connecting to the transport nodes in the network, considering both primary and backup hotels, is denoted as D_H:

$$D_H = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} h_{ij} + \sum_{i=1}^{s} \sum_{j=1}^{s} b_{ij} h_{ij} \tag{3.5}$$

where p_{ij} and b_{ij} are boolean variables that indicate if hotel i is assigned as primary or a backup for the group of RRUs at transport node j. h_{ij} represents the distance, in hops, between transport node i and j computed solving the shortest path problem. By multiplying equation (3.5) by the parameter α, the total cost for the distance is achieved:

$$C_H = D_H \alpha \tag{3.6}$$

Finally, to solve the problem, the proper number of BBU ports must be allocated in each hotel. The total number of primary and backup BBU ports and the related cost are calculated according to the following formulas:

$$P = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = P_P + P_B \tag{3.7}$$

$$C_P = P\gamma \tag{3.8}$$

where P_P and P_B are the total number of primary and backup ports respectively. C_P is the contribution of the total number of ports in each hotel multiplied by the cost parameter γ associated with each port.

Since the protection requires that each RRU is connected to two different BBU hotels, the total number of ports should be twice the number of RRUs, and consequently, the value for C_P can be fixed. However, only P_P is fixed, while P_B can be reduced. If exist RRUs have separate primary BBU hotels, they can share backup ports due to the single failure assumption done in this work. By sharing the backup ports among RRUs the value for C_P can be reduced, and further cost saving can be achieved.

In the following, two solutions for survivable fronthaul design are presented. First, the problem is solved by the heuristic and in the next subsection, an ILP formulation is introduced for comparison. All the notations used for the pseudo-codes are summarized in table 3.2.

Table 3.2: Notations used in the formulas and the VDA procedure.

N	Set of transport nodes, $|N| = n$
D_H	The overall distance between BBU hotels and RRUs.
p_{ij}	1 if BBU hotel i is assigned as primary for RRUs at node j, 0 otherwise.
b_{ij}	1 if BBU hotel i is assigned as backup for RRUs at node j, 0 otherwise.
H	$s \times s$ matrix. h_{ij} is the distance in hops between i and j computed with the shortest path.
C_H	The cost of overall distance between BBU hotels and RRUs.
P	Total number of required BBU ports.
x_i	Number of BBU ports required at hotel site i for primary purposes.
y_i	Number of BBU ports required at hotel site i for backup purposes.
P_p	Total number of primary ports.
P_B	Total number of backup ports.
C_P	The cost of overall ports in BBU hotels.
C	Set of nodes $i \in N$ considered as possible host for BBU hotel
B	Set of transport nodes hosting a BBU hotel, $|B| = b$
λ_i	Cost of opening a new BBU hotel in node $i \in C$
μ_{ij}	Cost of connecting the RRUs connected to node $j \in N$ to BBU hotel $i \in B$
F	Total cost
$Pconn_{ij}$	1 if node $i \in B$ is the primary BBU hotel for the RRUs connected to node $j \in N$, 0 otherwise
$Bconn_{ij}$	1 if node $i \in B$ is the backup BBU hotel for the RRUs connected to node $j \in N$, 0 otherwise
RRU_b	Set of nodes $j \in N$ whose RRUs have been assigned a backup BBU hotel $i \in B$
D_i	Nodal degree of node $i \in N$
BL_{ij}	1 if link $i \in L$ is assigned a backup wavelength to the connected to node $j \in N$, 0 otherwise
P_{Hotels}	Set of primary BBU hotels
H	$s \times s$ matrix. h_{ij} is the distance in hops between nodes i and j computed with the shortest path.
α	Weight of the hops in the cost function F.
β_i	Weight of the active BBU hotel i in the cost function F.
γ	Weight of the BBU hotel ports in the cost function G.

3.4.1 Heuristic

The facility location problem with protection aims at connecting $j \in N$ transport nodes, each containing a given amount of RRUs, through a list of possible $i \in C$ BBU hotel locations so that the total cost F is minimum. As for cost information, μ_{ij} is considered as the distance in hops between each RRU and BBU hotel pairs, while the cost λ_i, being it equal for all new BBU hotels, is not considered in the location procedure. The procedure starts by randomly choosing the candidate node

for hosting a BBU hotel $i \in C$ in line 4. The reason for choosing randomly is due to the fact that, differently from the FD approach, the starting point does not impact on the outcome of this procedure.

After opening a new BBU hotel in node i (line 5), in lines 6 to 8 all nodes $j \in N$ will be connected to it ($P_{conn_{ij}} = 1$) and node i will be considered as a primary BBU hotel for all of them (with dedicated primary ports and wavelengths). By having only one BBU hotel in the network and calculating F the worst-case cost for node i will be achieved (line 9).

The rest of the procedure aims to reduce the cost of F by adding further BBU hotels to the network. In the "for" loop starting at line 10, the procedure searching for a new placement, in addition to BBU hotel i. If placement with reduced cost exists, namely node $i' \in C$ (line 11), a new BBU hotel will be open in i' and the total cost will be updated accordingly (lines 12 and 13). In addition, those RRUs involved in the cost reduction in lines 14 to 16, will be disconnected from their former BBU hotel i and connected to the new BBU hotel i'. If there is no new location that exists such that by opening a new BBU hotel the total cost reduced, then the procedure achieved the lowest cost placement without protection, meaning that each RRU at this point is only connected to its primary BBU hotel.

The first step towards protecting this VD approach is, for each RRU search, to find another BBU hotel namely $i" \in B$, already open, in one hop distance ($h = 1$), different from the primary one, to obtain the lowest cost (line 20). If found, the second BBU hotel $i"$ will be considered as a backup hotel and backup ports and wavelengths will be assigned (lines 21 and 22). Also node j will be added in a set RRU_b which is the set of nodes in which their RRUs have been assigned both primary and backup BBU hotel ports (line 22). This step checks the possibility of using currently deployed resources to protect without adding any extra cost in terms of BBU hotels. In case, after this step, there are still some RRUs without protection, then the second check loop will be launched. This loop starts in line 26, searches for a maximum node degree ($k \in C$) which does not host a BBU hotel and opens a new hotel in that location to protect some RRUs.

Upon finding a node (k), the procedure will add it to the BBU hotel set and update the cost in lines 27 and 28. Backup ports and wavelengths will be assigned accordingly to all nodes $j \in N$ which are in one hop distance from the new BBU hotel k (lines 29-32). This loop will be repeated until all RRUs will be assigned their backup BBU hotels. The worst-case complexity of the VDA BBU hotel Placement procedure is estimated as $\mathcal{O}(N^3)$.

3.4.2 ILP Optimization

The core of our problem is based on the ILP formulation of the FLP introduced in [46]. The formulation in [46] has been modified to provide protection, using backup hotels, and to include the effects of BBU hotel ports. The problem is here formulated in such a way that, by properly tuning the parameter of the objective function, BBU ports can be minimized while solving the survivable fronthaul design problem.

Additional parameters:

- r_j number of RRUs at site j.

- M a large number.

Algorithm 1 Variable Distance BBU Hotel Placement

1: **Initialization:**
2: $C = N$
3: $B, F, RRU_b \leftarrow 0$
4: **Begin:**
5: //BBU hotel Placement procedure
6: **find** node $i \in C$ randomly
7: $B \leftarrow B \cup i$
8: **for** all nodes $j \in N$
9: $\quad Pconn_{ij} = 1$
10: **end for**
11: calculate F
12: **for** all nodes $i' \in (C - B)$
13: $\quad$ **if** exists $(F'_i < F_i)$
14: $\qquad B \leftarrow B \cup i'$
15: $\qquad$ update F
16: $\qquad$ **for** $j \in N$ such that $\mu_{i'j} < \mu_{ij}$
17: $\qquad\quad Pconn_{i'j} = 1$
18: $\qquad$ **end for**
19: $\quad$ **end if**
20: **end for**
21: //BBU hotel protection procedure
22: **for** each node $j \in N$
23: $\quad$ **if** exists a BBU hotel $i" \in B$ in $h = 1$ and $Pconn_{i"j} \neq 1$
24: $\qquad Bconn_{i"j} = 1$
25: $\qquad RRU_b \leftarrow RRU_b \cup j$
26: $\quad$ **end if**
27: **end for**
28: **while** $RRU_b \neq N$ **do**
29: $\quad$ **find** node $k \in C$ such that D_k is maximum and $k \neq B$
30: $\quad B \leftarrow B \cup k$
31: $\quad$ update F
32: $\quad$ **for** all nodes $j \in N$ in $h = 1$ from k s.t. $Bconn_{kj} \neq 1$
33: $\qquad Bconn_{kj} = 1$
34: $\qquad RRU_b \leftarrow RRU_b \cup j$
35: $\quad$ **end for**
36: $\quad$ remove k from C
37: **end while**
38: **Stop**

Additional variables:

- $c_{j,i,i'} = 1$ if source j is using destination i as primary and i' as backup hotel site; 0 otherwise.

Objective function:

$$Minimize\ G = C_B + C_H + C_P \tag{3.9}$$

The multi-objective function 3.9 is composed of three members. The first term takes into account the activation cost of each hotel (C_B). The second term accounts for the cost to connect RRUs to BBU hotels, both primary and backup (C_H) while the third term accounts for the cost of BBU ports required in each hotel (C_P).

The problem is subject to the following constraints:

$$\sum_{i \in N} p_{i,j} = 1, \forall j \in N \tag{3.10}$$

$$\sum_{i \in N} b_{i,j} = 1, \forall j \in N \tag{3.11}$$

$$p_{i,j} + b_{i,j} \leq 1, \forall i,j \in N \tag{3.12}$$

$$x_{i,j} \geq \sum_{i \in N} p_{i,j} r_i, \forall i \in N \tag{3.13}$$

$$c_{j,i,i'} \geq p_{j,i} + b_{j,i'} - 1, \forall i,j \in N, i' \in N - \{i\} \tag{3.14}$$

$$y_{i'} \geq \sum_{j \in N} c_{j,i,i'} r_j, \forall i \in N, i' \in N - \{i\} \tag{3.15}$$

$$B_i \cdot M \geq \sum_{j \in N} p_{i,j} + b_{i,j}, \forall i \in N \tag{3.16}$$

Constraints 3.10 and 3.11 ensure that there is one primary and one backup hotel for each RRU. Constraint 3.12 imposes primary and backup hotels to be disjoint. Constraint 3.13 counts the number of BBU ports to be installed in each primary hotel. Constraint 3.14 tells if a primary hotel is in common to a backup hotel for each source and is used in constraint 3.15 to ensure that there are enough BBU ports in each backup hotel. These two constraints, along with 3.9, allow minimizing the number of ports in each backup hotel. The number of BBU ports required at each backup hotel equals the largest number of RRUs that share the same primary hotel. Finally, constraint 3.16 activates hotels (i.e., tells if the hotel is a primary and/or backup for RRUs).
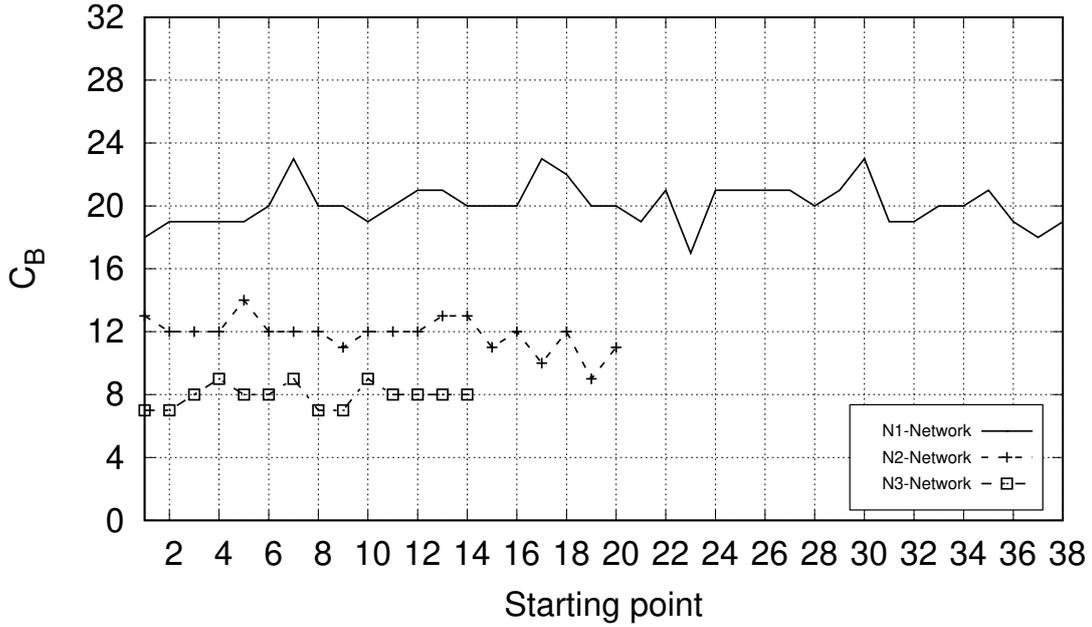
Figure 3.16: Effect of the starting point as a function of the number of BBU hotels (C_B) in $N1$, $N2$ and $N3$ networks.

3.4.3 Variable Distance Algorithm Results

The results related to the VDA technique and its comparison with the FDA one are reported in the following. Furthermore, the comparison between VDA and ILP presented over smaller size networks due to its scalability difficulties.

As stated before, the VDA technique does not relate to starting nodes as the FDA approach. On the contrary, it starts by choosing a node in the network and then checks all the placement alternatives. In the figure 3.16 all starting nodes are considered and the result of the placement in terms of BBU hotels is reported in $N1$, $N2$, and $N3$ networks. For the $N3$ network, a minimum of 7 BBU hotels and a maximum of 9 are found, which shows that the starting node introduces a maximum difference equal to 2 BBU hotels. In the case of $N2$ and $N3$ networks, this difference is larger which reflects the effect of topology on results. In the case of the $N2$ network, the difference between the maximum and the minimum is 5 BBU hotels and in the case of $N1$ network this difference reaches 6 BBU hotels but these values are due to singularities as a consequence of topology.

The benefit of the VDA approach is to find primary and backup hotels at the closest distance as possible. Since our emphasis is on the fronthaul segment based on CPRI, which puts strict requirements on delay, the methodology looks for primary BBU hotels as closest as possible to the RRUs and then proceeds for backup BBU hotels that could be further. Figure 3.17 shows the result of the application of this technique. It reports the comparison of average and maximum distance between each pair of RRU and primary and backup BBU hotels which indicate as *Ave-P*, *Ave-B*, *Max-P*, and *Max-B* respectively in the $N1$, $N2$, and $N3$ networks. The interesting fact from this figure is that in all networks the maximum distance between each RRU and primary BBU hotel pair is limited to one hop. This means that each RRU can find a primary BBU hotel either in the same node or at most one hop further. The backup BBU hotel, instead, can reach 5 hops distance with an average below 2
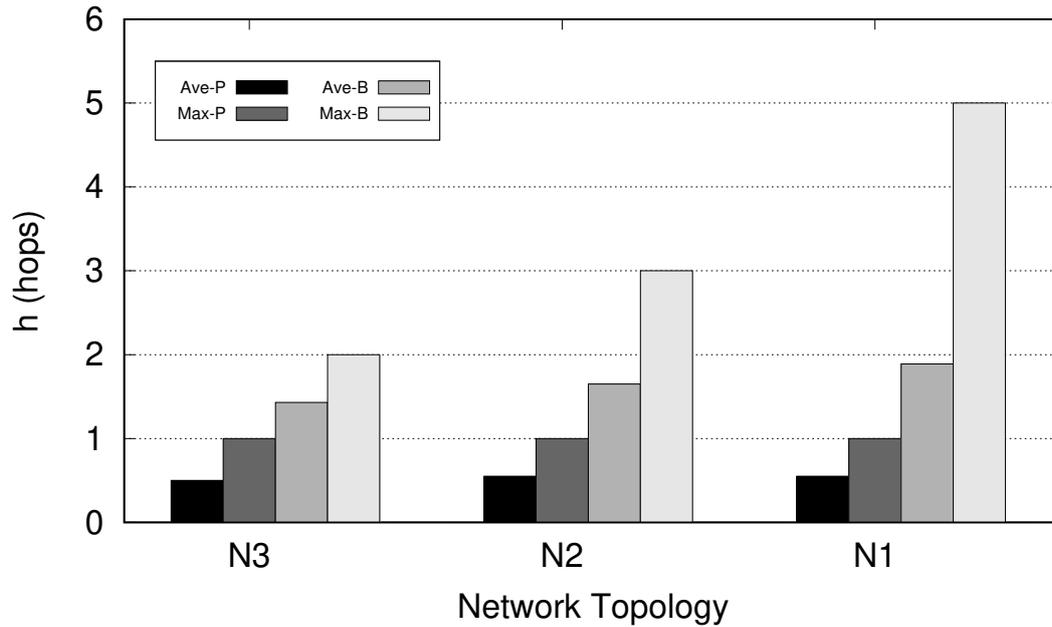
Figure 3.17: Comparison of average and maximum distance between an RRU and a primary and backup BBU hotels in $N1$, $N2$, and $N3$ networks.

hops to allow service continuity in case of failure.

To show the effectiveness of the VDA approach in finding optimal solutions, in figure 3.18 the total number of BBU hotels (C_B) as a function of topologies $N1$, $N2$ and $N3$ are shown. In this figure the Fixed Distance (FD) techniques, *Max-D* and *Min-D*, combined with textitP and B, are compared with Variable Distance (VD) results in the best and worst cases from figure 3.17. The best case is referred to as the lowest number of BBU hotels needed to have full coverage with protection. The worst case, on the contrary, is the case which requires the highest number of BBU hotels. These two values are extreme cases for each trend in figure 3.17. In all the network topologies the VD approach obtains better results than the FD one, even in the worst case. So it is concluded that VD not only finds the closest distance between RRU and BBU hotels, but it is also able to cover the whole network with resiliency by a less amount of BBU hotels.

In figure 3.19, using a similar methodology, the average number of ports per node (C_P) as a function of network topology for $N1$, $N2$ and $N3$ networks is reported for FD and VD techniques. The same conclusions can be drawn for the number of BBU hotels. By using VD a significantly lower number of ports, even in the worst case, is obtained compared to FD algorithms, thus supporting the effectiveness of the method. In relation to best and worst cases for the VD approach, a higher number of ports per node is observed for the best case, in $N1$ and $N2$ networks, which is related to the corresponding lower number of BBU hotels which gives less opportunity to share backup ports among BBU hotels. The relationship is different for the $N3$ network which could be explained as a consequence of specific topology.

Figure 3.20 shows the comparison of the average number of wavelengths per link (C_W) as a function of the three network topologies $N1$, $N2$, and $N3$. Since the VD technique can cover the whole network by a lower number of BBU hotels more wavelengths are needed to reach primary and backup nodes. As long as the network
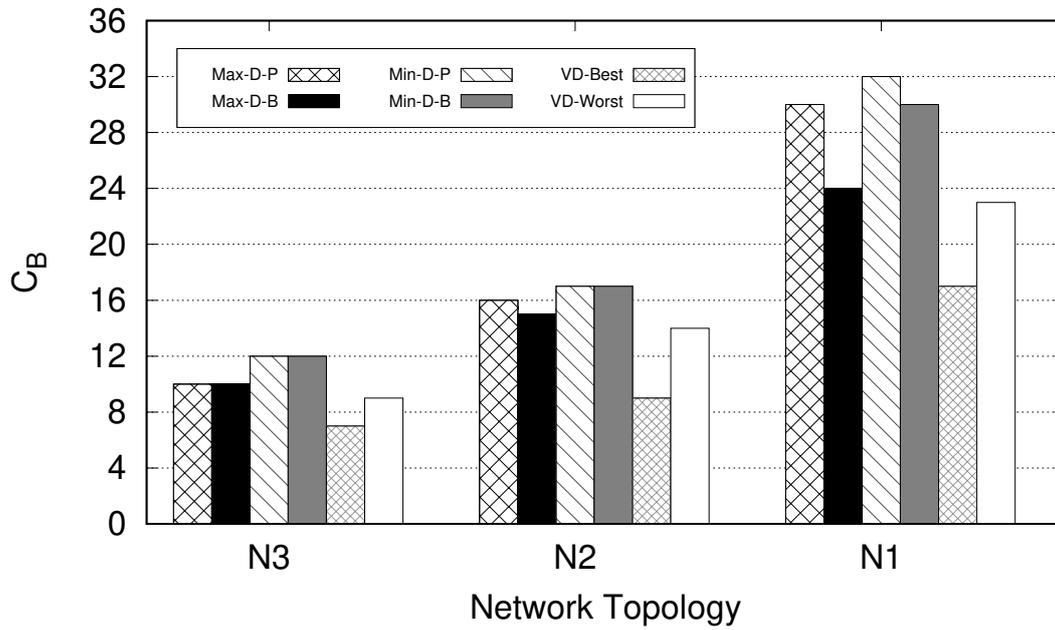
Figure 3.18: The total number of BBU hotels (C_B) for network topologies $N1$, $N2$, and $N3$, comparing different FD approaches with VD, by considering the worst and best cases from figure 3.17.
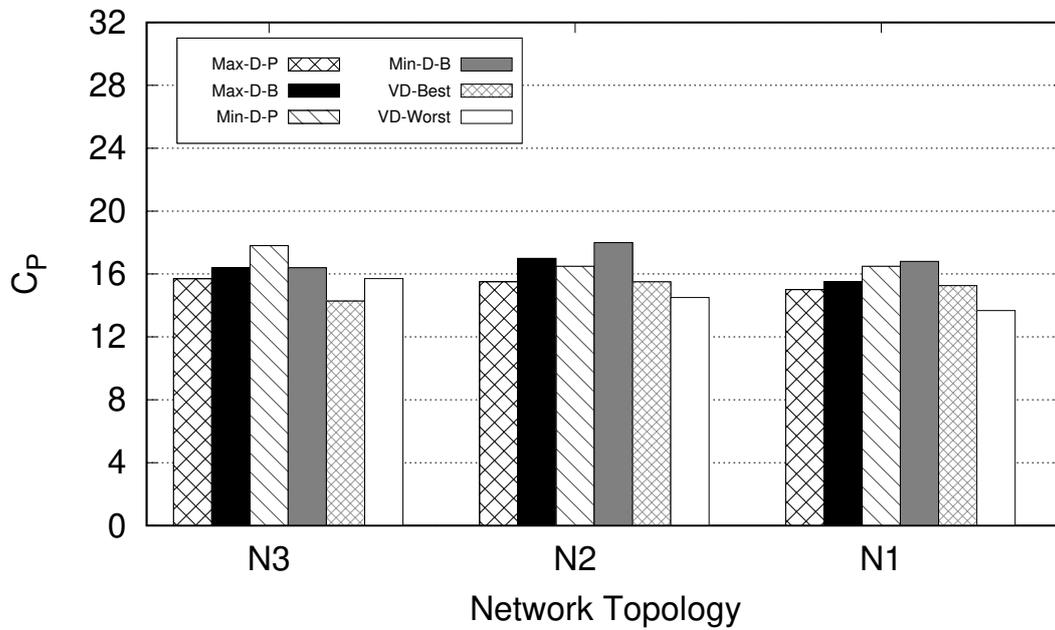
Figure 3.19: The average number of ports (C_P) for network topologies $N1$, $N2$, and $N3$, comparing different FD approaches with VD, by considering the worst and best cases from figure 3.17.
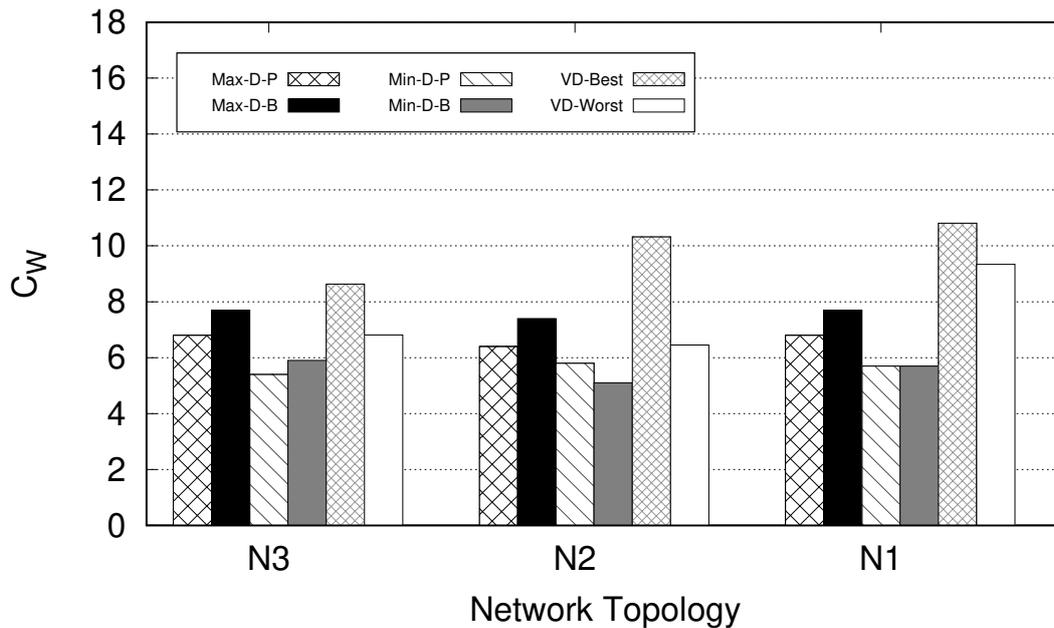
Figure 3.20: The average number of wavelengths (C_W) for network topologies $N1$, $N2$, and $N3$, comparing different FD approaches with VD, by considering the worst and best cases from figure 3.17.

size increases, the wavelength needed to connect each pair of RRU and primary and backup BBU hotel increases accordingly in the variable distance approach. In particular, the difference in the number of wavelengths for the worst and best BBU hotel assignments is particularly evident in limited size networks. This indicates that some trade-offs depending on costs are worth to be investigated.

VDA Results: Comparing ILP and Heuristic

In this part, an analysis of survivable fronthaul in C-RAN to evaluate the two strategies proposed and applied to different scenarios. The reference topologies of the optical transport network used in the performance assessment are presented in figure 3.21. Three metro/aggregation networks are considered with 16 nodes each but with different levels of connectivity. The connectivity N_i for network i is defined as follows:

$$N_i = \frac{\sum_{i=1}^{n} NO_i}{n} \tag{3.17}$$

where NO_i is the number of optical interfaces in node i and n is the total number of nodes, 16 for all networks in this evaluation.

In all the topologies each node represents a cell site, assumed to serve a value of the upstream traffic equal to 10 RRUs connected to the node, each one requiring two lightpaths, i.e., one connecting the RRU to the primary and one connecting the same RRU to the backup BBU hotel. Each edge in the graph represents a bidirectional fiber connection, all with the same length. The results discussed in this section are obtained using a Java-based simulator and compared with the optimal solution from ILP, obtained using CPLEX commercial tool. The results from the heuristic are averaged over all the possible combinations of BBU hotel pairs that can be used
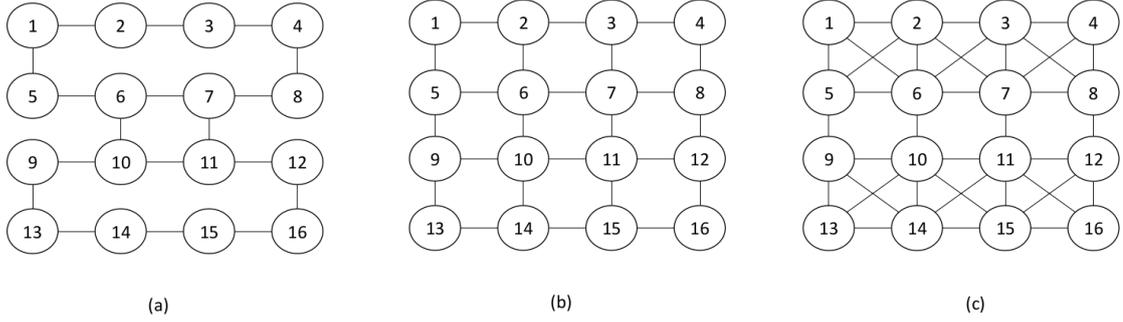
Figure 3.21: The reference network topologies, (a) network A with connectivity $N_A = 2.25$, (b) network B with connectivity $N_B = 3$ and (c) network C with connectivity $N_C = 4.5$.

as a starting point. Among the solutions, the maximum observed deviation from the average is 22% which shows the limited impact of the starting point on the results and allows the algorithm to start by random locations. In all the graphs reporting F and G, the results are normalized with respect to α (that was constant) and are reported in each case. All β_i were considered constant and equal to β. The following parameters are used:

$$R = \frac{\beta}{\alpha} \tag{3.18}$$

$$Q = \frac{\gamma}{\alpha} \tag{3.19}$$

Figure 3.22 reports the total cost of the survivable fronthaul design solution (i.e., the cost function F). In the figure, the two contributions to F are shown for each network when $R = 1$, and the total cost is normalized with respect to α. The cost obtained with the heuristic is compared to the one of the ILP when $\gamma = 0$ so that F has the same meaning as G. The total cost is lower for the ILP, with different contributions of BBU hotels and distance. While the ILP cost is constant with respect to different network connectivities, the cost of the heuristic is slightly higher when the network connectivity is higher. The reason is that the heuristic can activate less BBU hotels than the ILP, which causes the number of hops to grow, and results in an increased overall cost.

Similarly, figures 3.23 and 3.24 show the total cost function F, normalized respect to α when R equals 2 and 10, respectively. By increasing R, the hotel activation cost becomes more relevant in F, therefore the number of selected hotels decreases when R increases. For $R = 2$ the contribution of the BBU hotels to F is less than in the case $R = 1$. When $R = 10$, the number of active BBU hotels keeps decreasing but their contribution to the total cost becomes higher than in the case $R = 2$, due to the large R factor. As a final note, the heuristic provides a good approximation of the ILP when the activation cost and the distance have similar weight in F ($R = 1$) and when the activation cost is much more relevant than the distance ($R = 10$). In the case $R = 2$ instead, the heuristic solution is up to 40% more expensive than the ILP.

The number of BBU ports, that is the number of functional interfaces to serve the related RRUs, is calculated based on the number and location of BBU hotels. The
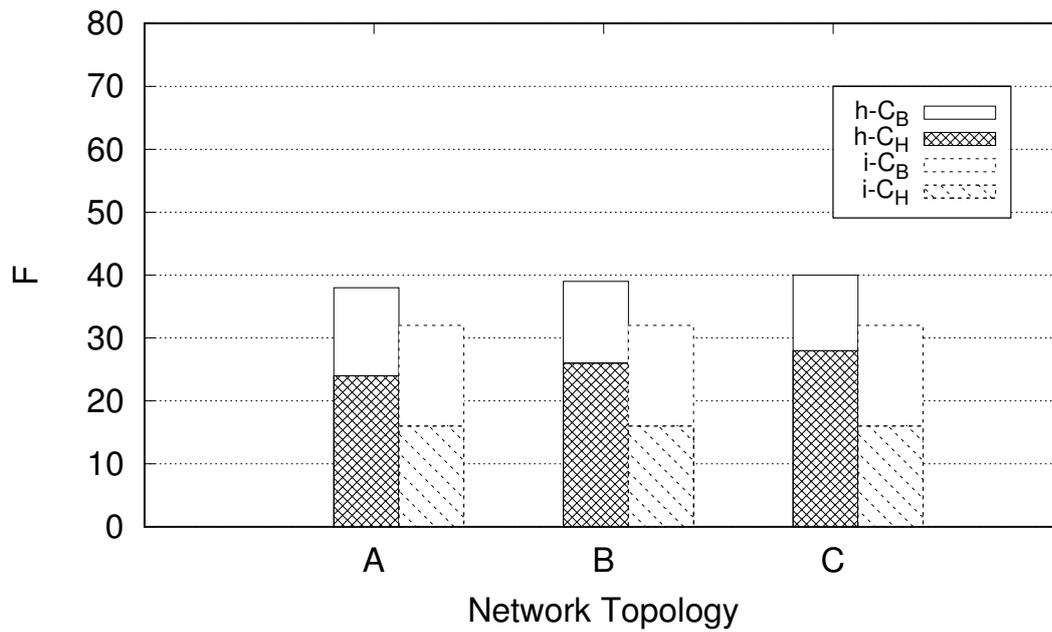
Figure 3.22: Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B, and C when $R = 1$.
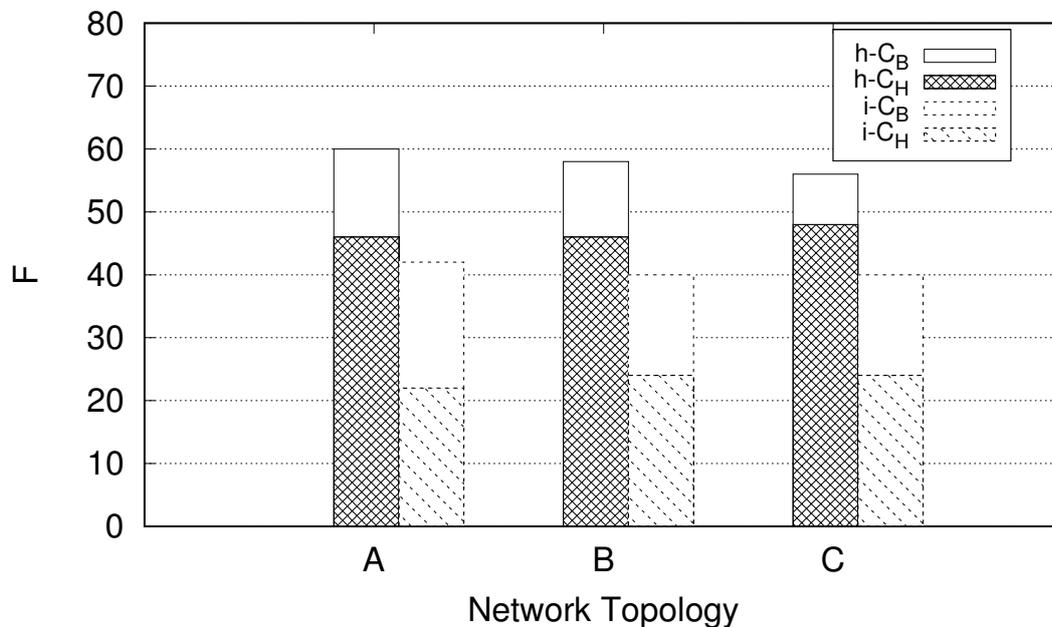
Figure 3.23: Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B, and C when $R = 2$.
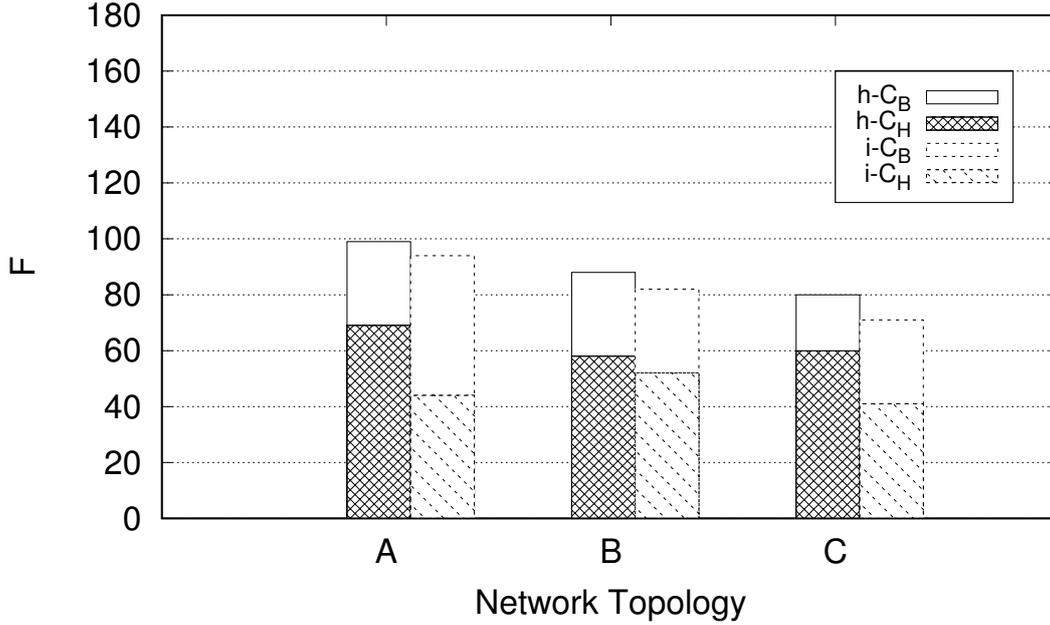
Figure 3.24: Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B, and C when $R = 10$.

previous results, obtained using F or G with $\gamma = 0$, do not include any consideration on the number of ports, not considered so far. To compare the results of the heuristic and ILP, the latter has been run once again to derive the minimum number of BBU hotel ports. α and β were all set to zero, γ was set to 1 and the hotel placement previously obtained was introduced in the ILP model as an additional constraint, to set the position of the BBU hotels. The overall number of backup ports obtained from the modified ILP is compared to the heuristic one, averaged over all the initial cases, and is reported in figures 3.25 and 3.26, for the three network topologies when R equal to 1 and 10, respectively. Since the total number of primary BBU hotel ports is fixed and equal to the number of RRUs, it is not included in these figures.

Figure 3.25 shows that the number of backup BBU hotel ports required by the ILP is lower than the heuristic one. In the case of $R = 1$, both ILP and heuristic have a large number of active BBU hotels, and since this number is higher for the ILP, ILP results more efficient in sharing BBU hotel ports. By increasing the network connectivity, the ILP easily assigns primary and backup BBU hotels such that the sharing of backup ports results higher than with the heuristic that, instead, assigns primary and backup hotels based only on F, and therefore is not aware of their impact on the number of shared backup ports.

Figure 3.26 shows that the sharing of BBU hotel ports is extremely difficult for the heuristic when R is high and the number of active hotels is very low. The total number of ports is high independently of the connectivity due to the fact that the solution obtained with the heuristic, averaged over all possible starting nodes, requires just two or three hotels to be active. The ILP instead, finds solutions with slightly more active hotels and therefore can limit the number of BBU hotel ports to lower values.
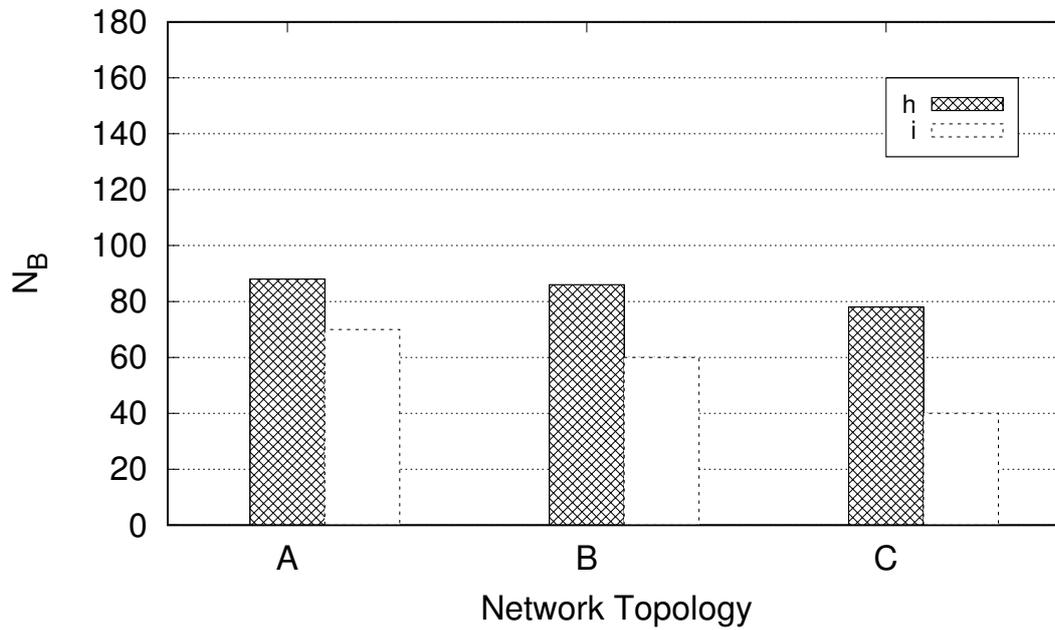
Figure 3.25: Total number of backup ports N_B for ILP (i) and heuristic (h) in networks A, B and C with $R = 1$.
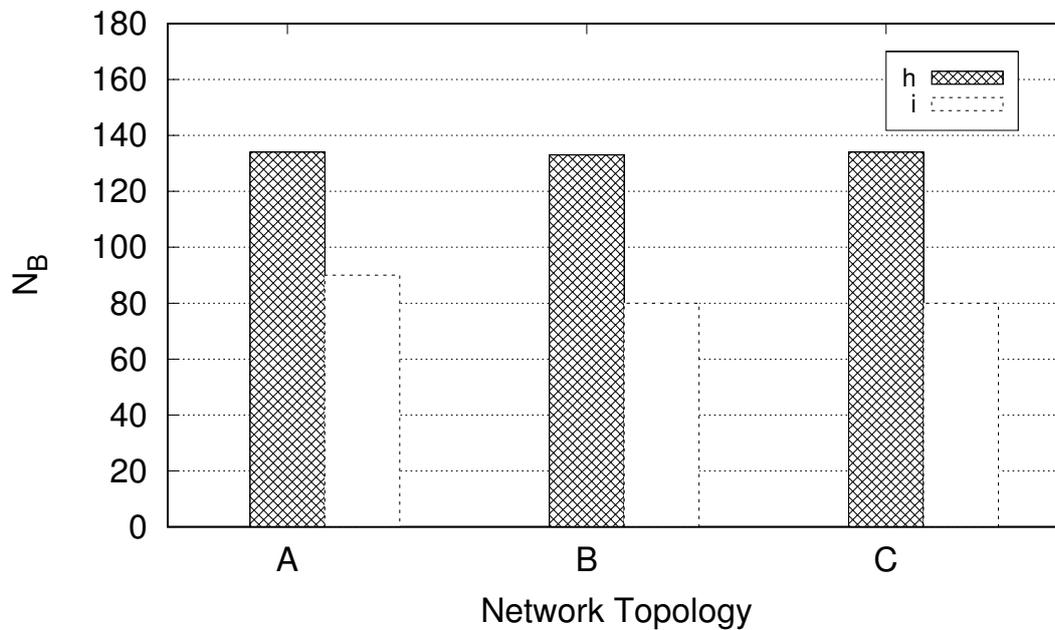
Figure 3.26: Total number of backup ports N_B for ILP (i) and heuristic (h) in networks A, B and C with $R = 10$.

In order to see the effects of γ on the placement, the value of the parameter Q is varied. Tables 3.2, 3.4, and 3.5 show the different values for F and G in the three networks when Q is equal to 0, 0.001 and 0.1, while R is considered constant and equal to 2. As expected, the total cost in each network increases by increasing Q, due to the cost introduced by the ports. For these values of Q, the sum of activation and distance costs are almost the same in the three cases, while their contribution changes. There may be solutions employing the different number of hotels and that leads to having slightly different costs like the case of $Q = 0.1$. The impact of γ on the cost is, therefore, to select the solution, among solutions with the same cost (measured by F), that minimizes also the total number of ports. The tables also show the average number of wavelengths per link without considering wavelength continuity. It is possible to notice how the required wavelengths per link decrease when the network connectivity increase, due to the higher number of available links to connect transport nodes.

In conclusion, when the contribution of the BBU hotel ports is considerably less relevant with respect to the activation and distance, which will represent a real case scenario, it is safe to neglect the contribution of the BBU hotel ports in the first computational phase. Then, when the hotels to activate are selected and the delay is minimized, a dedicated minimization can be performed to limit the number of BBU hotel ports.

Table 3.3: The effects of Q on the cost components of the objective function G for the network A ($R = 2$).

Network A							
Q	C_B	C_H	F	N_B	C_P	G	W
0	20	22	42	100	0	42	12.2
0.001	20	22	42	100	0.1	42.1	12.2
0.1	22	21	43	80	8	51	11.7

Table 3.4: The effects of Q on the cost components of the objective function G for the network B ($R = 2$).

Network B							
Q	C_B	C_H	F	N_B	C_P	G	W
0	16	24	40	80	0	40	10
0.001	16	24	40	80	0.08	40.08	10
0.1	18	23	41	70	7	48	9.6

Table 3.5: The effects of Q on the cost components of the objective function G for the network C ($R = 2$).

Network C							
Q	C_B	C_H	F	N_B	C_P	G	W
0	16	24	40	60	0	40	6.7
0.001	14	26	40	50	0.05	40.05	6.9
0.1	14	26	40	50	5	45	6.9

3.5 Sharing Backup Ports and Wavelengths

This part aims to share the backup ports and wavelengths to use network resources efficiently. As stated before, only a single BBU hotel or a single link failure considered at this time. For backup BBU hotel ports sharing the following rule should be applied: A backup BBU hotel port can be shared among some RRUs if and only if, those RRUs have primary BBU hotel ports in different BBU hotels. When a backup BBU hotel port shared between some RRUs, it will be reserved to be useful in case of failure. When a failure happens in a primary BBU hotel, the RRUs connecting to the failed BBU hotel, shifting to their backup BBU hotel. There should be enough BBU hotel ports in backup BBU hotel reserved to serve new RRUs. This is the reason backup BBU hotel ports can be only shared among RRUs from different primary BBU hotels so at each failure only one RRU uses the reserved backup BBU hotel port.

Algorithm 2 explains the procedure of BBU hotel ports sharing. It starts with the BBU hotel $i \in B$ in line 2 and identifies all nodes $j \in N$, contain a certain amount of RRUs, which have the backup ports in the BBU hotel i (line 3). In line 4, the algorithm searches for the primary BBU hotels $k \in B$ for each RRU in all nodes j because the sharing is only feasible if the RRUs have different primary BBU hotels. If these primary BBU hotels (k) are not the same, they will be recorded in a temporary set namely P_{Hotels} (line 6). Those RRUs in the nodes $j \in N$ which have the primary BBU hotels in the set P_{Hotels} can share their backup BBU hotel ports in BBU hotel i (line 10-12). The worst-case complexity of the sharing backup BBU hotel ports procedure is estimated as $\mathcal{O}(N^3)$.

Algorithm 2 Sharing Backup BBU Hotel Ports

1: **Begin:**
2: **for** each BBU hotel $i \in B$
3:     **for** each RRU in node $j \in N$ such that $BP_{ij} = 1$
4:         **if** exists a BBU hotel $k \in B$ such that $PP_{kj} = 1$
5:             **if** P_{Hotels} does not contain k
6:                 $P_{Hotels} = P_{Hotels} \cup k$
7:             **end if**
8:         **end if**
9:     **end for**
10:     **for** each RRU in node $j \in N$ such that $PP_{kj} = 1$ and P_{Hotels} contains k
11:         share backup port in BBU hotel i
12:     **end for**
13: **end for**

To have resilience against link failure, each RRU needs node and link disjoint lightpaths for its primary and backup BBU hotels. The sharing of backup wavelengths follows the same principle as the sharing of backup BBU hotel ports: a backup wavelength can be shared among some RRUs if and only if those RRUs have node and link disjoint lightpaths to reach their primary BBU hotels. Consequently, in case of a single failure, either BBU hotel or link, two or more RRUs can share the same backup wavelength in their backup lightpath, being their primary BBU hotels and primary lightpaths different. If a failure happens in the primary

BBU hotel, then the RRU can use the backup lightpath (and reserved backup wavelengths) to reach to the backup BBU hotel. Similarly, if a failure happens in any link of the primary lightpath, the RRU will use the backup lightpath toward its backup BBU hotel.

Algorithm 3 explains the procedure of sharing backup wavelengths. This algorithm starts by checking every link $i \in L$ in the network (line 2). For each link $i \in L$ it identifies all the RRUs in nodes $j \in N$ that use link i in the backup lightpaths (line 3). RRUs in node j that have different primary BBU hotels can share the backup wavelength in link i. In the specific, in line 4, the algorithm searches for the primary BBU hotels $k \in B$ for those RRUs sharing the same link $i \in L$ and check whether they are located in different nodes. In this case, the primary BBU hotels k will be added in the set P_{Hotels} (lines 5 and 6). The "for" loop starting from line 10 states that the RRUs in node j, which have a primary BBU hotel k in the set P_{Hotels}, can share the wavelength in link i. The set P_{Hotels} contains the different primary BBU hotels for the RRUs which share the backup link and eligible to share the same wavelength on that link (a condition in line 5). The worst-case complexity of the sharing backup wavelengths procedure is estimated as $\mathcal{O}(N^3)$.

Algorithm 3 Sharing Backup Wavelengths

1: **Begin:**
2: **for** each link $i \in L$
3:     **for** each RRU in node $j \in N$ such that $BL_{ij} = 1$
4:         **if** exists a BBU hotel $k \in B$ such that $PP_{kj} = 1$
5:             **if** P_{Hotels} does not contains k
6:                 $P_{Hotels} = P_{Hotels} \cup k$
7:             **end if**
8:         **end if**
9:     **end for**
10:     **for** each RRU in nodes $j \in N$ such that $PP_{kj} = 1$ and P_{Hotels} contains k
11:         share backup wavelength in link i
12:     **end for**
13: **end for**

3.6   Conclusion

The chapter presents a solution based on the Facility Location Problem (FLP) for BBU hotel placement in C-RAN to achieve protection in the fronthaul optical network segment against single BBU hotel failure. Different solutions have been proposed and compared in terms of relevant cost parameters, namely the number of BBU hotels, ports and wavelengths. Additional costs with respect to solutions without protection are evaluated showing the effectiveness of the proposed algorithms to maintain additional costs low. The proposed extension of a classical Facility Location Problem applied to support C-RAN resiliency, the VDA approach has been shown to achieve the lowest costs for both the BBU hotels and the required number of ports, which are the optical interfaces. In any case, the required amount of wavelengths is against the trend of the number of BBU hotels to support resiliency and trade-off depending on real deployment cost has to be found. Even though the VDA

algorithm does not put constraint on maximum distance, the primary node location is shown to be less than or equal to 1 hop distance for any topology, while, in case of failure, the backup node is any way at a limited distance from the served RRU, due to the minimization of the distance cost performed by the VDA algorithm. In the algorithm theoretical complexity, although both VD and FD approaches have the same complexity, depends on the network design requirements either technology can be applied.

Chapter 4

Design Methodologies for Reliable C-RAN - Distributed Approach

4.1   Introduction

The optimal placement of baseband functions in BBU hotels is challenging, especially in dynamic scenarios where these functions require to be activated in relation to access network topology changes. Also, service continuity in case of failures must be guaranteed. In Centralized Radio Access Network (C-RAN), the baseband functions placement can be performed by the Software-Defined Network (SDN) control/management plane based on a complete knowledge of the network. However, in access or aggregation networks that evolve, the knowledge of the network state may require frequent interactions between controller and network entities, thus making this approach unpractical. Besides, centralized approaches may result not scalable enough to meet computational requirements in networks with the high number of nodes, as it is expected to happen in 5G access networks. Distributed algorithms executed by network nodes can be adopted instead, which will be shown also to provide incremental solutions when adding or removing virtual or physical baseband resources.

Distributed Facility Location (DFL) problem has been proposed for the flexible configuration of wireless sensor networks [48] with no explicit solutions for survivability. Machine learning (ML) approaches are recently emerging as a viable solution to cope with dynamic contexts such as those represented by C-RAN. Applications of ML algorithms to self-organizing cellular networks have been recently described in [49], where 5G C-RAN has been also addressed as a potential future research direction. However, to the best of our knowledge, no application of ML to the DFL problem in C-RAN has been developed yet. The ML approach is expected to be effective also in C-RAN dynamic reconfiguration needs. ML provides a framework to define algorithms that proceed in learning some properties of the system to obtain some performance target [50]. In the meantime, the elements of the system are enriched with information that turns to be useful in the evolution of the system. These characteristics make the approach suitable for the BBU hotel assignment in C-RAN.

In this chapter, the concept of resiliency for C-RAN will be investigated with the Machine Learning (ML) approach and the benefits over the centralized one will be shown.
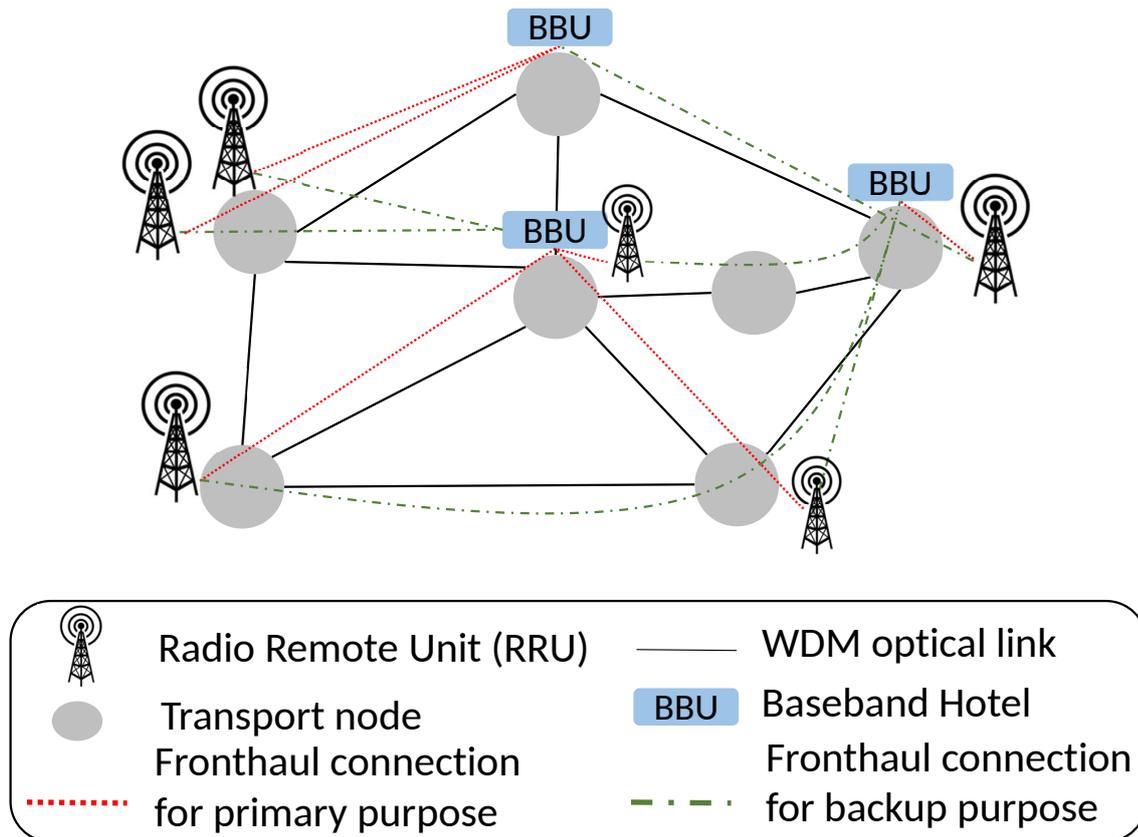
Figure 4.1: C-RAN architecture.

4.2 Distributed Method - ML approach

The reference C-RAN architecture used in the chapter is shown in figure 4.1. A set of RRUs in an area is divided into groups and connected to different nodes of the transport network, called transport nodes. Transport nodes are interconnected using optical fibers, creating the so-called fronthaul network. Each transport node hosts an edge Data Center (DC) with computational resources that can be used to execute the distributed algorithm locally and to virtualize baseband resources, whenever needed. Transport nodes are all potential candidates to host BBU hotels, where several BBUs, either physical or virtualized and accessed through ports, can be deployed. All RRUs directly connected to the same transport node is assumed to be assigned to the same BBU hotel to apply interference mitigation in the considered area [51].

Each RRU has a dedicated BBU port in the primary BBU hotel and an additional port on a backup BBU hotel assigned for reliability, possibly shared with other RRUs as it will be explained later. The fronthaul network imposes strict latency and capacity demands [52], and usually requires dedicated resources to satisfy these requirements. The traditional packet-based backhaul provides connectivity between BBU hotels and the core network of mobile network providers (not reported in the figure).

The survivable BBU hotel placement problem addressed in this chapter is defined as follows:

- **Given:** a set of transport nodes, each of which contains the information re-

Table 4.1: List of cost variables and parameters.

Parameters:	
S	Set of transport nodes, $|S| = s$.
C_B	The cost of activation of BBU hotel.
C_H	The cost of distance between each BBU-RRU pair.
C_P	The total cost backup and primary BBU ports.
H	$s \times s$ matrix. h_{ij} is the distance in hops between nodes i and j computed with the shortest path.
α	Weight for the distance in the cost function.
β	Activation cost for a single BBU hotel.
γ	Cost of a BBU hotel port.
Variables:	
B_i	1 if node $i \in S$ hosts a BBU hotel, 0 otherwise.
p_{ij}	1 if BBU hotel i is assigned as primary for RRUs at node j,0 otherwise.
b_{ij}	1 if BBU hotel i is assigned as backup for RRUs at node j, 0 otherwise.
x_i	Number of BBU ports required at hotel site i for primary purposes.
y_i	Number of BBU ports required at hotel site i for backup purposes.

garding i) total number of connected RRUs, ii) directly connected neighbor transport nodes, iii) maximum number of wavelengths in each link, and iv) maximum allowed distance to connect RRUs with BBUs.

- **Find:** a minimum cost primary BBU hotel placement so that each RRU is assigned to a BBU according to limitations on distance and wavelength availability for fronthaul links, and a minimum cost backup BBU hotel placement for reliability against single BBU hotel failure.

In the following, some useful parameters and variables are defined, while the notation used throughout this section is summarized in table 4.1.

To provide reliability against single BBU hotel failure, each RRU is connected simultaneously to two BBU hotels placed in different transport nodes, one for primary and one for backup purposes. The activation cost of BBU hotels in transport nodes needed to provide full coverage and resiliency of the target area is calculated using the following formula:

$$C_B = \beta \cdot \sum_{i \in S} B_i \tag{4.1}$$

where B_i is a boolean variable equal to 1 when the transport node hosts a BBU hotel, that is when it requires BBU functionalities, either for primary or backup purposes, related to one or more RRUs. β is a parameter associated with the activation cost for a BBU hotel in transport nodes and can be set accordingly.

To account for the delay introduced in the fronthaul network, a cost can be associated with the distance between BBU hotels and RRUs connecting to them. Distance between adjacent transport nodes is here assumed to be equal to 1 hop for all links. The overall cost for the distance is expressed as follows:

$$C_H = \alpha \cdot \sum_{i \in S} \sum_{j \in S} p_{ij} h_{ij} + \alpha \cdot \sum_{i \in S} \sum_{j \in S} b_{ij} h_{ij} \qquad (4.2)$$

where p_{ij} and b_{ij} are boolean variables that indicate if BBU hotel i is assigned as primary or a backup, respectively, for the group of RRUs at transport node j. h_{ij} represents the distance, in hops, between transport node i and j computed solving the shortest path problem. Both contributions (i.e., the overall distance for the primary path and backup path) are multiplied by cost parameter α, which represents the cost for the 1 hop link.

Finally, the proper number of BBU ports must be allocated in each hotel. The total number of primary and backup BBU ports, and the related cost, are calculated according to the following formula:

$$C_P = \gamma \cdot \sum_{i \in S} x_i + y_i \qquad (4.3)$$

C_P is the contribution of the total number of primary x_i and backup y_i ports in each hotel multiplied by the cost parameter γ associated with each port. Since the protection requires that each RRU is connected to two different BBU hotels, the total number of ports should be twice the number of RRUs, and consequently, the value for C_P can be fixed. However, only the number of primary ports is fixed and equal to the number of RRUs. On the contrary, the number of backup ports can be reduced. RRUs can share backup ports if they have different primary BBU hotels. When a single hotel failure occurs, RRUs assigned to that primary hotel switch to their backup hotel, hence it is forbidden to share backup ports among RRUs assigned to the same primary. By sharing the backup ports, the value for C_P can be reduced, and further cost saving can be achieved.

4.3   Design methodologies

In this section, two network design strategies to solve the survivable BBU hotel placement problem are presented in detail. First, a conventional centralized network deployment strategy based on ILP is presented. This strategy is intended to be executed on the top of the SDN controller since it requires complete knowledge of network topology and resources. Every time that there are changes to the network, the strategy must be re-executed to compute the new optimal state, and the controller is in charge of activating the required resources and configure new paths.

In the second subsection instead, a distributed algorithm based on a heuristic approach is proposed. As opposed to the centralized strategy, this algorithm is meant to be executed independently by every single node. When a new node is connected to the network, this strategy is executed by the new node to find suitable primary and backup BBU hotel autonomously. Even though the network controller is not needed in this phase, its presence is required, for example, to monitor the network state, to allocate network resources (i.e., new BBUs) after failures and to install new paths. Table 4.2 reports additional parameters and variable used in this section.

Table 4.2: List of variables and parameters for algorithm definition.

	Parameters:
r_i	Number of RRUs at site $i \in S$.
δ_{ij}^l	1 if shortest path between i and j is using link l, 0 otherwise.
M_W	Maximum number of wavelengths available in each link.
M_H	Maximum allowed distance between RRU and BBU (in hops).
L	Set of links.
$N_i(M_H, M_W)$	Set of eligible nodes within M_H and M_W constraints from node $i \in S$.
w_l	Number of wavelengths in use in the link $l \in L$.
N_i	Set of directly connected nodes to node $i \in S$.
TTL	Time-To-Live in hops.
CS_n	Current Set, used in Algo. 4 containing nodes to be considered at iteration n.
SA_j	Array of cell sites sharing the same primary BBU hotel $j \in S$.
Max	Parameter storing the largest number of cell sites sharing the same primary.
M	A large number.
	Variable:
$c_{iji'}$	1 if RRUs at node j are using destination i as primary and i' as backup hotel site; 0 otherwise.

4.3.1 Centralized ILP

Objective function:

$$Minimize\ G = C_B + C_H + C_P \tag{4.4}$$

The multi-objective function 4.4 is composed of three members. The first term takes into account the activation cost of each hotel (C_B). The second term accounts for the cost to connect RRUs to BBU hotels, both primary and backup (C_H) while the third term accounts for the cost of BBU ports required in each hotel (C_P). The problem is subject to the following constraints:

$$\sum_{i \in S} p_{ij} = 1, \forall j \in S \tag{4.5}$$

$$\sum_{i \in S} b_{ij} = 1, \forall j \in S \tag{4.6}$$

$$p_{ij} + b_{ij} \leq 1, \forall i, j \in S \tag{4.7}$$

$$B_i \cdot M \geq \sum_{j \in S} p_{ij} + b_{ij}, \forall i \in S \tag{4.8}$$

$$(p_{ij} + b_{ij}) \cdot h_{ij} \leq M_H, \forall i, j \in S \tag{4.9}$$

$$\sum_{a \in S} \sum_{b \in S} (p_{ab} + b_{ab}) \cdot \delta_{ab}^l \cdot \delta_{ij}^l \cdot r_b \leq$$
$$M_W + M \cdot (1 - p_{ij} + b_{ij}), \forall l \in L, i, j \in S \tag{4.10}$$

$$x_i \geq \sum_{j \in S} p_{ij} \cdot r_j, \forall i \in S \tag{4.11}$$

$$c_{iji'} \geq p_{ij} + b_{i'j} - 1, \forall i, j \in S, i' \in S - \{i\} \tag{4.12}$$

$$y_{i'} \geq \sum_{j \in S} c_{iji'} \cdot r_j, \forall i \in S, i' \in S - \{i\} \tag{4.13}$$

Constraints 4.5 and 4.6 ensure that there is one primary and one backup hotel for each RRU. Constraint 4.7 imposes primary and backup hotels to be disjoint. Constraint 4.8 activates hotels (i.e., tells if the hotel is a primary and/or backup for RRUs), while constraint 4.9 ensures that the maximum allowed distance M_H (in hops) is not exceeded. Constraint 4.10 limits the number of wavelengths over each link to M_W. Constraint 4.11 counts the number of BBU ports to be installed in each primary hotel. Constraint 4.12 tells if a primary hotel is in common to a backup hotel for each source and is used in constraint 4.13 to ensure that there are enough BBU ports in each backup hotel. These two constraints, along with 4.4, allow minimizing the number of ports in each backup hotel. In fact, the number of BBU ports required at each backup hotel equals the largest number of RRUs that share the same primary hotel.

4.3.2 Distributed heuristic

The proposed strategy to solve the survivable BBU hotel placement problem is performed in two phases. In the first phase, the algorithm decides where to activate primary and backup hotels. In the second phase, BBU ports are shared, whenever possible, to further minimize the total cost.

The distributed procedure proposed for BBU hotel placement in C-RAN is presented as algorithm 4, while the notation used throughout the chapter is summarized in table 4.2. The following assumptions are made.

In the beginning, each transport node is assumed to have information only regarding the number of directly connected RRUs and transport nodes, and the availability of wavelengths in each directly connected links. In order to provide coverage and resiliency for all RRUs in the network, the exchange of information among nodes is required. More specifically, the nodes interact to learn information regarding i) wavelengths availability and ii) if nodes are already active, i.e., if they are hosting active hotels. After the procedure is performed, the application running in the node asks the network controller to establish the connections with the selected nodes (two, one for primary and one for backup purposes) and to activate/reserve baseband resources in their local DC.

The neighbor nodes set of a transport node i is defined as the set of nodes to which an RRU, attached to i, can be connected, i.e., transport nodes with baseband resources within distance M_H and with sufficient wavelengths along the path. Algorithm 4 is executed in each transport node upon the needs of connecting a new

RRU to two BBU functionalities in separate BBU hotels, for primary and backup purposes. Algorithm 4 calls algorithm 5 to find the neighbors of a node. Since the probability of activating two or more RRUs at the same time is rare, only one transport node at a time is assumed to execute algorithm 4. This is due to the fact that distribution of the traffic in the network is not uniform and also the network can adopt and periodically optimize by different traffic pattern.

The procedure presented by algorithm 4 is executed in each transport node as long as there is a new request for primary or backup BBU hotel connection. The starting node, namely node i, is chosen randomly. The algorithm starts at line 2. Function NFF is called from node i in line 3 in order to extract set of neighbor nodes $N_i(M_H, M_W)$. This set contains nodes within the maximum number of hops (M_H) from node i that have enough wavelengths to allow the connection of new RRUs. If node i hosts an active BBU hotel (line 5) it is selected to act as a primary BBU hotel for RRUs at node i (line 6). For the backup connection, if exists an active BBU hotel at node j in the set of neighbors for i $N_i(M_H, M_W)$ (line 7), RRUs at node i connects to it for backup purpose (line 8). Lines 9 to 11 reserve the required wavelengths in all the links of the path from i to j, in order to accommodate the traffic from new RRUs. If no node in the set $N_i(M_H, M_W)$ has an active BBU hotel (line 12), then one node will be chosen randomly from the set (line 13) to act as a BBU hotel (line 14) in order to be backup BBU hotel for node i (line 15). The wavelengths in all the links between nodes i and j will be updated accordingly (lines 16-18).

If node i does not have any BBU hotel in its cell site (line 21), three possible situations might happen: the first is the case when two active BBU hotels exist in the set $N_i(M_H, M_W)$ namely nodes j and z (line 23). If there are more active BBU hotels in the set, two are chosen randomly. In this case node i will connect to them one as primary (line 24) and the other as backup BBU hotel (line 25). Consequently, all the wavelengths to be used in the links forming the path to primary and backup BBU hotels are be updated accordingly (lines 26- 28).

The second possible situation happens when only one node, namely node j, in the set $N_i(M_H, M_W)$ has an active BBU hotel (line 30). In order to keep the primary BBU hotel as close as possible, node i activates a BBU hotel in its cell location (line 31) and connects its RRUs to it as primary BBU hotel (line 32). The backup BBU hotel is node j (line 33). In lines 34 to 36 all the wavelengths in the links between nodes i and j are updated accordingly. The last case happens when no active hotel is found in the set $N_i(M_H, M_W)$ (line 38), then with the same line of reasoning of keeping primary BBU hotel as close as possible, in line 39 node i activates its BBU hotel as primary BBU hotel (line 40). One random node, namely node j, from the set $N_i(M_H, M_W)$ is chosen (line 41) and assigned for backup purposes (lines 42 and 43). Like in the other cases, all the wavelengths in the links between nodes i and j are updated accordingly (lines 44 - 46).

In order to keep track of the maximum number of hops, algorithm 5 sets Time-To-Live (TTL) parameter equal to M_H (line 2) and stops when TTL reaches to zero (line 4). The current set CS_n of nodes to be considered at first iteration ($n = 0$) is initialized with initial node i. While the CS_n is not empty (line 5), a random node (k) contained in this set is considered (line 6). All the neighbor nodes of k are iteratively considered (line 7) and placed in the neighbor set of the initial node i (line 9) if there are enough wavelengths to accommodate the request (line 8). Neighbor

---

Algorithm 4 Distributed Location Algorithm

1: **Begin**
2: $i =$ a random node in set S
3: **call** $NFF(i)$
4: //Procedure when node has active hotel:
5:     **if** $B_i = 1$
6:        $p_{ii} = 1$
7:        **if** exists a node $j \in N_i(M_H, M_W)$ s.t. $B_j = 1$
8:           $b_{ji} = 1$
9:           **for** all $l \in L$ between nodes $i, j \in S$
10:             $w_l = w_l + r_i$
11:           **end for**
12:        **else**
13:           $j =$ a random node in set $N_i(M_H, M_W)$
14:           $B_j = 1$
15:           $b_{ji} = 1$
16:           **for** all $l \in L$ between nodes $i, j \in S$
17:             $w_l = w_l + r_i$
18:           **end for**
19:        **end if**
20: //Procedure when node does not have active hotel:
21:     **else**
22: //There are two hotels in the neighbors set:
23:        **if** exist $j, z \in N_i(M_H, M_W)$ s.t. $B_j = B_z = 1$
24:           $p_{ji} = 1$
25:           $b_{zi} = 1$
26:           **for** all $l \in L$ between nodes i and $j, z \in S$
27:             $w_l = w_l + r_i$
28:           **end for**
29: //There is one hotel in the neighbors set:
30:        **else if** exists $j \in N_i(M_H, M_W)$ s.t. $B_j = 1$
31:           $B_i = 1$
32:           $p_{ii} = 1$
33:           $b_{ji} = 1$
34:           **for** all $l \in L$ between nodes $i, j \in S$
35:             $w_l = w_l + r_i$
36:           **end for**
37: //No hotel exists in the neighbors set:
38:        **else**
39:           $B_i = 1$
40:           $p_{ii} = 1$
41:           $j =$ a random node in set $N_i(M_H, M_W)$
42:           $B_j = 1$
43:           $b_{ji} = 1$
44:           **for** all $l \in L$ between nodes $i, j \in S$
45:             $w_l = w_l + r_i$
46:           **end for**
47:        **end if**
48:     **end if**
49: **Stop**

---

node j is then inserted in the set of nodes to be considered in the next iteration (line 10). After all the neighbors of k have been identified, k is removed from the current CS_n (line 13), and these instructions are repeated until CS_n is empty. At this point, TTL is updated (line 15) and the iteration index n is updated (line 16). This procedure is repeated until the limit set by TTL is reached, then the set of neighbors of i is returned (line 18).

Algorithm 5 Neighbor Finder Function (NFF)

1: **Given:** node $i \in S$
2: **Initialization:** $TTL = M_H$, $CS_0 \leftarrow i$, $n = 0$
3: **Begin**
4: **while** $TTL \neq 0$
5:    **while** $CS_n \neq \{\}$
6:       **get** random node k from CS_n
7:       **for** all nodes $j \in N_k$
8:          **if** $w_l + r_i \leq M_W$
9:             $N_i(M_H, M_W) \leftarrow j$
10:            $CS_{n+1} \leftarrow j$
11:          **end if**
12:       **end for**
13:       **remove** k from CS_n
14:    **end while**
15:    $TTL = TTL - 1$
16:    $n = n + 1$
17: **end while**
18: **Return** $N_i(M_H, M_W)$
19: **Stop**

BBU Port Sharing

After finding the BBU hotel placement, RRUs are re-assigned to further reduce the number of ports by sharing backup BBU ports. For this phase, nodes have to interact to exchange information regarding the primary hotel for RRUs that share the same backup BBU hotel. The rule to perform port sharing is that RRUs assigned to different primary hotels can share the same backup BBU port. Therefore, the minimum number of backup BBU ports equals the maximum number of RRUs that share the same primary hotel. These ports are sufficient to guarantee backup service to all the RRUs connected to the backup hotel and can be used when a single hotel failure occurs. This procedure is reported in algorithm 6.

Algorithm 6 is executed in every node which has an active BBU hotel. In this pseudocode, the considered active hotel is node i. Algorithm 6 starts at line 2 by introducing SA_j as a set of RRU sites sharing the same primary hotel j, initially set to zero for each node j in the network. Also, a parameter Max is initially set to zero, introduced to stores the maximum value of RRUs sharing the same primary. In line 3 BBU hotel located at node i identifies all nodes $j \in S$ that are using BBU hotel at node i as backup. Lines 4 to 7 aims at finding all the other nodes, like node $k \in S$, such that both node j and k have the same primary BBU hotel (line

4). If such hotel exists, namely BBU hotel located at node $z \in S$ (line 5), then the value for SA_j increases by one (line 6). Since all RRUs at nodes j and k have their primary BBU ports in the same BBU hotel, they must have distinct backup BBU ports in BBU hotel i, so if BBU hotel z fails, there are enough ports at hotel i to accommodate the new RRUs. After checking all the nodes that share the same primary BBU hotel with node j, at line 9 the number of antennas sharing the same primary is compared with Max to store the maximum value (line 10). In line 13 the minimum number of backup BBU port (y_i) that node i must have to guarantee protection for all RRUs connected to it is set to be equal to the maximum value (Max).

Algorithm 6 Sharing Backup BBU Hotel Ports

1: **Begin:**
2: **Initialization:** $SA_j = 0, \forall j \in S$, $Max = 0$
3: **for** all nodes $j \in S : b_{ij} = 1$
4:     **for** all nodes $k \in S : b_{ik} = 1$
5:         **for** all nodes $z \in S : B_z = 1$ and $p_{zj} = p_{zk} = 1$
6:             $SA_j = SA_j + 1$
7:         **end for**
8:     **end for**
9:     **if** $SA_j \cdot r_j > Max$
10:         $Max = SA_j \cdot r_j$
11:     **end if**
12: **end for**
13: $y_i = Max$
14: **Stop**

4.4   Numerical results

The reference topologies of the optical transport network used in the performance assessment are presented in figure 4.2. The ILP and heuristic are evaluated firstly in a 16 nodes network (figure 4.2). Then, a 17 nodes network is considered, where a node is added to the 16 nodes network to evaluate the capabilities of the two strategies to deal with dynamic scenarios. Finally, to evaluate a larger scenario, also a 36 nodes network is considered (figure 4.3).

The results discussed in this section are obtained using a Java-based simulator and compared with the optimal solution from ILP, obtained using CPLEX commercial tool. The ILP results are obtained in the case $\beta >> \alpha >> \gamma$, so to prioritize the minimization of BBU hotel activation, then the distance and finally the number of BBU ports. Given the intrinsic randomicity of the distributed strategy, 50 different simulations are performed.

Figure 4.4 reports the number of BBU hotels required by ILP and heuristic in the best and worst case, i.e., the cases when the active BBU hotels are minimum and maximum, respectively. The ILP provides always the best solution and the number of required BBU hotels decreases when the distance constraint increases. This is because, when the allowed distance increases, RRUs can be connected to hotels in farther nodes, requiring fewer BBU hotels to be activated. In the case
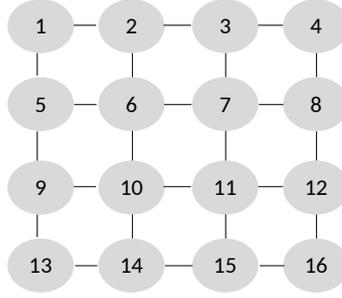
Figure 4.2: Network with 16 nodes.
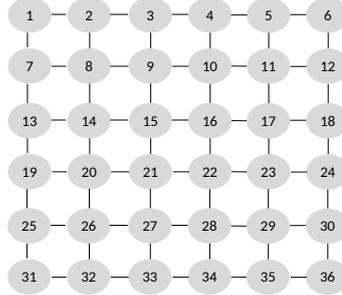
Figure 4.3: Network with 36 nodes.

of heuristic instead, different behaviors are experienced. In particular, in the best case for the heuristic, i.e., when the choices due to randomness are favorable to reduce the number of active hotels, the proposed strategy follows the trend of the ILP, requiring only one additional hotel, in case of 1 hop constraint, with respect to the optimal solution. In the worst-case instead, the number of active BBU hotels is larger and follows the trend of the ILP only until 3 hops constraint. For larger distance constraints (4 to 6), the amount increases due to the limit on the number of wavelengths. The distributed strategy tries to connect RRUs to the farthest BBU hotel that can reach, increasing the wavelength need over the links. When the distance constraint is large, some of the links are saturated and therefore closer BBU hotels must be selected, increasing the number of active hotels and decreasing the BBU hotel sharing.

Figure 4.5 depicts the number of BBU hotels required by ILP and heuristic, averaged over all the 50 cases, with and without the hop constraint (*h-80-avg* and *h-inf-avg*, respectively). The case without wavelength limitation follows the decreasing trend of the ILP, reaching optimal solutions when the maximum allowed distance is 5 and 6 hops. In the wavelength limited case instead, in these last two cases the number of active BBU hotels increases, following the trend of the maximum case reported in the previous figure.

The wavelengths usage, in the most used link and on average (over links), for ILP and heuristic, with and without hop constraint, in the 16 nodes network is reported in table 4.3. When the wavelength limit (set to 80) applies, both the ILP and the heuristic require all the wavelengths in the most used link, when the maximum allowed distance is higher than 3 hops. In this case, two hotels are enough to ensure protected service for all RRUs (see figure 4.4), so the links directly attached to the selected hotels become fully used. In the case of ILP, the average wavelength usage increases until it reaches a value of 26.7, which is the minimum cost case.

Figure 4.4: The number of active BBU hotels required by ILP and heuristic in the best and worst case for different distance constraints in the 16 nodes network, with wavelength constraint equal to 80.
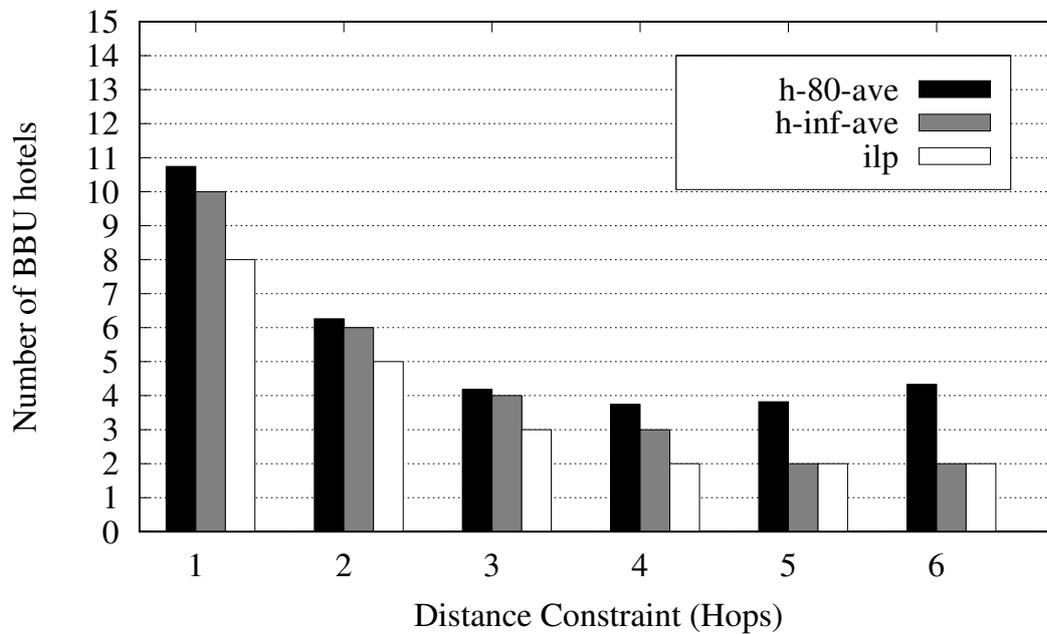
Figure 4.5: The number of active BBU hotels required by ILP and heuristic, with and without wavelength constraint, averaged over 50 cases for different distance constraints in the 16 nodes network.

Table 4.3: The number of wavelengths per link (maximum and average cases) required by ILP and heuristic, with and without wavelengths constraint, for different limits over distance in the 16 nodes network.

| Maximum distance in [hops] | Number of wavelengths per link | | | | | |
| | ILP | | H-80 | | H-inf | |
	Max	Avg	Max	Avg	Max	Avg
1	20	10	20	8.9	20	8.9
2	40	15.8	50	16.9	50	17
3	50	21.7	70	24	80	23.8
4	80	26.7	80	30.1	90	28.8
5	80	26.7	80	33.8	100	30
6	80	26.7	80	34	130	33.9

The cost for the heuristic, instead, keeps increasing, even though the active BBU hotels increase when the distance constraint is equal to 5 and 6. This is because each node runs the algorithm only once and when the links reaching active BBU hotels are full, BBU hotels in different nodes are selected. However the primary and backup hotels, and the wavelengths already assigned, cannot be changed, even if the new hotels are closer to nodes already assigned, thus increasing the overall link resources usage.

The table 4.3 reports also the case with is no wavelength limitation in the heuristic. In this case both the maximum number of wavelengths allocated in the most used link and the average usage increases over 80, allowing this strategy to reach near-optimal solutions in terms of active BBU hotels.

Table 4.4 reports the average distance, in hops, between RRUs and BBUs for ILP and heuristic with different limits over distance in 16 nodes network and maximum wavelengths limit equal to 80. The hops, both maximum and average, required by ILP increases with the maximum allowed distance and reaches the maximum value of 4 and an average of 2. In the heuristic instead, the maximum distance increases up to 6, while the average reaches a maximum of 2.5. The increasing trend in both cases is because, when the maximum allowed distance increases, both strategies try to reach farther BBU hotels to reduce the activation of new hotels, therefore reducing the cost.

On the one hand, when the number of wavelengths is limited, a proper choice of the sequence of nodes in which the distributed algorithm is performed can lead to near-optimal solutions. On the other hand, this choice requires complete knowledge of the network and can be performed only at a higher level (i.e., in the network controller), thus the distributed strategy is not always capable of reaching optimality.

Figure 4.6 reports the number of backup BBU ports as a function of different distance constraints obtained with ILP and heuristic, averaged over the 50 cases, in 16 nodes network. From the figure, it is possible to notice that the two strategies are capable of reducing the number of backup ports with respect to the case in which there is no port sharing, which is 160 ports. For low values of distance constraint or, alternatively, when the number of active hotels is large, the number of required BBU ports is low, with the ILP that provides a better solution that the distributed strategy. On the other hand, when the distance constraint is large, the number of

Table 4.4: The maximum and the average number of hops, between RRUs and BBUs for ILP and heuristic with different limits over distance in the 16 nodes network with wavelengths constraint equal to 80.

Maximum distance in [hops]	Distance [hops]			
	ILP		H-80	
	Max	Avg	Max	Avg
1	1	0.75	1	0.66
2	2	1.19	2	1.25
3	3	1.63	3	1.78
4	4	2	4	2.22
5	4	2	5	2.5
6	4	2	6	2.5

ports reach 160 in case of ILP, because with only two active hotels no port sharing is possible. In these cases instead, the heuristic performs slightly better, having more than 2 active hotels and therefore allowing some port sharing.

To show how the two strategies react to network changes, figures 4.7 and 4.8 show a sample of transition from 16 to 17 nodes network when the maximum allowed distance is 1 hop. When a centralized view of the network is available, to reach the optimal solutions is necessary to run the algorithm and find the new optimum. Using the ILP (figure 4.7) may lead to many changes in the active nodes, requiring heavy migration of BBUs from one hotel to another. In this case, BBU hotel 3,5,8,14 can be deactivated and BBU hotels 4,6,11,13,17 must be activated. With the heuristic instead (figure 4.8), only node 17 is activated in addition to already active hotels. In both strategies, the new node (17) must be activated, since it cannot be connected to 2 active hotels (one for primary and one for backup purposes). However, the distributed approach allows the node to find hotels independently, leaving the rest of the network untouched. This approach is more incremental than the ILP, and therefore more suitable to dynamic scenarios.

Figures 4.9 and 4.10 depicts another example of the same transition, but with a limit on the maximum allowed distance equal to 3 hops. Once again, the ILP case (figure 4.9), which accounts for the cost of BBU hotels activation, the distance between RRUs and BBUs, and BBU ports, requires changes in the network configurations, even if no BBU hotel is activated. The heuristic instead (figure 4.10), even though it requires one more active hotel than ILP, keeps the network configuration untouched when passing from a 16 to a 17 nodes network.

Figure 4.11 shows the active BBU hotels in the 36 nodes network required by ILP and heuristic, on average. From the figure, it is possible to notice that the ILP requires less active BBU hotels when the distance constraint increases. The hotels required by the heuristic follow the ILP trend, but they increase when the distance constraint is 6. Similarly to the 16 nodes network, this is due to the limit on the wavelengths, that forces the algorithm to activate more BBU hotels. The case of 5 and 6 hops are not reported in the ILP case due to the problem of complexity. While in the case of 16 networks the time required to solve the model with the ILP is in the order of tens of seconds, in the 36 nodes network this time increases to tens of minutes for a distance constraint less than 4 hops, while for larger values the
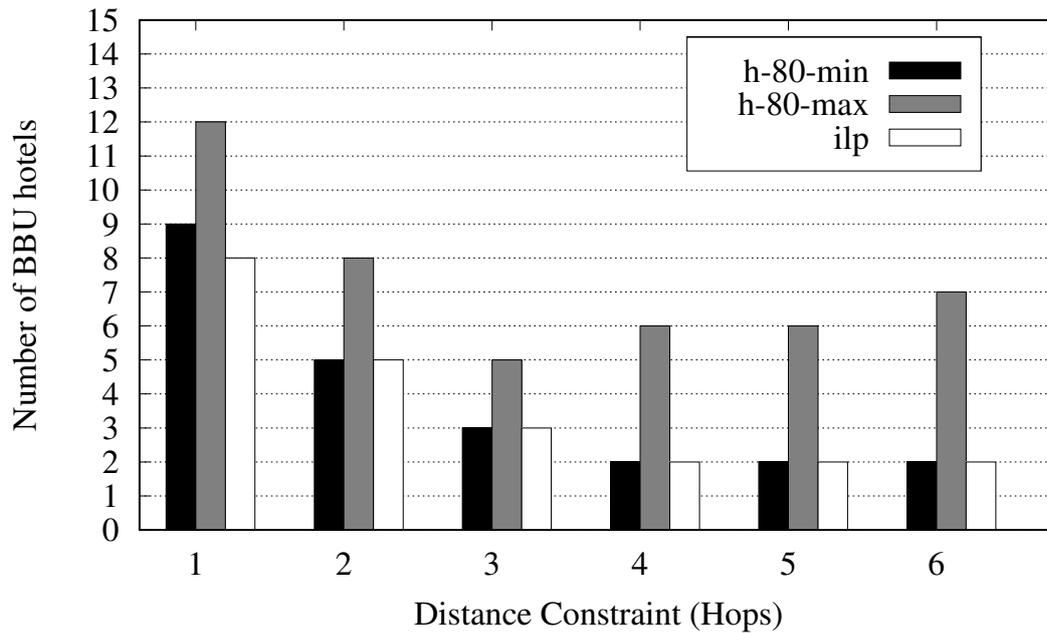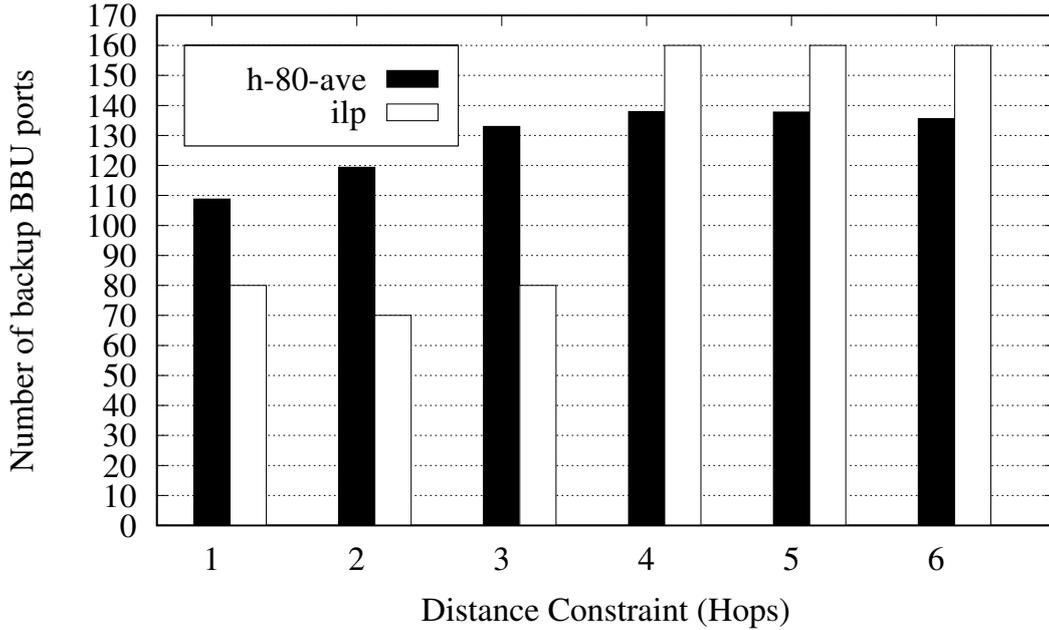
Figure 4.6: The number of backup BBU hotel ports required by ILP and heuristic, averaged over 50 cases for different distance constraints in the 16 nodes network, with wavelength constraint equal to 80.

complexity of the instances makes not possible to find a solution.

Figure 4.12 depicts the total number of wavelengths required by ILP and heuristic, averaged over all the 50 cases. The total amount of wavelengths increases when the distance constraint increases. When the distance constraint increases, farther BBU hotels can be reached, increasing the overall amount of wavelengths that are needed to connect RRUs and BBUs. The absolute difference between the two strategies also increases with the distance constraint, due to the inability of the heuristic to properly choose locations for BBU hotels.

4.5 Case study I: Distributed Machine Learning Location Algorithm

Classical optimization approaches, either heuristics or ILP, typically assume that network topology is known and related procedures can find optimal or sub-optimal solutions given constraints and elementary costs. These approaches require a certain amount of global knowledge, i.e. the network topology, that is impractical when dealing with evolving C-RAN configurations. Even though these networks are typically managed by an SDN controller, the self-configuration capability is recommended to support flexibility, scalability and service continuity in the presence of failure.

Machine learning (ML) approaches are recently emerging as a viable solution to cope with dynamic contexts such as those represented by C-RAN. Applications of ML algorithms to self-organizing cellular networks have been recently described in [53], where 5G C-RAN has been also addressed as a potential future research direction.

Figure 4.7: An example of evolution from 16 to 17 nodes network using ILP with the maximum allowed distance equal to 1 hop. The active BBU hotels are highlighted in blue.

The ML approach is expected to be effective also in C-RAN dynamic reconfiguration needs. ML provides a framework to define algorithms that proceed in learning some properties of the system to obtain some performance target [54]. In the meantime, the elements of the system are enriched with information that turns to be useful in the evolution of the system. These characteristics make the approach suitable for the BBU hotel assignment in C-RAN.

In this context, a two-phase ML Distributed Facility Location Algorithm (ML-DFL) is here proposed to locate BBU hotels in a C-RAN while supporting reliability against a single failure. The description of Training Data Sets (TDS) and tasks are given for the two phases and the performance targets are defined. Relationships between data and SDN control planes are outlined and results to quantify the degradation with respect to optimal ILP discussed.

4.5.1 Definition of the ML-DFL Algorithm

The definition of an ML algorithm requires to identify a sequence of tasks that sequentially operate on a TDS to produce updated performance evaluation. An aspect of the system that the ML algorithm helps to learn needs to be identified. The proposed ML-DFL algorithm for BBU hotel location in C-RAN is represented in figure 4.13. Two sequences of tasks indicated as *Phase 1* and *Phase 2*, each with a specific learning objective, are outlined. The SDN orchestrator in the control plane is assumed to initiates the procedures and takes advantage of the learning achieved in each phase. This procedure reduces the rate of interactions between the data plane and the SDN controller, improving scalability and performance [55].

Reinforcement Learning(RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its actions and experiences. An agent takes actions; for example, a drone making a delivery, or the algorithm is the agent. Action in this context is the set of all possible moves the agent can make. An action is almost self-explanatory, but it
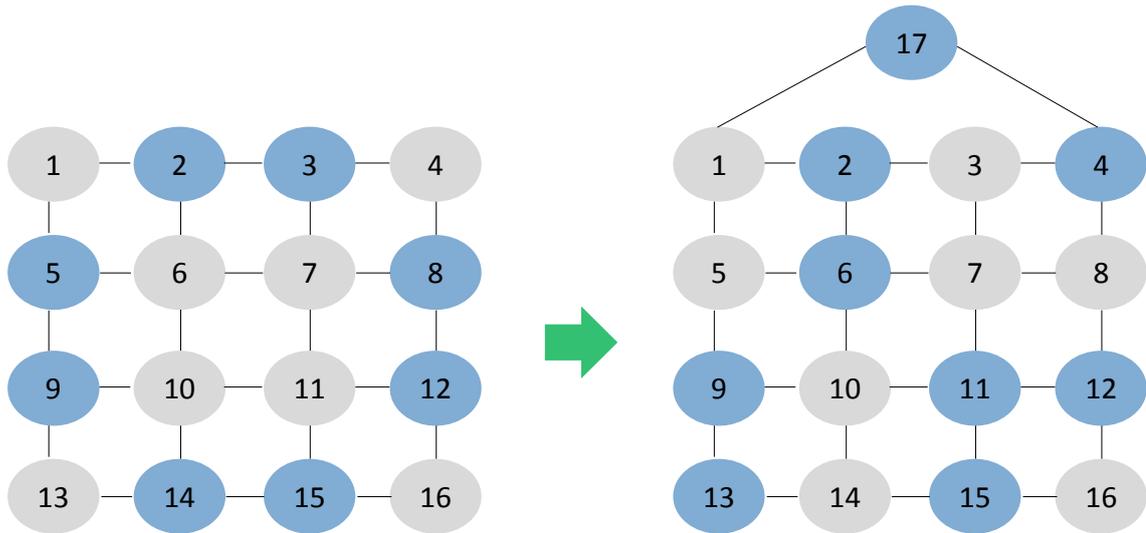
Figure 4.8: An example of evolution from 16 to 17 nodes network using heuristic with the maximum allowed distance equal to 1 hop. The active BBU hotels are highlighted in blue.

should be noted that agents usually choose from a list of discrete, possible actions [56].

In Supervised learning, you train the machine using data that is well "labeled." It means some data is already tagged with the correct answer. It can be compared to learning which takes place in the presence of a supervisor or a teacher. A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data. Successfully building, scaling, and deploying accurate supervised machine learning Data science model takes time and technical expertise from a team of highly skilled data scientists. Moreover, Data scientist must rebuild models to make sure the insights given remains true until its data changes.

Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data. Unsupervised learning algorithms allow you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning deep learning and reinforcement learning methods [57].

Though both supervised and reinforcement learning use mapping between input and output, unlike supervised learning where feedback provided to the agent is the correct set of actions for performing a task, reinforcement learning uses rewards and punishment as signals for positive and negative behavior.

As compared to unsupervised learning, reinforcement learning is different in terms of goals. While the goal in unsupervised learning is to find similarities and differences between data points, in reinforcement learning the goal is to find a suitable action model that would maximize the total cumulative reward of the agent.

ML-DFL, initiated by the SDN controller, starts from a C-RAN configuration where no BBU functionality is assigned to the node. The algorithms performed by each phase are completely distributed, meaning that no global knowledge is required. Each node i is assumed to know the RRUs connected to it and, after an
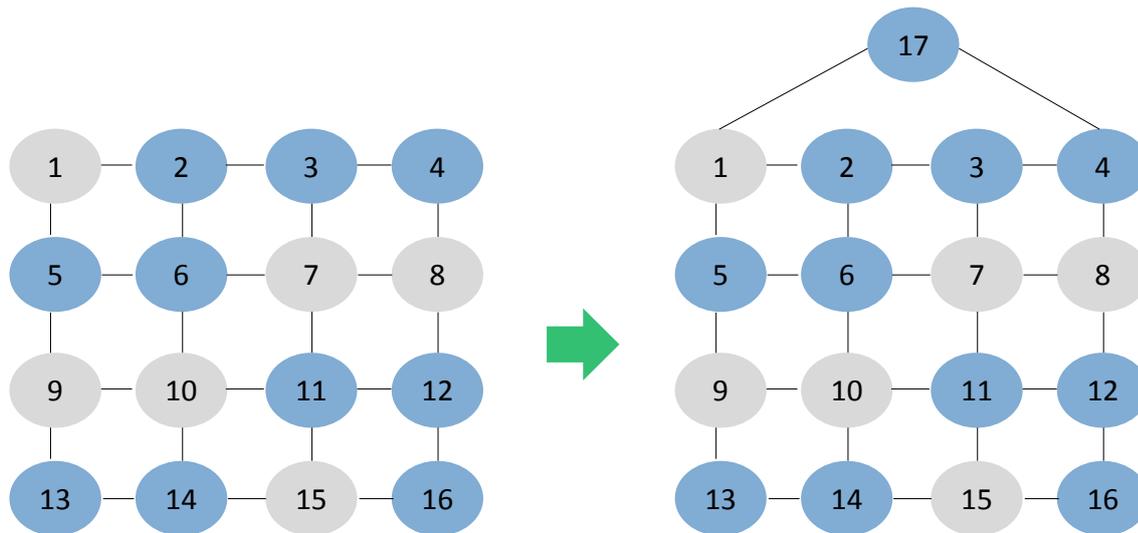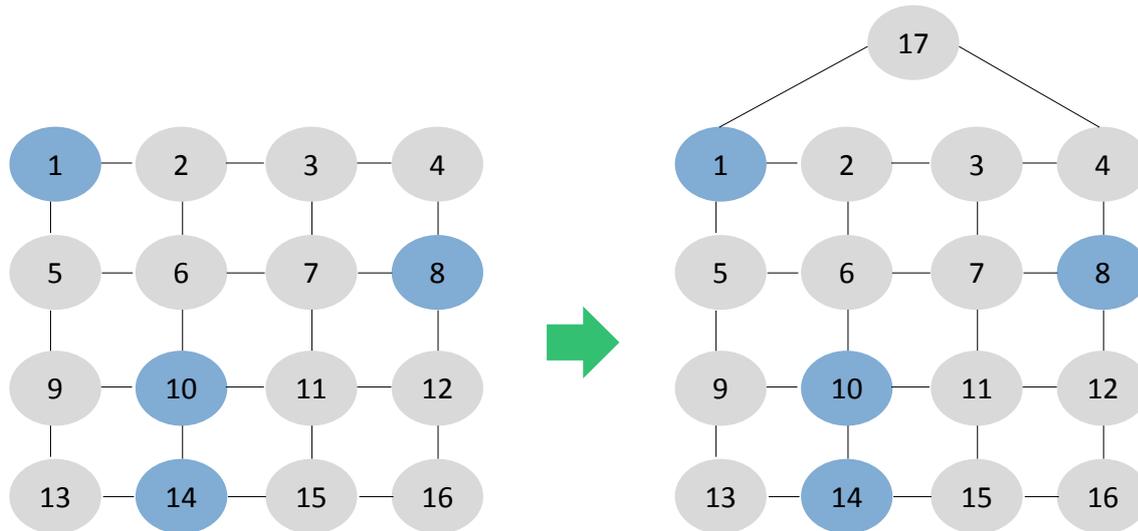
Figure 4.9: An example of evolution from 16 to 17 nodes network using ILP with the maximum allowed distance equal to 3 hops. The active BBU hotels are highlighted in blue.

Table 4.5: Definition of ML-DFL elements.

Phase	TDS	Performance	Target
1	M, A	$\frac{A_i}{A_{i-1}}$	1
2	H, F	F	Minimize F

initial neighbor discovering phase, the set of neighboring nodes. Each phase can be described as supervised learning whose TDS and target performance are presented in table 4.5

In particular, *Phase 1* performs its tasks to achieve complete service and protection for installed antennas. The training dataset for *Phase 1* is defined by matrix M, representing the links between each couple (i,j) of C-RAN nodes, and by total number A of served antennas. What is learned during phase 1 is the content of the matrix M and the number of antennas A, which are updated by performing each task. The performance of *Phase 1* is defined by the ratio $\frac{A_i}{A_{i-1}}$ of served antenna before and after the execution of task i, whose target value is 1, meaning that the complete set of RRU is connected and protected. *Phase 2*, after having learned about antennas, aims at achieving BBU hotel location sharing for cost optimization. During phase 2, the learning of the matrix H, containing information on primary and backup BBU hotels, is performed by discovering how to share backup BBU hotel among RRUs with distinct primary BBU hotels. The performance of *Phase 2* is represented by the cost function $F = C_B + C_H$, where C_B and C_H are the costs of activating a BBU hotel and the cost in hops of the solution, respectively, whose target is to be minimized.

4.5.2 Numerical Results

The network considered to present sample results is the same presented in figure 4.2. Each node is assumed to have 10 RRUs. F is calculated with both the ML-DFL
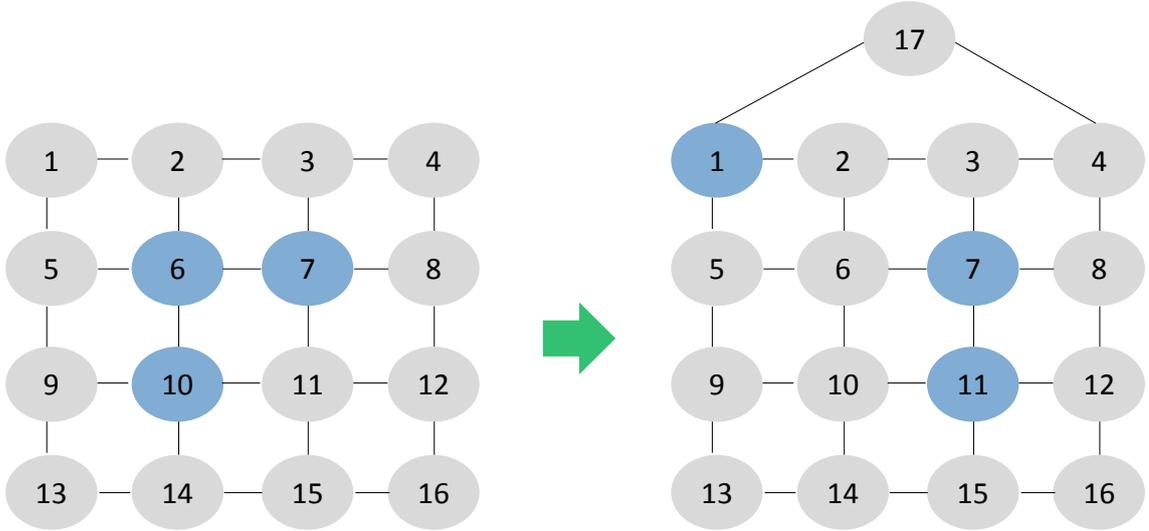
Figure 4.10: An example of evolution from 16 to 17 nodes network using heuristic with the maximum allowed distance equal to 3 hops. The active BBU hotels are highlighted in blue.

Table 4.6: Total number of wavelengths.

MAX_D	ML-DFL	ILP
1	250	240
2	470	420
3	580	600
4	670	640
5	700	640

and a suitably ILP as shown in figure 4.14. The number of wavelengths needed to interconnect C-RAN nodes is also presented in table 4.6 with different distance constraints (MAX_D). As expected, the ML-DFL, as a sub-optimal approach, overestimates costs but finds values very close to those calculated by ILP.

4.6 Case study II: Reliable Deployment for Vehicular Networks

The globally connected car market is growing rapidly. Novel services will be offered to vehicles, many of them requiring low-latency and high-reliability networking solutions. The Cloud Radio Access Network (C-RAN) paradigm, thanks to the centralization and virtualization of baseband functions, offers numerous advantages in terms of costs and mobile radio performance. C-RAN can be deployed in conjunction with a Multi-access Edge Computing (MEC) infrastructure, bringing services close to vehicles supporting time-critical applications. However, a massive deployment of computational resources at the edge may be costly, especially when reliability requirements demand the deployment of redundant resources. In this context, cost optimization based on ILP may result in being too complex when the number of involved nodes is more than a few tens. This section proposes a scalable approach
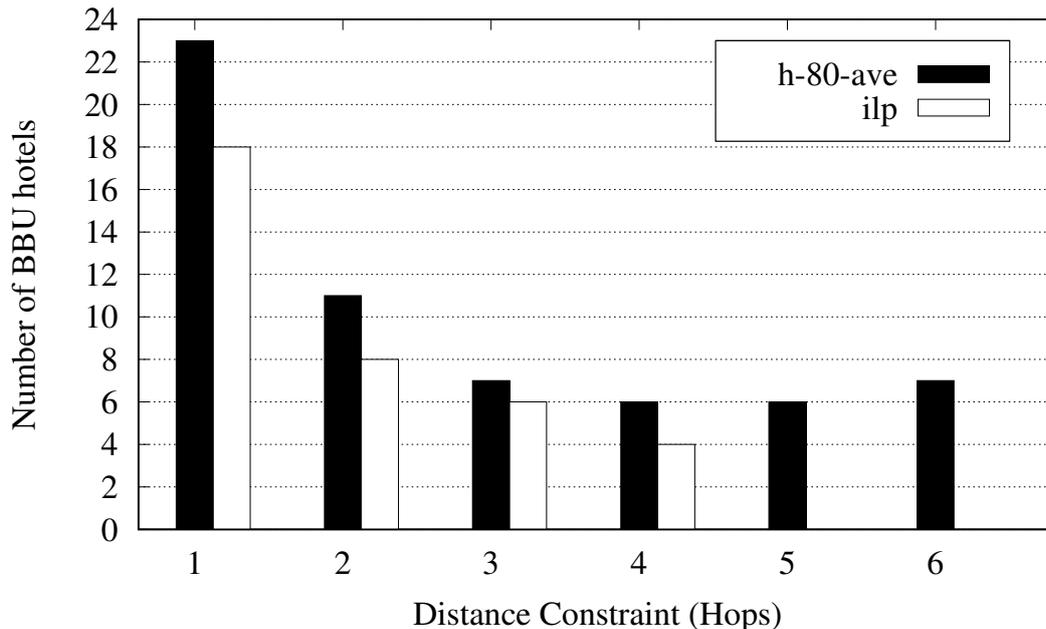
Figure 4.11: The number of active BBU hotels required by ILP and heuristic, averaged over 50 cases, for different distance constraints in the 36 nodes network, with wavelength constraint equal to 80.

for C-RAN and MEC computational resource deployment with protection against single-edge node failure.

C-RAN architecture can be used as an enabler for vehicular communications providing network assistance and commercial services, as depicted in figure 4.15. Vehicles communicate directly with the mobile network or with Road Side Units (RSUs), that send collected data through the mobile network. Data concerning low-latency applications can be elaborated directly in the edge nodes, thanks to the computational resources offered by the MEC. Computational resources in edge nodes can be used for (i) virtual baseband processing; (ii) virtual mobile core network functions; and (iii) edge application services [58]. Non-time-sensitive data can be delivered to applications performed in remote locations (not reported in the figure).

The traffic destined to remote cloud resources is user-dependent and requires lower bandwidth with respect to fronthaul requirements [59] and is out of the scope of this chapter. In this work, we propose to co-locate, within the same edge node, cloud and BBU processing functions. An edge node is considered to be active when it hosts physical or virtual functions, either for BBU processing or edge core/cloud services.

To provide a reliable C-RAN against single node failures, a $1 + 1$ protection solution is desirable to avoid temporary service outages due to resource restoration. Primary and backup path resources must be allocated to provide resiliency against hardware failures. This work considers single active edge node failures (i.e., a failure of all servers placed in an active edge node). The formulation of the joint BBU hotel and edge cloud processing location problem with resiliency is as follows:

- **Given** a set of RRUs to be connected to active edge nodes, a set of edge nodes (candidates to host BBU and edge processing resources), and a set of links connecting edge nodes.
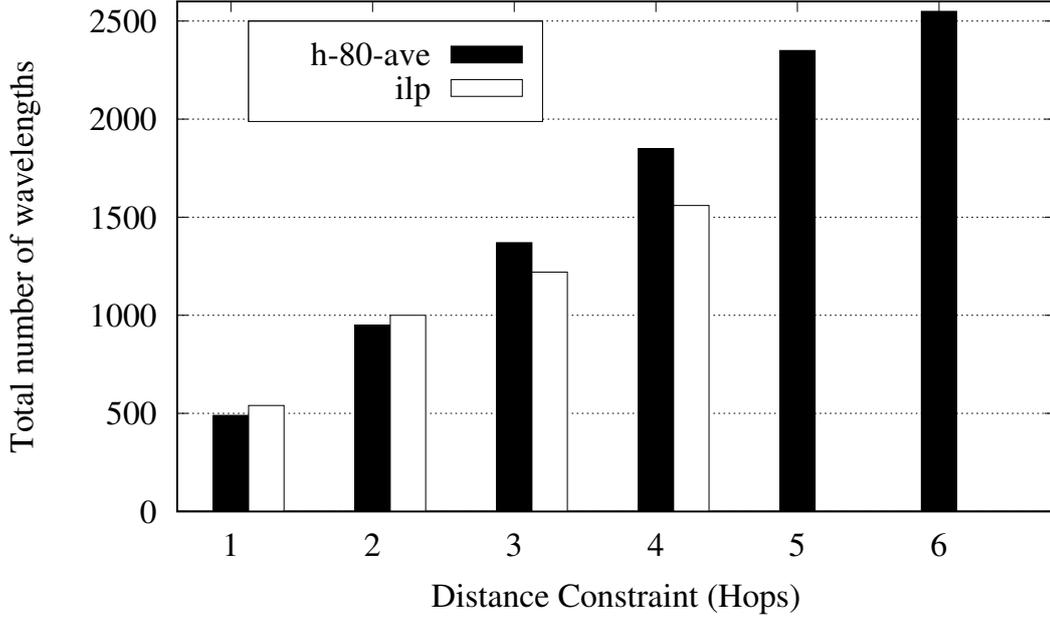
Figure 4.12: The total number of wavelengths required by ILP and heuristic, averaged over 50 cases for different distance constraints, in 36 nodes network with wavelength constraint equal to 80.
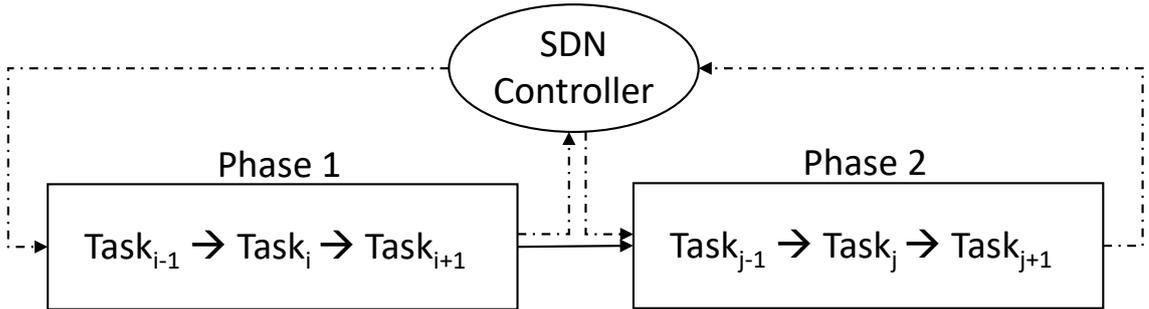
Figure 4.13: Architecture and phases of the ML-DFL.

- **Find** active edge nodes and suitable optical resource assignment such that (i) the number of active nodes and (ii) total wavelengths are minimized.

- **Ensure** that each RRU is connected to two active edge nodes (one for primary and one for backup purposes) and that the maximum available wavelengths per link and maximum allowed distance to provide target service are not exceeded.

4.6.1 Two-Phases Hybrid Approach

The hybrid approach proposed here is performed in two phases. In the first phase, a heuristic is proposed to provide a computationally simple but reliable C-RAN coverage by guaranteeing that each RRU has both a primary and a backup node and that minimum delay is achieved. The second phase is an optimization process, based on a modified version of the ILP proposed in the paper [60], which aims at reducing the number of active nodes found in *Phase 1*. The details of the hybrid algorithm are reported below.
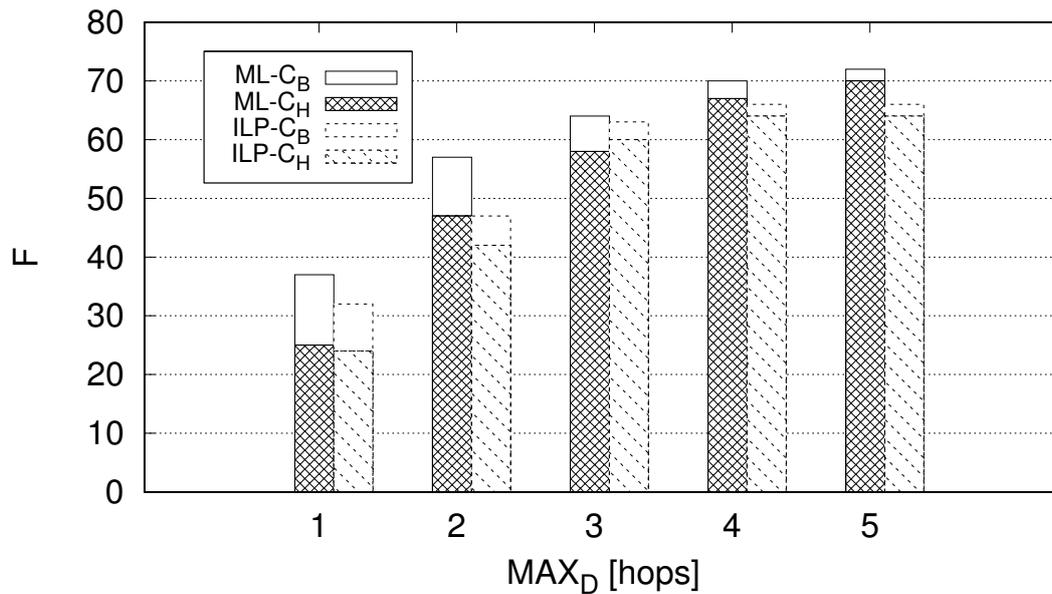
Figure 4.14: Cost F: ML-DFL vs. ILP.

textitPhase 1 is assumed to start from a C-RAN configuration where no edge node is active, i.e., BBU and edge functionalities have yet to be assigned to nodes. This has, anyway, no impact on the generality of the approach. In this phase, the edge node activation is performed within a 1 hop distance or, equivalently, RRUs can be connected only to the node itself or to a neighbor edge node. This implicitly assumes that there are enough resources on the links connecting neighbors and guarantees that delay constraints are always satisfied. It should be noted that to solve the deployment problem, primary and backup nodes must be selected. Therefore, not satisfying the aforementioned condition on the link resources does not guarantee a solution to the problem.

In addition to the C matrix needed to model the physical links (see table 4.7), two additional structures are introduced here:

- H matrix: This is a $n \times 2$ matrix, where each row represents a node of the network; the first column indicates which is the primary edge node chosen by the node on that row, while the second column indicates which is the backup node.

- W matrix: This is a $n \times n$ matrix which keeps track of the use of the links between nodes. In W, there is one row for each source edge node (where the RRUs are physically connected). W has one column for each edge node, that is, the possible locations for the edge server performing baseband and services for the specific RRUs. This matrix is needed to provide a feasible solution at the end of *Phase 1* but is not used in *Phase 2*.

Algorithm 7 presents the pseudo-code of the algorithm executed by each node of the network during *Phase 1*. In the beginning, the algorithm starts with empty H and W matrices (line 2). This algorithm executed in a sequence for each node until all nodes in the network have both primary and backup connections (a condition in line 4). Then, node i checks some conditions for the primary and the backup
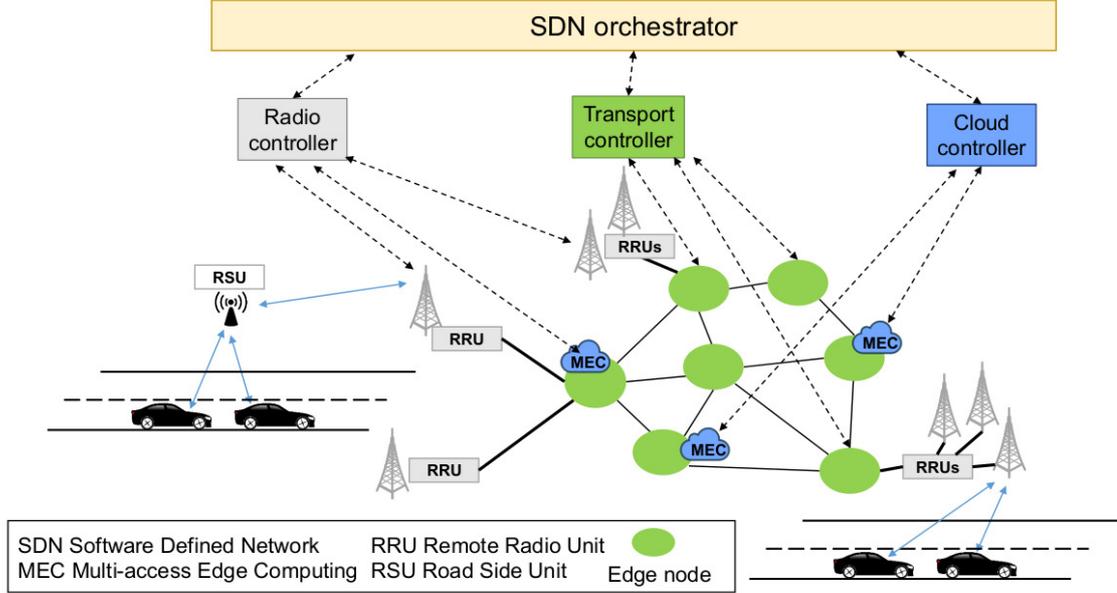
Figure 4.15: Software-Defined Networking (SDN)-controlled Cloud Radio Access Network (C-RAN) architecture for vehicular communications.

Table 4.7: Notation used in this section.

Parameter	Definition
N	Set of edge nodes in the network, $|N| = n$.
R_s	Number of sources (RRUs) directly connected to $s \in N$.
C	$n \times n$ matrix. $c_{ij} = 1$ if node i is directly connected to node j, 0 otherwise.
h_d	Binary variable equal to 1 if edge node $d \in N$ is active, 0 otherwise.
M^W	Maximum available wavelengths in each link.
M^H	Maximum allowed distance between RRUs and edge nodes.

connection to find suitable edge nodes. If node i is already active (line 6), it can use itself as the primary edge node (line 7). Otherwise, node i must search among its neighbors to find an already active node (line 8) and, if it succeeds, makes the primary connection to the edge node j (line 9) and updates W matrix accordingly (line 10). The updating phase stores in the position i, j of the matrix the required wavelengths over link i–j. If no neighbor is active (line 11), node i activates itself and makes the primary connection to itself (lines 12 and 13).

After establishing the primary connection, node i executes a set of instructions to find the backup edge node. There are two possible situations. The first situation is when node i is already active and plays the primary role for the RRUs connected to itself or not active at all (line 16). In this case, node i either finds a directly connected neighbor node (j), which is already active and satisfies the distance restriction, and connects to it (lines 17–19) or chooses randomly one of the neighbors as a backup, defines the backup connection, and updates W matrix accordingly (lines 20– 23). The other situation happens when node i is active (line 25). Node i can take advantage of this situation and makes the backup connection to the local edge node

(lines 26 and 27). *Phase 1* stops when all nodes in the network have both connections to primary and backup nodes.

The objective of the second phase is to minimize the number of active nodes. This is achieved by reassigning the RRU connections and shutting down active nodes by further centralizing BBU and edge processing functions within the distance constraints (M^H).

Algorithm 7 C-RAN reliable coverage (*Phase 1*).

1: **Initialization:**
2: $H, W \leftarrow \emptyset$
3: **Begin:**
4: **while** exists node $i \in N$ s.t. $(H_{i0} = 0) \vee (H_{i1} = 0)$
5: //Primary connection assignment:
6:     **if** $h_i = 1$
7:         $H_{i0} = i$
8:     **else if** $\exists$ node j s.t. $c_{ij} = 1$ **and** $h_j = 1$
9:         $H_{i0} = j$
10:         **update** W
11:     **else**
12:         $h_i = 1$
13:         $H_{i0} = i$
14:     **end if**
15: //Backup connection assignment:
16:     **if** $(h_i = 1$ **and** $H_{i0} = i)$ **or** $(h_i = 0)$
17:         **if** $\exists$ node j s.t. $c_{ij} = 1$ **and** $h_j = 1$
18:             $H_{i1} = j$
19:             **update** W
20:         **else**
21:             **activate** random neighbor j $(h_j = 1)$
22:             $H_{i1} = j$
23:             **update** W
24:         **end if**
25:     **else**
26:         $h_i = 1$
27:         $H_{1i} = i$
28:     **end if**
29: **end while**
30: **End**

4.6.2   Numerical Results

Numerical results are obtained in different networks to evaluate the effectiveness of the ILP and hybrid solutions in terms of active edge nodes and of the centralization gain, G_C, that is the advantage related to centralizing BBU and cloud functionalities, expressed by the following formula:
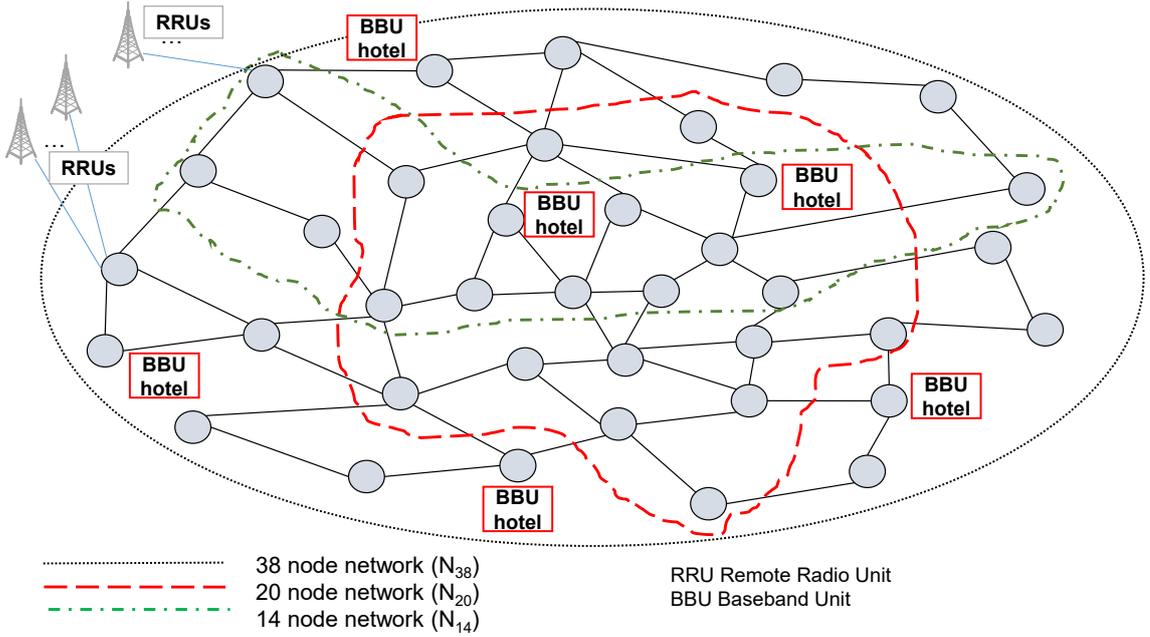
Figure 4.16: N_{38}, N_{20}, and N_{14} C-RAN topology for numerical evaluations.

$$G_C = \frac{n - \sum_{d \in N} h_d}{n} \tag{4.14}$$

where n and h_d have been defined in table 4.7. Three sample networks, N_{38}, N_{20}, and N_{14}, consisting of 38, 20, and 14 nodes, respectively, are considered, as represented in figure 4.16. Evaluations assume here that 10 RRUs are physically connected to each node to provide mobile network coverage and transmission capacity for the vehicular network, and the adoption of CPRI. The proposed algorithms and evaluations can be extended to different numbers of RRUs, possibly unbalanced among edge nodes and suitably adapted to different functional split, which is left for future works.

In figures 4.17, 4.18, and 4.19, comparisons are reported between the hybrid and the ILP approaches by plotting the results in terms of the number of active edge nodes as a function of the allowed distance, expressed in hops. The cost of the hybrid solution depends on the node from which the heuristic procedure starts: the maximum and minimum costs in terms of the total number of active nodes obtained are both reported in the plots. Besides, the results at the end of *Phase 1* of the hybrid strategy are also shown, as lines and denoted as H, to outline the effect of the optimization phase. These lines are constant because they do not depend on the distance, as they provide a solution within 1 hop distance. The costs obtained with the hybrid and ILP approaches decrease with the distance in all networks. The minimum value that can be achieved is 2 because one primary and one backup node must be always present to cope with single edge node failure. In case of tight distance constraints (e.g., 1 or 2 hops), data cannot be transported far in the network; thus, many edge nodes must be activated. When the distance constraint increases, farther nodes in the network can be reached and, consequently, the number of total active nodes decreases.

From the figures, it can be seen also the influence of the starting node, represented by the difference between the maximum and the minimum costs. In the worst cases,
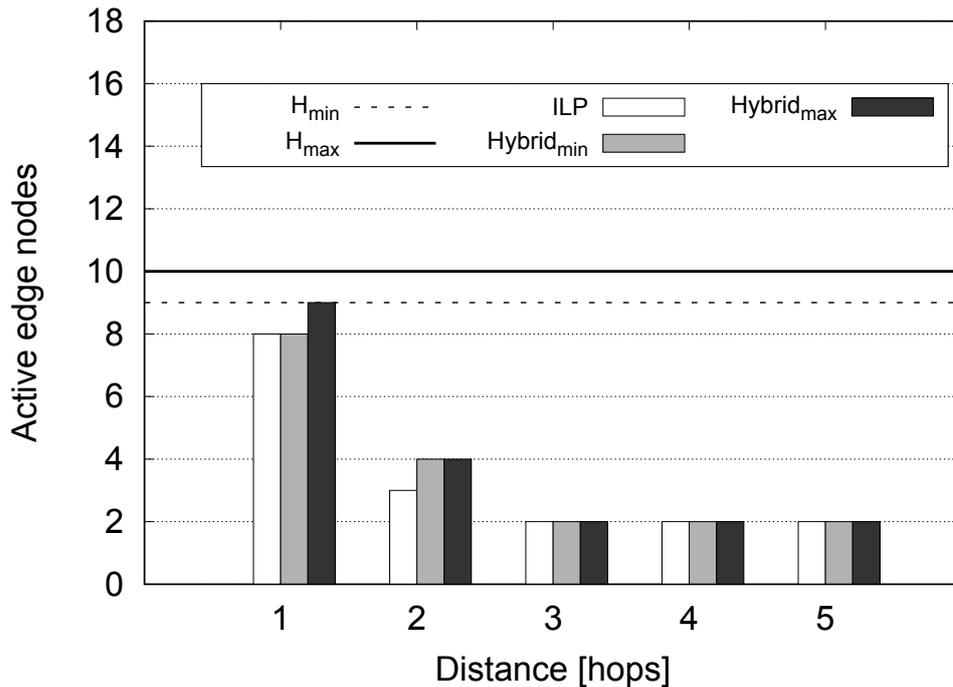
Figure 4.17: The total number of active edge nodes as a function of the allowed distance between RRUs and edge nodes for network N_{14}: Maximum and minimum costs of the hybrid results are reported after both phases.

only one additional node must be activated. Also, the results of the hybrid are shown to be the same as the optimal ones in most cases. However, in very few cases, the hybrid approach cannot achieve optimal solutions due to the choices performed in *Phase 1*, where some nodes are excluded by the pool of possible active nodes and cannot be activated in *Phase 2*.

In figure 4.20, the gain of centralization of BBU and edge cloud functionalities is presented as a function of the allowed distance from RRUs by comparing the ILP results with the results of the hybrid approach at the end of *Phase 1* (denoted as H) and *Phase 2* in the maximum-cost case. This gain is relevant both for ILP and hybrid, with the hybrid being very close or coincident to the optimal solution. In the worst case (i.e., distance constraint equal to 1 hop), the hybrid provides only 8% gain reduction. As expected, *Phase 1* provides only suboptimal solutions. It is, therefore, evident the role of *Phase 2* of the hybrid approach in achieving a high centralization gain with respect to the plain coverage achieved in *Phase 1*.
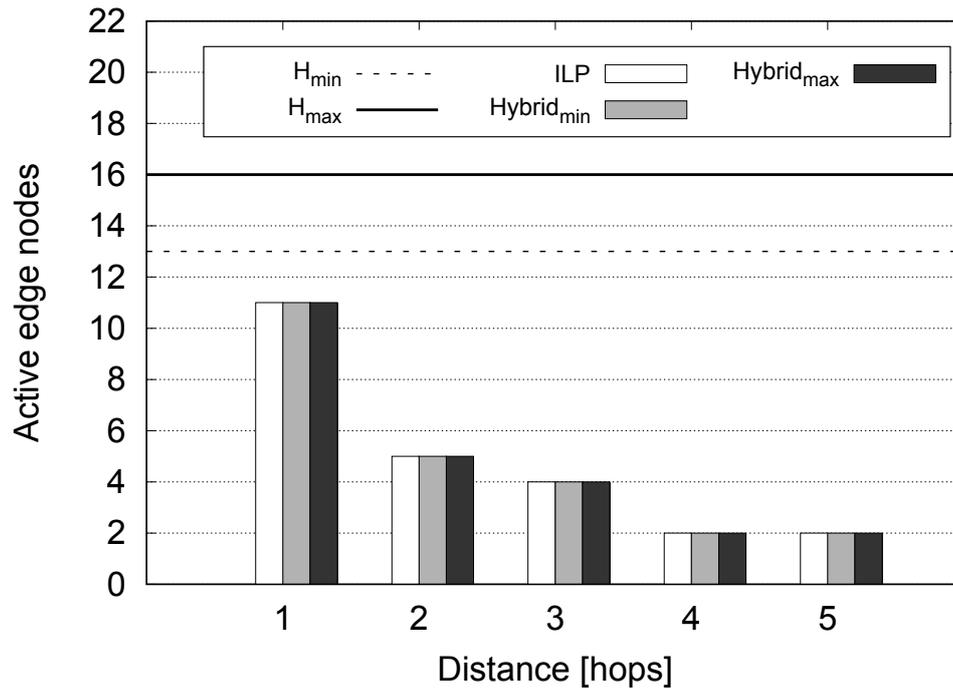
Figure 4.18: The total number of active edge nodes as a function of the allowed distance between RRUs and edge nodes for network N_{20}: Maximum and minimum costs of the hybrid results are reported after both phases.
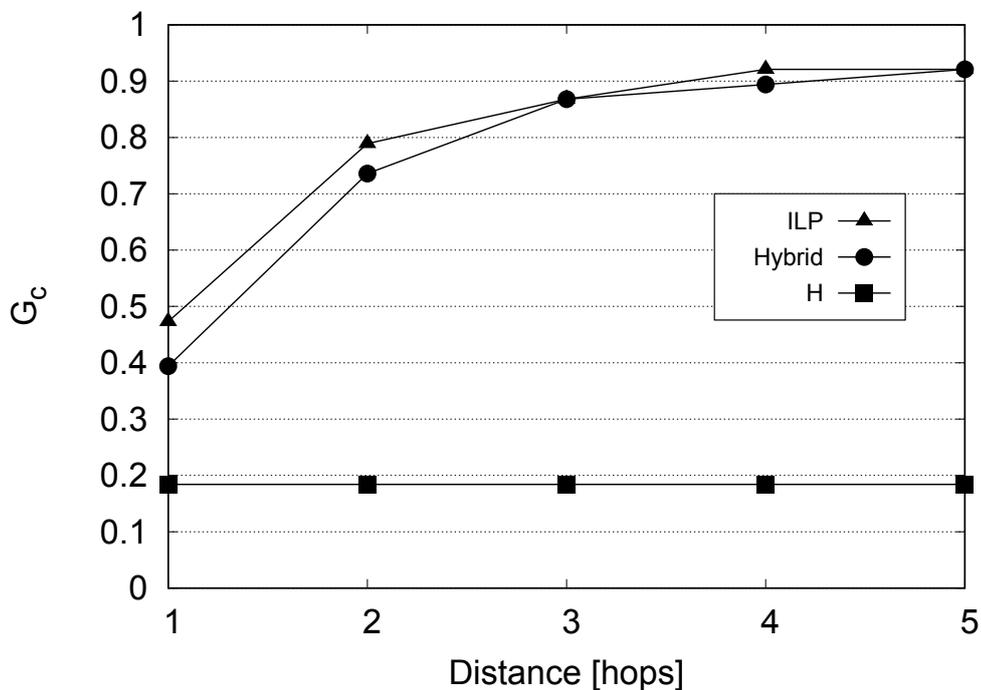
Figure 4.20: Centralization gain as a function of the allowed distance between RRUs and edge nodes for network N_{38}: Results are reported for the maximum cost for hybrid (*Phase 1* and *Phase 2*), and ILP.

Table 4.8 reports the number of active links, wavelengths over the most used link,
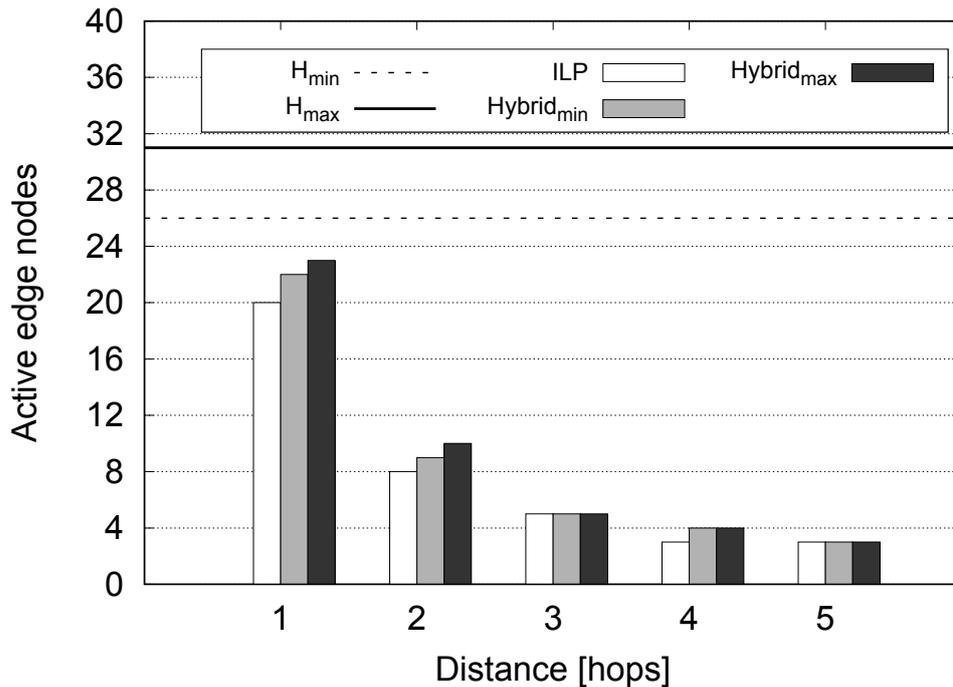
Figure 4.19: The total number of active edge nodes as a function of the allowed distance between RRUs and edge nodes for network N_{38}: Maximum and minimum costs of the hybrid results are reported after both phases.

and overall wavelengths in network N_{38} for the two strategies. By comparing the strategies, it is possible to observe that the ILP requires a slightly higher number of wavelengths with respect to the hybrid approach when the number of active nodes is lower (distance constraints 1, 2, and 4). Nevertheless, because the activation cost of a node is much larger than the cost of a wavelength, the ILP solution always reaches a lower cost solution compared with the hybrid approach. When the ILP and hybrid require the same amount of active nodes (distance constraints 3 and 5) the ILP requires fewer wavelengths than the hybrid approach due to a wider set of choices. This happens for similar reasons also for the wavelengths required over the most used link.

Table 4.8: The number of active links, wavelengths over the most used link, and total wavelengths for the hybrid and ILP for different distance constraints in network N_{38}.

Dist.	Hybrid			ILP		
[hops]	Active	Max	Total	Active	Max	Total
1	45	10	530	48	10	560
2	51	40	950	50	40	1040
3	49	70	1370	51	60	1350
4	52	70	1530	48	80	1830
5	51	80	1790	52	80	1780

4.7 Conclusion

This chapter addresses the problem of providing low latency and reliable services in a cost-efficient way using 4G and 5G networks. Centralized and distributed algorithms have been proposed and compared in this section to solve the BBU hotel location problem in C-RAN. The results obtained by the distributed algorithms are sub-optimal with respect to the centralized approach.

Chapter 5

Baseband Functional Splitting Analysis for 5G Access Network

5.1 Introduction

5G has become the hottest study topic in both industry and academia in the past few years. Compared with 4G Long-term Evolution (LTE) networks, future 5G is expected to provide end-users with unprecedented user experience in terms of data rate, ultra-low latency, and universal access. In addition to enhanced Mobile Broadband (eMBB) service, 5G will exceed 4G systems with better support of two other kinds of applications: ultra Reliable Low-Latency Communications (uRLLC) and massive Machine-Type Communication (mMTC). With the capabilities, 5G is deemed to bring a fundamental transformation to human society [61].

It is clear that both core and Radio Access Networks (RAN) should evolve to accommodate 5G vision. Accordingly, new designs have to properly address key challenges and requirements to successfully achieve the vision of an inclusive, cohesive, and sustainable society. The future networks should be capable of handling the complex context of operations characterized by a tenfold increase in traffic [62], various mobility levels, and interference. In addition, multiple requirements need to be met including Quality of Experience (QoE) satisfaction, energy-efficient operation (90% improvement by 2020 [63]), resource efficiency, and cost-efficiency.

In this chapter, the technology evolution toward a possible architecture of the 5G access network namely Xhaul will be investigated. The evolution in designing a BS which fits into strict requirements drown. A section is dedicated to the function splitting techniques its advantages and requirements. At the end of this chapter, the new transport protocol will be introduced. In the end, the main advantages and challenges of Xhaul are then described in detail.

5.2 Toward 5G Architecture Implementation

Given the mass volume of wireless cells that will be deployed in the 5G network, transporting huge amounts of data between thousands of cells and network core with low latency in a cost effective manner is a major challenge. To address the aforementioned challenges, the Xhaul architecture, aimed at developing the next generation of 5G integrated backhaul and fronthaul networks enabling a flexible
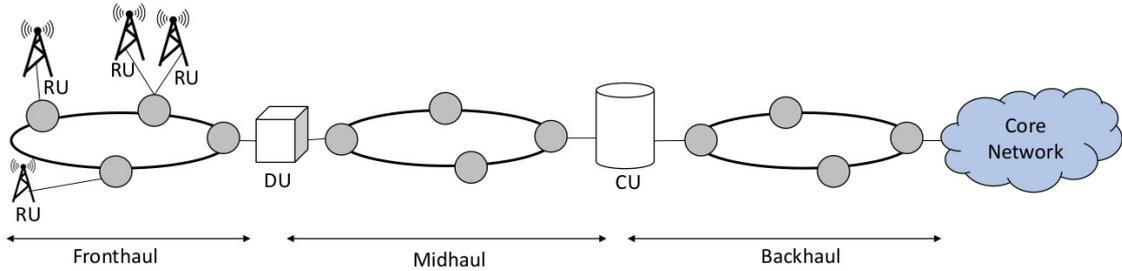
Figure 5.1: Scheme of the Xhaul network.

and software defined reconfiguration of all networking elements in a multi tenant and service oriented unified management environment. The envisioned Xhaul transport network will consist of high capacity switches and heterogeneous transmission links (e.g., fiber or wireless optics, high-capacity copper, or millimeter wave) interconnecting remote radio units, pooled processing units (mini data centers), and Points of Presence (PoPs) of the core networks of one or multiple service providers.

This requires completely new physical layer technologies or a radical evolution of existing ones, such that the challenging 5G performance requirements can be met. The Xhaul architecture will use a novel unified data plane protocol able to transport both backhaul and fronthaul traffic, regardless of the functional RAN split [64].

The methodology presented here is referred to as the 5G network architecture as defined by 3GPP [65]. This architecture consists of two parts: the radio access network and the core network. The radio access network is expected to be based on the Xhaul concept which differs from current implementation in many ways. First, it extends between the user and the base station, which is called "gNodeB" (gNB). The gNB consists of three logical entities: Central Unit (CU), Distributed Unit (DU) and Remote Unit (RU). One gNB could contain one CU and multiple DUs and several RUs. In this sense, a gNB is a kind of mini-C-RAN. Each split option comes with different requirements such as latency, bandwidth, and usage of Processing Units (PU).

Figure 5.1 shows a 5G logical network architecture as divided into 3 parts. Fronthaul is the network segment from RU till the corresponding DU. The distance of these two entities can not be more than 20 km due to the delay-sensitive functionalities which will be executed in DU. Normally the bandwidth in this segment is the highest because of the low layers splits. The network segment between DU and responsible CU, where upper layer BBU functionalities are performed, is called midhaul. Several DUs can reside in this part of the network which is connected to the same CU. The distance in this segment is more relaxed (80-100 km), compared to the fronthaul, due to more relaxed delay requirements of upper layers splits. The third part is the backhaul which is extended between gNB and the core network.

In order to relax the stringent fronthaul requirements, functional splits between the DU and CU are defined [66], [67]. The functional split refers to a division of signal processing functionalities between the DU and CU. 3GPP has identified eight functional splits with different suboptions. Besides, CPRI released a new version of CPRI called eCPRI [68], which already uses new splits. However, CPRI and eCPRI do not deliver a full interface standardization that would allow true interoperability among different vendors. On the other hand, the recently formed *xRAN* fronthaul working group supports an open, interoperable and efficient fronthaul interface.
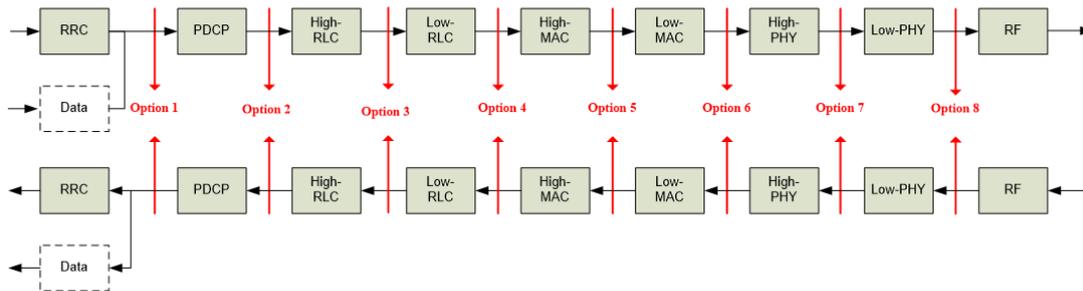
Figure 5.2: 3GPP functional splits.

Table 5.1: Functional splits analysis.

Use case	One-way latency	DL bandwidth	UL bandwidth
RRC-PDCP	30 ms	151 Mbps	48 Mbps
PDCP-RLC	30 ms	151 Mbps	48 Mbps
RLC-MAC	6 ms	151 Mbps	48 Mbps
Split MAC	6 ms	151 Mbps	48 Mbps
MAC-PHY	250 μs	152 Mbps	452 Mbps
PHY-RF	250 μs	1966 Mbps	1966Mbps

Several different functional splits are currently being investigated to be used for a New Radio access network (NR). In NR the radio processing and baseband functions from 3GPP protocol stack are split up into a DU and a CU. Figure 5.2 illustrates the LTE protocol stack for reference, as the NR protocol stack has not yet been announced. In figure 5.2, the processing functions closest to the antenna ports are located in the bottom, and moving upwards the signal is going through more and more processing before it is sent into the fronthaul network. 3GPP has proposed eight functional split options including several sub options. The arrows within figure 5.2 illustrate different options for functional splits, and the functions below arrow will be the functions implemented in the DU, where the functions above the arrow will be performed in the CU. The functions left in the DU are very close to the users as they will be located at the antenna mast, the functions located in the CU will benefit from processing centralization, and high processing powers within a data center referred to as the CU pool. The more functions located in the DU, the more processing has already been done before data is transmitted on the fronthaul network and the lower bit rate on the fronthaul network.

Table 5.1 illustrates the trade-off between (qualitative) gains and (quantitative) network requirements for different splits in LTE. Option 8 is equivalent to pure C-RAN, i.e., all functions are centralized enabling maximum gain, namely, interference coordination mechanisms such as Coordinated MultiPoint (CoMP) are enabled, computational resources are pooled and can be scaled based on demand, etc., at the cost of the toughest network requirements [69].

5.3   Function Split

The possibility of splitting up the BS functions in other ways than the traditional RRU-BBU split has been investigated in several papers. The majority of existing papers focus only on one or a few functional splits. The description of the functional splits follows the LTE protocol stack known from the traditional BSs. The lower part of this protocol stack includes three layers; the lowest is the physical layer, then follows the data link layer and on top is the network layer. These layers, together with the consequences of placing a functional split between specific elements, are introduced in the following.

In 5G *Phase 2* and *Phase 3* projects, several key technologies defined in 3GPP are taken as the baseline and enhanced with additional specific extensions to meet the requirements of the individual projects. Most implementations include CU-DU split, with some going further to also include exposure of Common Public Radio Interface (CPRI) creating a split in the radio equipment between a Remote Unit (RU), Distributed Unit (DU) and Centralized Unit (CU). Option 8 from figure 5.2 is the exposure of the CPRI interface, while Option 7 is referred to as enhanced CPRI (eCPRI). Both amount to the separation of the RU from the Base Band Unit (BBU). Of the remaining options, only *Option 2* has resulted in significant further work. *Option 2* consists in the separation of a distributed unit and a centralized unit, with the $F1$ reference point defined to connect the CU and DU. This point is worth emphasizing that depends on the network requirements and traffic pattern in some cases the entities can be jointly localized such as CU and DU or DU and RU.

5.4   Next Generation Fronthaul Interface

There have been lots of efforts put on addressing the fronthaul issues in both the industry and academia [70]. For example, different compression algorithms are proposed and analyzed to reduce the CPRI data rate [71]. The CPRI forum has begun the discussion on "Radio over Ethernet". The basic idea is to use Ethernet to transport the CPRI stream. In NGMN, schemes of the BBU-RRU function split are analyzed, aiming at reducing the fronthaul bandwidth to facilitate C-RAN deployment. In IEEE, a task force called *IEEE 1904.3* was founded recently, targeting the design of CPRI encapsulation on Ethernet packets [72].

The CPRI interface helps to separate the BBU and the RRU to enable the deployment of distributed base stations. CPRI has been working well for traditional mobile networks including *2G*, *3G*, and *4G* and has the following *3* traits.

- the CPRI line rate is constant regardless of traffic;

- the mapping between BBU and RRU is fixed one-to-one and not flexible;

- the sampling IQ data rate is dependent on the number of antennas.

With networks evolving to 5G, CPRI is becoming more and more unsuitable to accommodate evolution. It is well known that mobile traffic varies in the temporal dimension, which is called the tidal wave effect. For example, the data traffic in an office area is high in the daytime and yet plummets at midnight. For dense urban areas, the tidal wave effect is noticeable. However, the CPRI data stream

is synchronous, which means that it is constant regardless of the change of traffic. Even when there is no user traffic in the network, there are still CPRI streams running between the BBU and the RRU. This is a waste of bandwidth and leads to low utilization efficiency. Second, with CPRI an RRU is one-to-one correspond to a BBU. The relationship is configured offline. It may cause concern in the context of C-RAN. In C-RAN, the baseband units are centralized and virtualized in a pool.

Reliability becomes extremely important as each pool takes care of thousands of users. Therefore for the sake of protection, it would be desirable if, in C-RAN, one RRU could be automatically switched to another BBU hotel. Current CPRI however, does not support such flexible and automatic re-routing. Finally, the CPRI bandwidth is dependent on the number of antennas. As the number of antennas increases, the CPRI data rate increases in proportion. This could become a major hindrance to CPRI's applicability in 5G as far as multiple antenna technologies are concerned [73].

Based on the analysis above, the CPRI interface needs to redefine, leading to a new fronthaul interface called Next-Generation Fronthaul Interface (NGFI). The fundamental way to realize NGFI is to redesign the function split between the BBU and the RRU, which makes it different from the traditional fronthaul and backhaul. Given that for different scenarios and applications, different function split schemes exist, which leads to different kinds of NGFI realization.

5.4.1   Different split options

The shortcomings of the CPRI mentioned above are mainly caused by the current BBU-RRU function split. For the current BBU-RRU function split, the baseband-related functions are processed by the BBU while the RRU processes radio frequency related functions. Therefore, the NGFI design should start with a paradigm shift by rethinking and redesigning the functions split between BBU and RRU. Moreover, the function split between BBU and RRU may be different according to the bandwidth and latency of fronthaul, which could be adaptive to different scenarios. For example, if the low bandwidth and high latency fronthaul are provided, more functions should be moved from BBU to RRU. Accordingly, fewer functions should be move from BBU to RRU for the high bandwidth and low latency fronthaul.

- **Option 8: RF/PHY** 3GPPs split option 8 is what has already been introduced as the traditional RRU-BBU split. This split has been known for several years and the literature in this area is very comprehensive. Therefore, several directions within using this split are investigated focusing on the CPRI transport interface: both the traditional CPRI transport, the option of transporting CPRI over the Ethernet network and the option of compressing the CPRI signal are considered.

- **Option 7: Low PHY** In this functional split, the Fast Fourier Transformation (FFT) is included locally in the DU. Due to the Fourier transform, the data to be transmitted over the fronthaul interface is represented by subcarriers. By removing the cyclic prefix and transforming the received signal to the frequency domain using the FFT, guard subcarriers can be removed in the DU. In this split, the fronthaul bitrate is lowered compared to option 8, but it is still constant as the resource element mapping is executed in the CU, and

the resource element mapping is necessary to detect unused subcarriers, and thereby achieve a variable bitrate.

- **Option 6: MAC-PHY** This split separates the data link layer from the physical layer. All physical processing is handled locally and the MAC scheduler is centralized. The resulting CU pooling gain is thereby only including the data link layer and network layer functions, which represent approximately (implementation-specific) 20% of the overall baseband processing [74]. This results in no possible energy savings for the physical layer. The payload, to be transmitted over the fronthaul, using this split is transport blocks and this leads to a large reduction in the bandwidth on the fronthaul link. The load on the fronthaul link is dependent on the load at the $S1$ interface.

- **Option 5: Intra MAC** In this split, an overall scheduler is centralized in the CU, and a MAC sublayer is local in each DU to handle time critical processing. From this split and below, the time critical procedures in the HARQ are performed locally in the DU, and also the functions where performance is proportional to latency. In split option 5, the CU-pool is communicating with the DUs through scheduling commands and HARQ reports. The reduced delay requirements on the fronthaul interface ensure that the distance to the CU pool can be longer.

- **Option 4: RLC/MAC** This split receives RLC Protocol Data Units (PDUs) in the Down Link (DL) direction and transmits MAC Service Data Units (SDUs) in the Up Link (UL) direction. The possibility of a virtualized RLC will lead to resource sharing benefits for both storage and processor utilization. The shorter subframe sizes expected in 5G will allow for more frequent decisions by the scheduler, adapting better to traffic demands or channel conditions, however, this results in more frequent notifications to RLC from MAC specifying the size of the next batch of RLC PDUs. This option may be more robust over non-ideal transmission conditions and during mobility because the Automatic Repeat Request (ARQ) is centralized in the CU.

- **Option 3: Intra RLC** In this split, the RLC is separated into high RLC and low RLC. The low RLC is composed of segmentation functions and the high RLC is composed of ARQ and other RLC functions. The processing of PDCP and asynchronous RLC processing takes place at the CU. All other functions remain in the DU including synchronous RLC network functions. This option reduces the fronthaul latency constraints as realtime scheduling is performed locally in the DU. This option may be more robust over non-ideal transmission conditions and during mobility because the ARQ is centralized in the CU.

- **Option 2: RLC/PDCP** In this split, the PDCP and RRC are centralized while the other functions are performed locally in the DU. This split receives PDCP PDUs in the DL direction and transmits RLC SDUs in the UL direction. This split uses an already standardized interface which makes the inter-operation between elements simpler. In this split, the traffic is divided into multiple flows, which can be directed to various access nodes, making the split support multi-connectivity. In this split, all real-time aspects are located in the DU, and this makes the link requirements for this split the most relaxed.

- **Option 1: PDCP/RRC** In this split, the entire functions are located in the DU. This gives the benefit that the user data is close to the transmission point which can be beneficial for caching. This split will not support several features such as those providing inter-cell coordination, therefore this split might not be beneficial for implementations where many cells are connected to a CU pool. A benefit of centralizing the RRC is that many functions are handled locally, but the user will still benefit from faster mobility management and the operator from not needing to manage and maintain the $X2$ interface.

5.5 Advantages and challenges

In future 5G networks, the number of cells will increase to an extreme number. This means that with C-RAN, one CU pool will probably be connected to hundreds or even thousands of DUs. By using the traditional RRU-BBU split for all those DUs, great advantages are obtained giving the largest amount of shared resources and very simple and scalable DUs. On the other hand, by using a lower split, fewer resources can be shared and the DU will be more complex, but the load on the fronthaul network will be lower and vary with the user load. This is a trade-off between localizing and centralizing the BS functions. The latter scenario will also prove more resilient compared to a traditional BS, as there will be more processing power available in the CU pool, and thereby backup options.

The higher numbered splits have the advantages that they support advanced functions such as CoMP and they are more robust to non-ideal transport conditions. At the same time, they have very strict latency requirements and higher bitrates. The lower splits have moved almost all functions locally, close to the user. This results in high utilization of the fronthaul link, but only a few resources shared in the CU pool. In short, the higher split the more resources shared in the CU pool and the lower split the more resources shared on the fronthaul link. But also other things need to be taken into consideration: For example, under certain circumstances, it will be more efficient to have a longer distance between the DU and CU than the 40 km limited by the HARQ process. This could be to cover a rural area or to cover a certain road by one CU pool and benefit from fast handovers. The possibility of having multiple local schedulers as in split 1 to 4 can be beneficial when a lot of processing power is required locally.

Chapter 6

Adaptive Function Chaining for Resilient 5G RAN

6.1 Introduction

One of the motivations of the Xhaul network is its ability to split different functions and executes them in separate entities. Functional splits determine the number of functions which stay locally and the number of baseband functions that are centralized in the relatively well-connected locations in the network. There is a vast number of works already done in the literature which study different functional splits. In order to mitigate the fronthaul requirements imposed by the Xhaul architecture, several functional splits, each characterized by a different demarcation point between the centralized and the distributed units, have emerged. However, the selection of the appropriate centralization level (i.e., the functional split) remains a challenging task, since several parameters have to be considered in order to make such a decision.

This chapter proposes a novel function location algorithm, which adopts dynamic function chaining in relation to the evolution of the traffic estimate. The obtained results show remarkable improvement in terms of bandwidth saving and multiplexing gain with respect to conventional C-RAN fronthaul and suggest design criteria for the emerging 5G access network. This methodology can also guarantee service continuity in the case of single CU/DU or link failure, namely based on dedicated and shared path protection principles. The proposed techniques not only provide full protection against single failure but also exhibit significant savings in terms of network resources, by suitably sharing redundant backup resources.

6.2 Adaptive Algorithm for Dynamic Variation of User Traffic

In relation to the adoption of different functional split options, in this chapter, the classic residential-industrial traffic over 24 hours has been considered but the approach presented here can be adapted to any other variation of the traffic. The possibility of dynamically assigning different functions to different entities and nodes in the access network is studied according to the traffic profile. As can be seen in figure 6.1 this approach assumes that function splits are not statically assigned
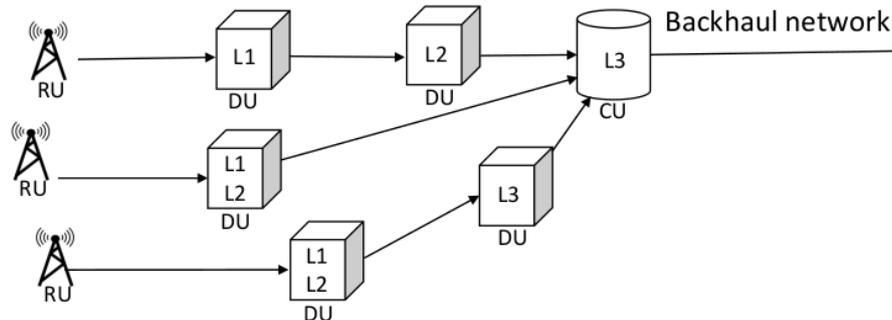
Figure 6.1: Sample Xhaul function chain configurations considered in the algorithm.

but instead, depending on the traffic demand and availability of the Processing Units (PUs), dynamic chaining of the function is configured based on the Xhaul to efficiently allocate network resources.

6.2.1 Function chain requirements

In order to make this function chaining feasible there are a few constraints that need to be taken into consideration:

- **Latency**: Among all the options for splitting, PHY and MAC layers are the most delay-sensitive. The main reason is due to the Hybrid Automatic Repeat Request (HARQ) which is controlled by lower layer MAC and executed in the PHY layer. Splitting the PHY and MAC layers lead to stricter requirements over latency. In this chapter, we evaluated the latency parameters as a function of distance in terms of hops.

- **Bandwidth**: As we mentioned several times one of the major benefits of functional splits in the Xhaul network, is the bandwidth usage reduction. Since NGFI is traffic dependent, in case of low traffic extra bandwidth can be used for other purposes. By implementing the functional chain, there is the possibility of executing the bandwidth hungry functions in the local or the closest DU. As a result, the outgoing low bandwidth signal can be routed throughout the network to be executed on another DU or in the CU.

- **Processing units (PU)**: in the previous generation of the access network, all functionalities were executed in either data center or BBU hotels with a high amount of resources. Xhaul, on the other hand, is introducing the possibility to perform some processing in DUs with a limited amount of processing resources (namely the number of PUs). In CUs, instead of unlimited processing resources are still considered, as in previous configurations. As a consequence, the proper dimensioning of the PUs in DUs is an important aspect of optimization.

6.2.2 Function Chain Algorithm

The heuristic algorithm presented here aims at locating baseband functionalities in the access network as a reconfigurable function chain, by efficiently adapting to traffic generated by active antennas, in relation to distance (hop) and PU constraints. The problem to solve is formally defined as follows:

Table 6.1: Notation used in the algorithm 8.

N	Set of nodes in the network $|N| = n$
D_i	Nodal degree of node $i \in N$
S_i	Set of active DUs under the hop and processing units constraint for node $i \in N$
S_n	Set of active nodes in the network.
DU_i	DU in the location of node $i \in N$
P_i	Total processing units of the DU in node $i \in N$
P_1	Number of PUs for $L1$
P_2	Number of PUs for $L2$
P_3	Number of PUs for $L3$
A_i	Number of active antennas at node $i \in N$
$Path_{ij}$	The shortest path between nodes i and $j \in N$
B_i	Total available bandwidth in link i
B_1	Required bandwidth for $L1$
B_2	Required bandwidth for $L2$
B_3	Required bandwidth for $L3$

- **Given** the physical network with interconnected nodes supporting antennas, the number and placement of CUs, the number of PUs in DUs and the daily traffic profile.

- **Find** the minimum number of active DUs according to delay (hops) and PU constraints in order to adapt to the daily traffic profile while dynamically reconfiguring the X-haul function chain.

The algorithm 8 is executed sequentially in all the nodes of the network. Each node can execute the baseband functionalities but they are all assumed to be deactivated before starting the algorithm. The algorithm stops after the last node in the network execute the algorithm (the condition in line 2). It is also assumed that the dimensioning of the PUs has been precomputed and all DUs have a certain amount of available PUs. All the notations used in the algorithm can be found in table 6.1

The algorithm starts in line 3 in the node $i \in N$ with the highest nodal degree. The effect of the starting point in the assignment algorithm has been already studied [75]. Depending on different constraints (maximum distance and available PUs) the set S_i is created in line 4. This set is composed of all the possible DU candidates under the requirement constraints. If node i is the first node that executes the algorithm or, the constraints are so tight that there is no possible DU candidate, then the set S_i turns out empty. Lines 5 to 7 to investigate this situation. If node i cannot find any DU, then it activates the DU in its location and the active DU_i will be added to the set S_n in line 6. This set contains all active DUs in the network. In line 7 node i uses the available PUs in DU_i. Since DU_i is just opened, it has enough PUs to executes all the layers (line 7). On the other hand, if there are some possible DU candidates exist, a decision has to be made regarding the assignment (line 8).

The decision making logic is based on finding the DU with the highest available PUs to execute all the layers and prevent routing and assigning the bandwidth throughput of the network. In line 9 each DU in the set S_i namely DU in node j is

checked for the availability of the PUs. If DU in node j has enough PUs that can execute all the layers (line 10) then node i will be connected to the DU in node j and related PUs will be assigned to it (line 11). For assigning the bandwidth, the algorithm finds the shortest path between nodes i and j which has been precomputed and allocates the required bandwidth to all the links associated with the $Path_{ij}$ (lines 12 to 14). Lines 15 to 31 consider the situation when the chosen DU has only enough PUs to executes layers 1 and 2 (line 15). In that case, node i will be connected to DU in node j and uses the available PUs for layers 1 and 2 (line 16). Upon the connection to DU in the node j, all the links in the $Path_{ij}$ also get the required bandwidth (lines 17 to 19).

For the execution of layer 3, the algorithm first searches for all the possible DUs namely $z \in S_i$ under the required constraints (line 20). If such DU exists then node i uses its PUs for executing layer 3 functions (line 21). The required bandwidth also will be assigned to all the links in the shortest path between nodes j and z (lines 22 to 24). Otherwise, the shortest path towards all predefined CUs will be computed and the closest one will be identified (line 26) so that the rest of the functions will be routed and executed in that CU (lines 27-30).

In line 32, the last possible scenario will be tested. If the available DU only has enough PUs for the execution of layer 1, then node i will be connected to DU in node j and executes layer 1 functions (line 33). The bandwidth in the shortest path between nodes i and j also will be updated in lines 34 to 36. For the rest of the functions again the algorithm looks for all the possible DUs in the set S_i (line 37). If such DU exists, in line 38, the assignment for layers 2 and 3 is presented and the related bandwidth will be updated accordingly (lines 39 to 41).

Otherwise, the algorithm connects node j to the closest CU in line 43 (based on the shortest path) and uses the available PUs for executing layer 2 and 3 (line 44). The related bandwidth will be updated accordingly in lines 45 to 47. In line 51, node i will be removed from the set N and the algorithm passes the control to the next highest nodal degree node in the network.

The worst-case complexity of the algorithm is $\mathcal{O}(N^3)$. It is calculated by considering the maximum number of iterations for all the loops in the algorithm.

6.3   Resilient schemes

To obtain a favorable combination of a node and link resilience yielding high connection availability and resource usage efficiency, we propose two algorithms namely Dedicated Path Protection (DPP) and Shared Path Protection (SPP) in the network. The DPP enforces provisioning the node and links disjoint paths for every request in the network. The established paths were dedicated network resources namely bandwidth (BW) and processing units (PUs) for different baseband processing execution. This results in keeping a significant amount of resources ideal for the case of failure. Therefore, the second algorithm, SPP introduced to reduce network resource usage. In particular, this method shares all the possible reserved backup resources under some constraints. So as a result, survivability still intact but with less overall network cost and more overall connection availability.

The heuristic algorithms presented in this section aim at allocating the baseband functionalities in the access network as a function chain, depending on different requests generated by active antennas, in relation to the constraints such as latency,

---

Algorithm 8 Function Chain.

1: **Initialization:** $S_i, S_n \leftarrow \emptyset$
2: **while** $(N \neq \emptyset)$ **do**
3:     **find** node $i \in N$ s.t. D_i is maximum
4:     **create** S_i
5:     **if** $S_i \leftarrow \emptyset$
6:        $S_n \leftarrow DU_i$
7:        $P_i = P_i - [P_1 + P_2 + P_3] * A_i$
8:     **else**
9:        **for** each node $j \in S_i$
10:           **if** $P_j \geq [P_1 + P_2 + P_3] * A_i$
11:             $P_j = P_j - [P_1 + P_2 + P_3] * A_i$
12:             **for** each $l \in Path_{ij}$
13:                $B_l = B_l + [B_1 + B_2 + B_3] * A_i$
14:             **end for**
15:           **else if** $P_j \geq [P_1 + P_2] * A_i$
16:             $P_j = P_j - [P_1 + P_2] * A_i$
17:             **for** each $l \in Path_{ij}$
18:                $B_l = B_l + [B_1 + B_2] * A_i$
19:             **end for**
20:             **if exists** node $z \in S_i$ s.t. $z \neq j$ and $P_z \geq [P_3] * Ai$
21:                $P_z = P_z - [P_3] * A_i$
22:                **for** each $l \in Path_{jz}$
23:                   $B_l = B_l + [B_3] * A_i$
24:                **end for**
25:             **else**
26:                **find** closet CU
27:                $P_{CU} = P_{CU} - [P_3] * A_i$
28:                **for** each $l \in Path_{jCU}$
29:                   $B_l = B_l + [B_3] * A_i$
30:                **end for**
31:             **end if**
32:           **else if** $P_j \geq [P_1] * A_i$
33:             $P_j = P_j - [P_1] * A_i$
34:             **for** each $l \in Path_{ij}$
35:                $B_l = B_l + [B_1] * A_i$
36:             **end for**
37:             **if exists** node $z \in S_i$ s.t. $z \neq j$ and $P_z \geq [P_2 + P_3] * Ai$
38:                $P_z = P_z - [P_2 + P_3] * A_i$
39:                **for** each $l \in Path_{jz}$
40:                   $B_l = B_l + [B_2 + B_3] * A_i$
41:                **end for**
42:             **else**
43:             **find** closet CU
44:             $P_{CU} = P_{CU} - [P_2 + P_3] * A_i$
45:             **for** each $l \in Path_{jCU}$
46:                $B_l = B_l + [B_2 + B_3] * A_i$
47:             **end for**
48:           **end if**
49:        **end for**
50:     **end if**
51:     **remove** node i from N
52: **end while**

Table 6.2: List of parameters used in algorithms 9 and 10.

N	Set of nodes in the network $|N| = n$
D_i	Nodal degree of node $i \in N$
S_i	Set of nodes on the maximum distance from node $i \in N$
MD	Maximum allowed distance
T_i	Computational capacity required by node $i \in N$. $T_i = (\sum_{x=1}^{5} L_x) \times RU_i$
S_A	Set of active nodes in the network.
P_{ij}	1 if node j is the primary DU for the RUs connected to node i, 0 otherwise.
B_{ij}	1 if node j is the backup DU for the RUs connected to node i, 0 otherwise.
$Path_{i,j}$	The shortest path between nodes i and $j \in N$
BW_x	Required bandwidth for split layer x

bandwidth, and processing unit limitations. The solution must be resilient against a single node or link failure. The problem to solve is formally defined as follows:

- **Given** the physical network with interconnected nodes supporting antennas, and maximum possible traffic in each node.

- **Find** the minimum number of active DU/CUs according to delay (hops), bandwidth (BW) and PU constraints.

- **To ensure** the full coverage of the network as well as guaranteeing the survivability against a single node or link failure.

While heuristic strategies for C-RAN survivable deployment have been presented in the past for both DPP and SPP [76], in this section first a heuristic approach for resilient function splitting is presenting. In the following subsection, the cost-saving methodologies in terms of network resources implemented to reduce the total network costs. All the notations used in the two strategies are reported in table 6.2.

6.3.1 Dedicated Path Protection (DPP)

The function chaining algorithm with the DPP protection technique is presented in algorithm 9. The algorithm is executed sequentially in all the nodes of the network. Each node can execute the baseband functionalities but they are all assumed to be deactivated before starting the algorithm ($S_A \leftarrow \emptyset$). The algorithm terminates after the last node in the network executes the algorithm (the condition in line 2). The algorithm starts at line 3 in the node $i \in N$ with the highest nodal degree which expects to achieve better results in terms of node activation. Depending on the maximum allowed distance constraint the set S_i is created in line 4. For the primary and backup support, two nodes j_1 and j_2 from the set S_i are chosen (line 5). If nodes j_1 and j_2 have enough amount of PUs to completely process all the traffic from node i, both nodes j_1 and j_2 will be activated (line 6) one for the primary and the other for the backup purpose (line 7). The corresponding functionalities and the amount of PUs in both nodes are updated accordingly in lines 8 and 9. Moreover,

the bandwidth on all links on the shortest paths is also correspondingly updated (lines 10 and 11).

If the number of PUs is not enough for executing the whole functionality, the process of forming the function chain will be started by activating nodes j_1 and j_2 (line 13) one for the primary and the other for the backup purpose (line 14). The remaining required PUs after executions in nodes j_1 and j_2 is updated (lines 16 and 17). All the links in the shortest path between nodes i and j are also updated. The chaining functionalities continue as long as it does not reach the distance limitation or all the required functionalities executed properly (the condition starting at line 18). For each node j_1 and j_2 the sets of S_{j1} and S_{j2} are created using new constraints in terms of distance allowance (line 19). In line 21, two nodes namely z_1 and z_2 will be picked such that have enough PUs for the rest of baseband execution. They will be activated (line 22) one for the primary and one for the backup purpose (line 23). Both the PUs and the bandwidth on the shortest path between the pair of j_1, j_2 and nodes z_1 and z_2 will be updated accordingly (lines 24-27). The chaining process continues by updating the constraint on maximum allowed distance in line 28 and swapping the nodes in lines 29 and 30. After all the executions, node i is removed from the set and control of the algorithm moves to the next node.

6.3.2 Shared Path Protection (SPP)

This technique aims to share the backup network resource efficiently. As stated before, in this study we consider a single node (CU/DU) or a single link failure at a time. For sharing the backup resources weather it is PU or bandwidth (BW) in a backup path, the following rule should be applied: A backup PU or bandwidth can be shared among some RUs if and only if, those RUs have dedicated and different PUs and bandwidth in their primary paths. When a failure happens in any part of the primary path, the RUs using the resources in that path, shifting to their backup one. More details on this technique are presented in algorithm 10. This procedure checks all active nodes to evaluate the possibility of sharing. In line 3, the algorithm checks all the nodes j_1 and j_2 which using PUs in active node i to see if they can satisfy the condition for sharing backup PUs. In line 4, if exists a node $z \in S_A$ such that it is not part of the primary path for nodes j_1 and j_2 then in line 5 the PUs can be shared.

With the same line of reasoning, the sharing for backup bandwidth is presented in the second part of algorithm 10. This procedure evaluates all links namely link $l \in L$ in line 7. If exist two or more nodes such as j_1 and j_2 which using the bandwidth in l for backup purpose, then only in the case of not having an identical primary path they can share the bandwidth in link l.

6.4 Results

To show the effectiveness of the algorithm, a set of results is here presented organized into two parts. Firstly, the results related to the dynamic function chaining is presented without any survivable applied technique. In the second part, the results obtained with the two protection techniques are also shown.

Algorithm 9 Function Chain with DPP.

1: **Initialization:** S_i and $S_A \leftarrow \emptyset$
2: **while** $(n \neq \emptyset)$
3:     **find** node $i \in N$ s.t. D_i is maximum
4:     **create** S_i
5:    **if** exist nodes j_1 and $j_2 \in N$ s.t. PU_{j1} and $PU_{j2} \geq T_i$
6:       $S_A \leftarrow j_1$ and j_2
7:       $P_{i,j_1} = 1$ and $B_{i,j_2} = 1$
8:       $PU_{j1} = PU_{j1} - T_i$
9:       $PU_{j2} = PU_{j2} - T_i$
10:       **for** all links $l \in Path_{i,j1}$ and $Path_{i,j2}$
11:         **update** BW_l
12:       **end for**
13:     **end if**
14:     **else**
15:       $S_A \leftarrow j_1$ and j_2
16:       $P_{i,j_1} = 1$ and $B_{i,j_2} = 1$
17:       $T_{i1} = T_{i1} - PU_{j1}$
18:       $T_{i2} = T_{i1} - PU_{j2}$
19:       **for** all links $l \in Path_{i,j1}$ and $Path_{i,j2}$
20:         **update** BW_l
21:       **end for**
22:       **while** $MD \neq 0)$ or $(T_{i1}, T_{i2} \neq 0$
23:         **create** S_{j1} and S_{j2}
24:        **if** exist nodes $z_1 \in S_{j1}$ and $z_2 \in S_{j2}$ s.t. $PU_{z1} \geq T_{i1}$ and $PU_{z2} \geq T_{i2}$
25:           $S_A \leftarrow z_1$ and z_2
26:           $P_{j,z_1} = 1$ and $B_{j,z_2} = 1$
27:         $PU_{z1} = PU_{z1} - T_{i1}$
28:          $PU_{z2} = PU_{z2} - T_{i2}$
29:          **for** all links $l \in Path_{j1,z1}$ and $Path_{j2,z2}$
30:           **update** BW_l
31:          **end for**
32:     $MD = MD - 1$
33:     $j_1 \leftarrow z_1$
34:     $j_2 \leftarrow z_2$
35:     **end while**

Algorithm 10 SPP Techniques.

 1: PU sharing:
 2: **for** each node $i \in S_A$
 3:    **for** all nodes j_1 and j_2 s.t. $B_{ij1} = 1$ and $B_{ij2} = 1$
 4:       **if** exists node $z \in S_A$ s.t. $P_{zj1} \neq P_{zj2}$
 5:          **share** PU_i
 6:       **end if**
 7:       **end for**
 8:    **end for**
 9: **end for**
10: BW sharing:
11: **for** each link $l \in L$
12:    **if** exist nodes j_1 and j_2 s.t. $l \in B_{path,j1}$ and $B_{path,j2}$
13:       **if** exists node $z \in S_A$ s.t. $P_{zj1} \neq P_{zj2}$
14:          **share** BW_l
15:       **end if**
16:       **end if**
17:    **end for**
18: **end for**

6.4.1 Numerical Results for Dynamic Function Changing

Figure 6.2 shows the reference network for all the evaluations, consisting of 38 nodes and 48 high capacity transport links. Three CUs are considered to serve the network where data centers are located. The figure also is shown two simple examples of the decision logic of the algorithm. The connection between RUs and CUs for service purposes can happen through a chain of intermediate DUs. All the nodes in the network including the one hosting CUs can produce traffic and need to be assigned to proper entities for processing.

The main motivation of the algorithm 8 is to adopt the proper amount of network resources to the traffic pattern evolution, assumed in figure 6.3 as the number of active antennas per node over 24 hours of the day. The pattern presents low traffic in the early hours of the day, a peak in the middle and then decreases while reaching the end of the day. In this chapter, we assumed the same amount of traffic at each hour for all the nodes in the network. This value is the highest amount of traffic predicted on each specific hour of a day.

The total number of the PUs is calculated based on the average of this pattern, in relation to the requirements of each functional layer and then averaged on the number of nodes in the network. The algorithm 8 aims to find suitable chaining of the functions throughout the network while using the available resources and avoiding any blocking of requests.

Figure 6.4 shows the comparison of C-RAN and Xhaul in terms of activated nodes, namely BBU hotels, DUs and CUs respectively, varying the distance constraints. It is assumed the latency constraint is not violated for both architectures due to the diameter of 20 km of the use case in terms of hops. For this comparison, the Xhaul network does not have any limitation over the available PUs in DUs which is the same situation in the C-RAN architecture. As a consequence, the variation of the traffic during the day does not affect the number of active DUs in the network.
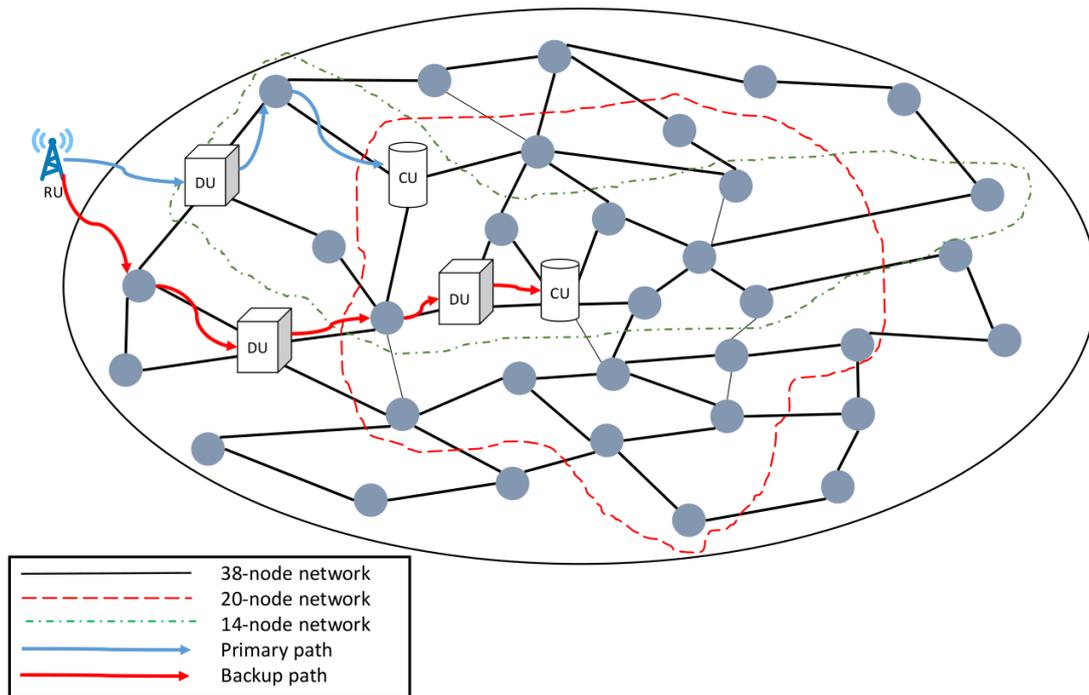

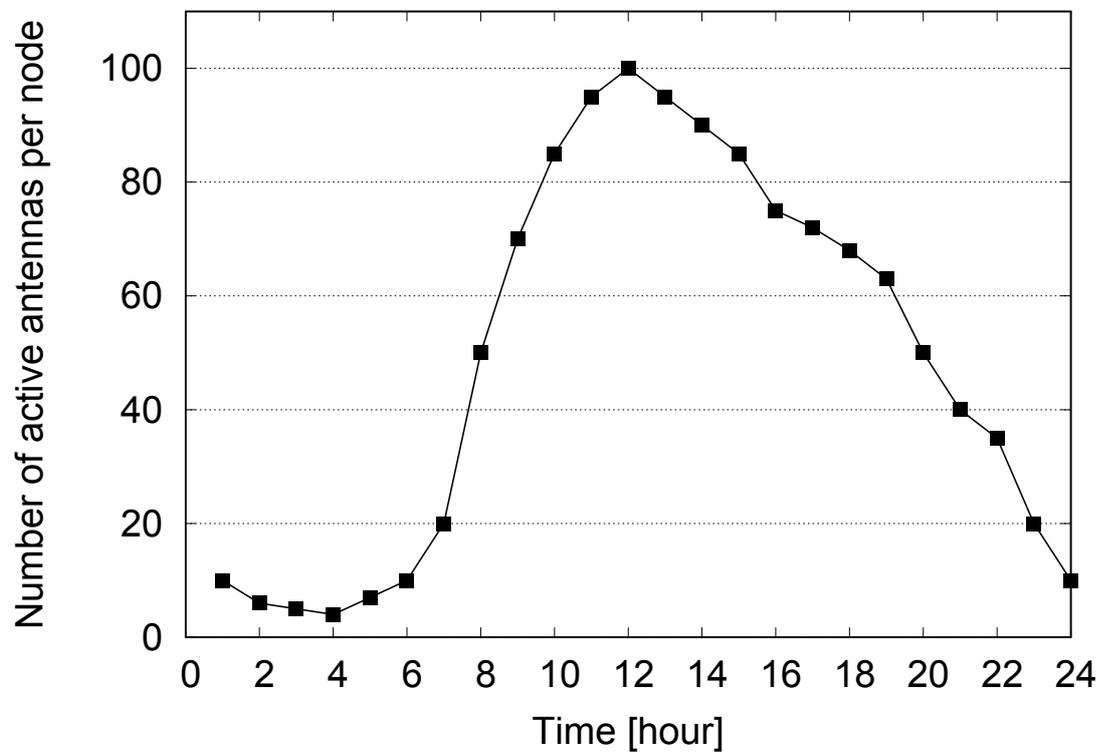Figure 6.2: Reference access network for evaluations.

Figure 6.3: Evolution of the number of active antennas per node during the 24 hours.
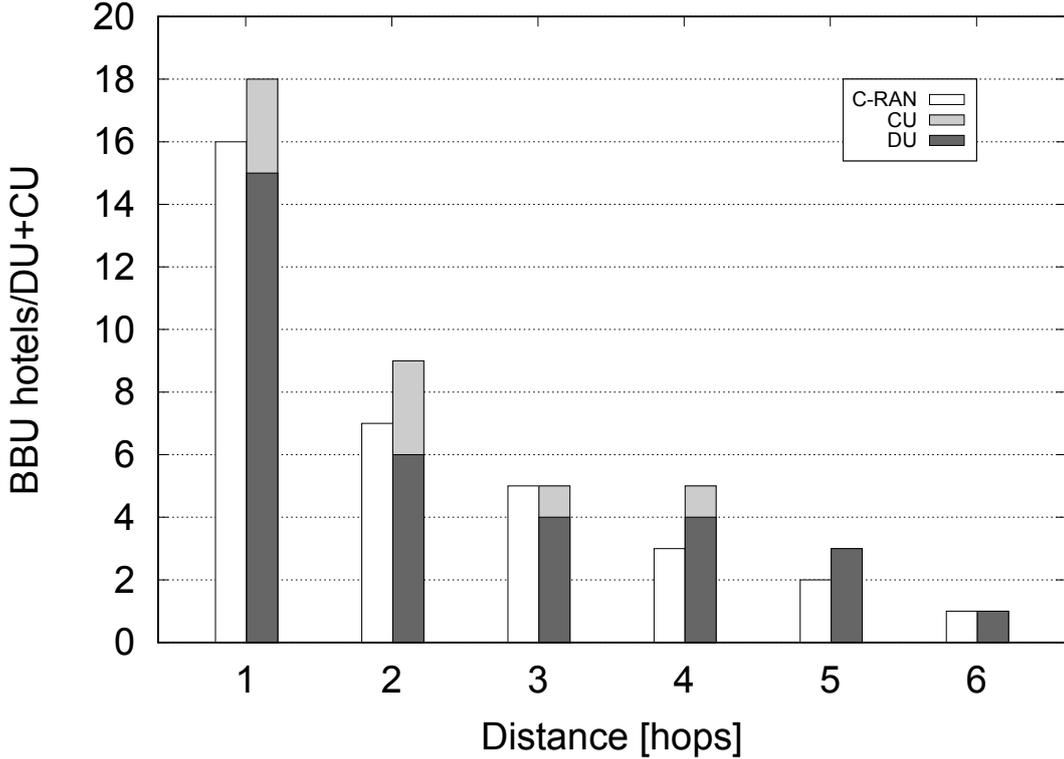
Figure 6.4: Comparison of the total number of active nodes (BBU hotels, DUs and CUs) as a function of the distance constraint for C-RAN and Xhaul with no limitation on the PU.

In all distance constraints, the two approaches achieve close results. In C-RAN the number of active BBU hotels decreases as the constraint on distant relaxes. This is also true for the Xhaul except in the cases that algorithm due to the physical network topology cannot find a better solution even by relaxing the distance (hops 3 and 4). In the Xhaul, the trend also shows the contribution of DUs and CUs. When distance constraint is strict (1 hop) the algorithm relies also on CUs for the execution of functions. In the very relax distance constraints (5 and 6 hops) the dependencies on CUs are eliminated due to the fact that the algorithm prioritizes using DUs over CUs. In the 6 hops constraint, C-RAN and Xhaul have the same requirement of activating only 1 node that corresponds to full centralization.

Figure 6.5 reports the evaluation of the required bandwidth in the same conditions of figure 6.4, showing the real advantage of the Xhaul architecture. This figure compares the total assigned bandwidth in the network as a function of the distance constraint, again for C-RAN and Xhaul, with no limitation on the PUs. By relaxing the distance constraint, the total bandwidth usage increases in all scenarios, which represents the well-known cost of centralization. Even though the PUs are assumed to be infinite, the variation of the traffic during the day sensibly affects the bandwidth in Xhaul. Instead, being C-RAN at a constant bit rate, the variation of the traffic does not affect the assigned bandwidth. In particular, Xhaul adapts to the traffic variations and in both the low (6 a.m.) and the peak (12 p.m.) traffic situations there are effective bandwidth savings in adopting the Xhaul approach.

This is also shown in figure 6.6 by plotting the multiplexing gain in terms of
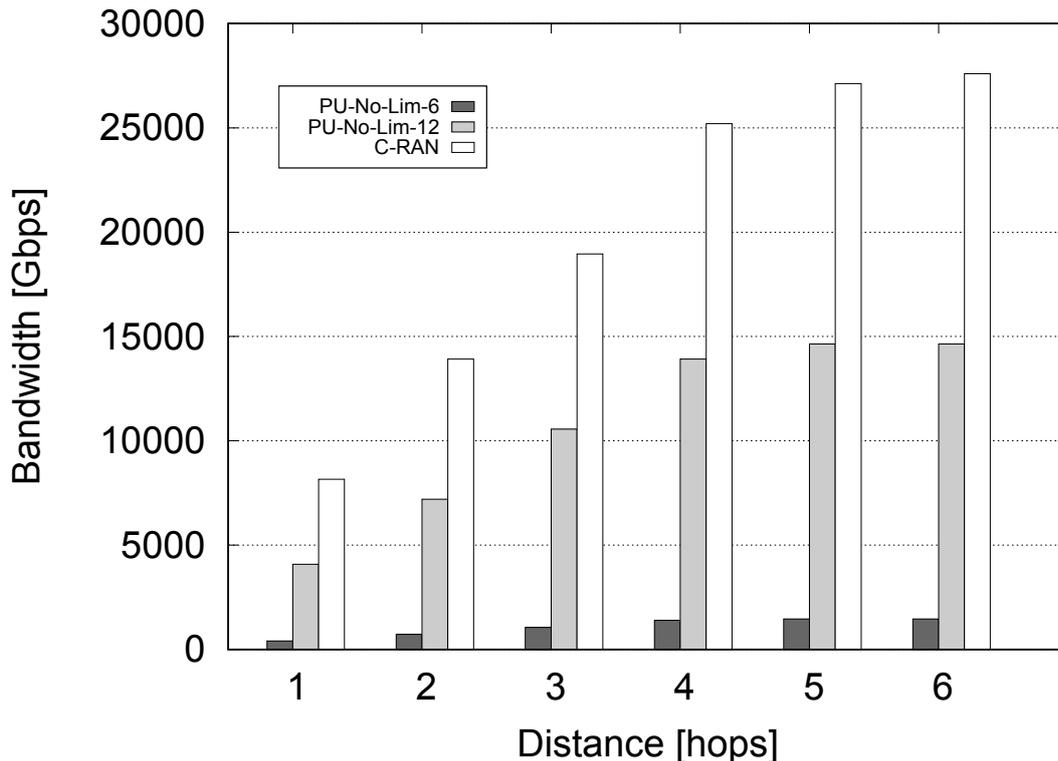
Figure 6.5: Comparison of the total bandwidth as a function of the distance constraint for C-RAN and Xhaul with no limitation on the PU in two different traffic situations (i.e. at 6 a.m. and 12 p.m. from figure 6.4).

bandwidth indicated by G in equation 1. It is defined as the difference between the total amount of used bandwidth in C-RAN (BW_c) and Xhaul (BW_x) scenarios divided by the value for the C-RAN. This value shows the statistical saving in the usage of bandwidth in Xhaul compares to the C-RAN. The multiplexing gain results almost independent of the distance constraint and much higher for Xhaul then for C-RAN. This means that with Xhaul the access network can allocate more services with respect to C-RAN, given a set of transport resources.

$$G = \frac{BW_c - BW_x}{BW_c} \qquad (6.1)$$

Figures 6.7 and 6.8 show the results for the Xhaul architecture when both limitations over distance and available PUs have been applied, in the low (6 a.m.) and the peak (12 p.m.) traffic situations, respectively. The figures are showing a comparison of the total number of active DUs and bandwidth per link as a function of different distance constraints. The dimensioning of the PUs in the nodes is based on the average traffic and the processing required for each layer. As the distance constraint is relaxed, the constraint on PUs leads to a higher number of active nodes with respect to the ideal case. As far as the bandwidth, even with the same number of active DUs, having a longer path means also more bandwidth needed. These figures both suggest designing the network to have a distance around 2 or 3 hops so that both aspects, number of nodes and bandwidth, can be optimized.
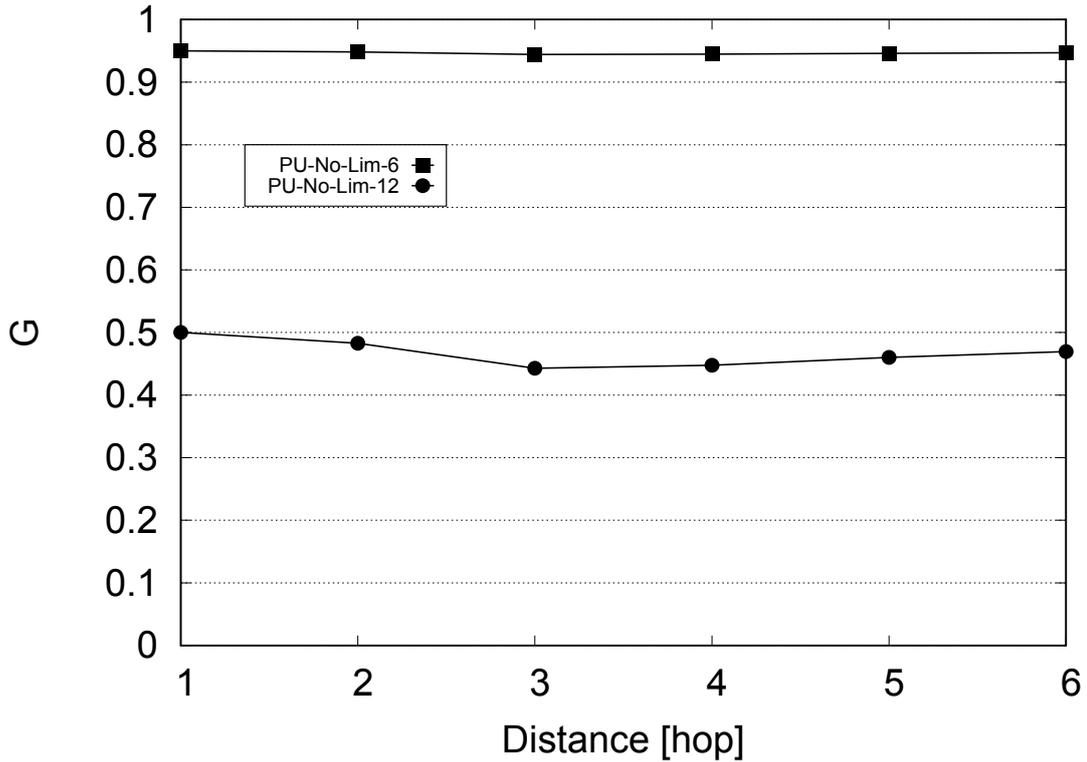
Figure 6.6: Comparison of the Xhaul multiplexing gain with respect to the C-RAN with no limitation on the processing units in low (6 a.m.) and peak (12 p.m.) traffic hours.

6.4.2 Numerical Results for Survivability Techniques

In order to show the effectiveness of the presented survivable algorithms, a set of results is here presented organized into two parts. The first part is dedicated to the comparison of Xhaul and C-RAN architecture. The results presented here have no limitation over the amount of PUs. This is due to the fact that in C-RAN architecture the assumption is that all baseband functionalities are executed in large BBU hotels with no limitation of required resources. Therefore, in order to have a fair comparison, there is no limitation over PU applied to Xhaul as well. n the second part, the effect of the limitation on the maximum allowed processing units (PU) and on the available bandwidth on each link (BW) is outlined.

Figure 6.9 presents the comparison of the total number of active nodes, namely BBU hotels and CU/DUs as a function of the distance constraint expressed in the number of hops, with reference to the two scenarios, namely C-RAN and Xhaul respectively. In this set of results, DPP has been applied to both architectures. In addition, the restriction over the maximum allowed bandwidth is set to 100 Gbps in all the transport links in both cases. As can be seen from the figure, both scenarios activate fewer nodes when the constraint over distance relaxes. This is due to the fact that both algorithms are not heavily bounded by the constraint over distance and can achieve more centralization. Despite this, Xhaul activates more nodes in comparison with C-RAN when the distance constraint is tight, namely, the number of hops is equal to 1 and 2. This happens because of the chaining mechanism which requires more nodes to sequentially execute baseband functionalities which in
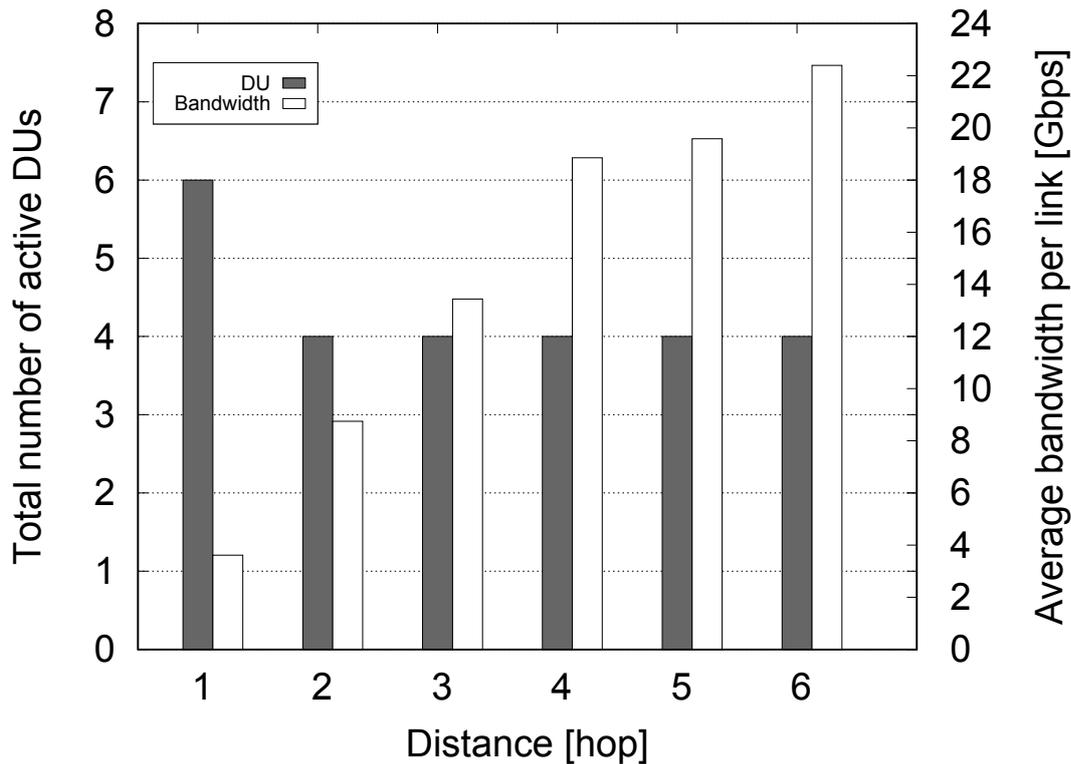
Figure 6.7: Comparison of the average bandwidth per link and total active DUs as a function of the distance constraints for Xhaul with the limitation on both processing units and hops in the low traffic (6 a.m.).

C-RAN case are executed in a single place.

The results presented in figure 6.10 allow the numerical comparison of the total used bandwidth with DPP in Xhaul and C-RAN over different distance constraints expressed in hops. As in the previous figure, no restriction over the processing units has been applied. The limitation over bandwidth is again fixed to 100 Gbps for each link.

The results here are complementary to the ones presented in figure 6.9 due to the fact that having less active nodes means more bandwidth for routing. More specifically, when the restrictions over the distance are tight, namely the hop limit is equal to 1 and 3, the Xhaul architecture has more active nodes due to the chaining in comparison to C-RAN. As a consequence, the total used bandwidth is lower than in C-RAN due to functional splitting. When the distance constraint relaxes, the real benefit of the Xhaul approach shows in the sense that the used bandwidth is significantly lower than C-RAN. The reason is that in C-RAN all the traffic is routed from RRU to BBU hotel, while in Xhaul the bandwidth-hungry functions are executed by the DU close to the users thus making the traffics routed to either another DU or CU significantly lower. At distances equal to 5 and 6 hops, both scenarios have an equal amount of active nodes but due to the execution of all layer traffic in C-RAN, a large gap in bandwidth usage is shown.

The main reason to implies the SPP to the network is to reduce the network cost in terms of bandwidth usage as well as keeping the whole architecture resilient to the possible failures. In this figure (Figure 6.11) the SPP technique is applied and
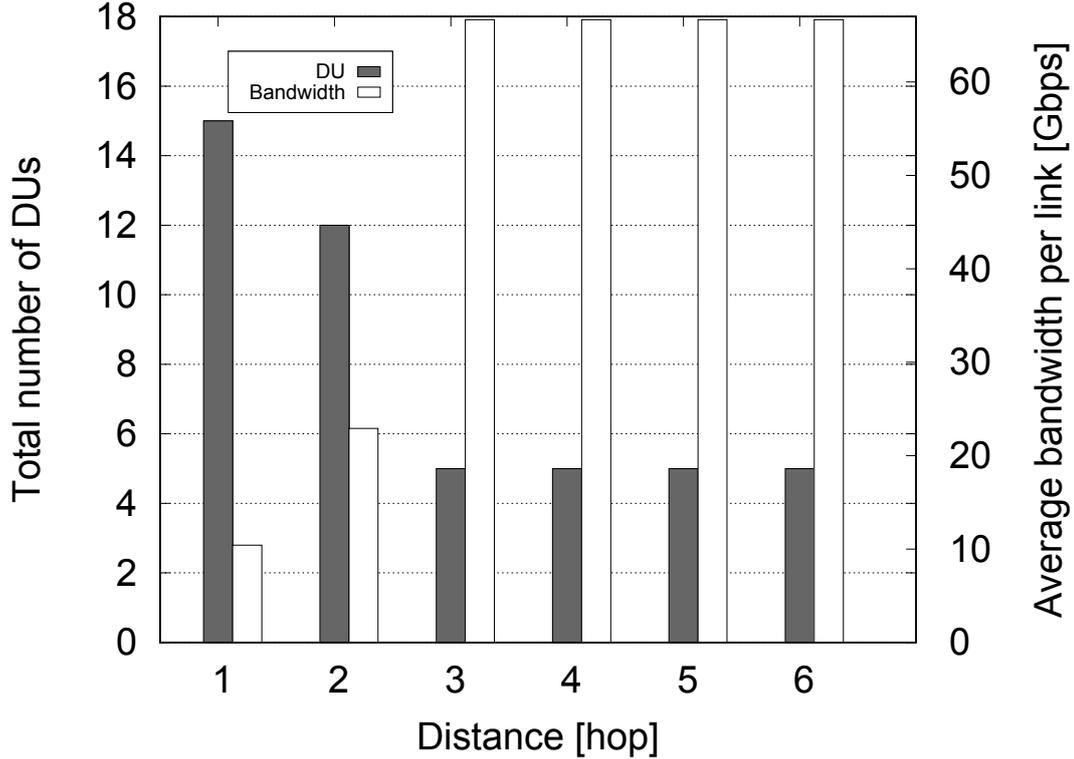
Figure 6.8: Comparison of the average bandwidth per link and total active DUs as a function of the distance constraints for Xhaul with the limitation on both processing units and hops in the high traffic (12 p.m.).

the related results presented in both scenarios. By comparing the results obtaining here and in the DPP case (figure 6.10), it is evident that SPP efficiently reducing the bandwidth usage for both architectures. This approach has more effect on Xhaul than C-RAN due to higher active nodes.

In Xhaul in particular, when the distance constraint is fixed to 1 hop, the SPP does not affect significantly due to really strict constraints. The more the distance constraint relaxes, SPP shows more effectiveness by sharing more bandwidth. At the higher distance constraints, 5 and 6 hops, since few nodes are active in the network algorithm can not be effective and result in the same value as DPP.

Table 6.3 reports the overall saving in terms of bandwidth when the protection technique changes from DPP to SPP for both Xhaul and C-RAN. As can be seen in Xhaul by relaxing the distance constraint the percentage of bandwidth sharing increases thanks to functional splitting which lets the baseband functionalities execute in different entities. As indicated before, for the high number of hops, the SPP does not have any effect because all the functions are centralized in a few nodes. In C-RAN on another hand, by increasing the distance the possibility of sharing bandwidth decreases due to the centralization.

This is also shown in figure 6.12 by plotting the multiplexing gain in terms of bandwidth indicated by G in equation 6.1. It is defined as the difference between the total amount of used bandwidth in C-RAN (BW_c) and Xhaul (BW_x) scenarios divided by the value for the C-RAN. This value shows the statistical saving in the usage of bandwidth in Xhaul compared to the C-RAN for DPP and SPP.
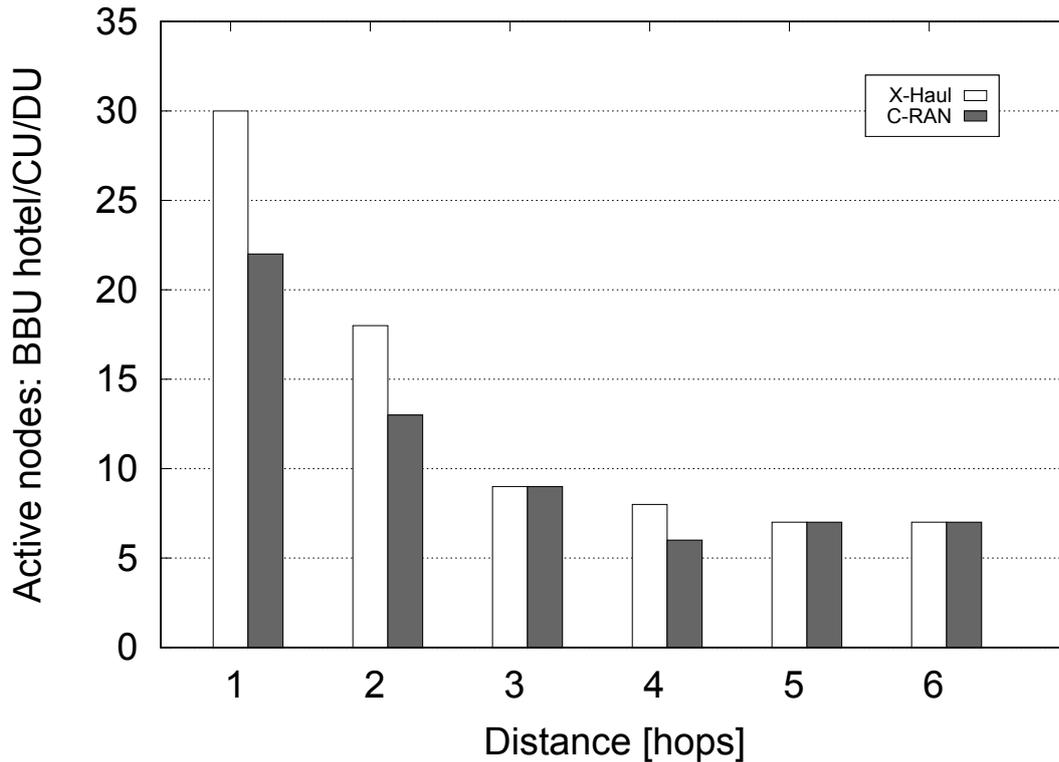
Figure 6.9: Comparison of the total number of active nodes in terms of BBU hotels or CU/DUs as a function of the distance in hops, for Xhaul and C-RAN with DPP protection.

A comparison of the effectiveness of different protection schemes in different situations is performed. In particular, when the distance constraint is small, namely 1 and 2 hops, it is advisable to use DPP due to low latency requirements and fast recovery. When the value for hops increases, 3 and 4 hops, the SPP is more efficient and can be applied in use cases in the area of enhanced Mobile Broadband (eMBB). When the distance is not strictly bounded such as hop equal to 5 and 6 then it is better to take advantage of the centralization benefits of C-RAN.

In the second part of the results, the effectiveness of the variation of processing units and maximum allowed bandwidth is investigated.

Figure 6.13 shows the impact of the available processing units in each node. The figure shows the total active nodes as a function of the amount of PUs for

Table 6.3: Comparison of the bandwidth saving in percentage from DPP to SPP for Xhaul and C-RAN in different distance constraints.

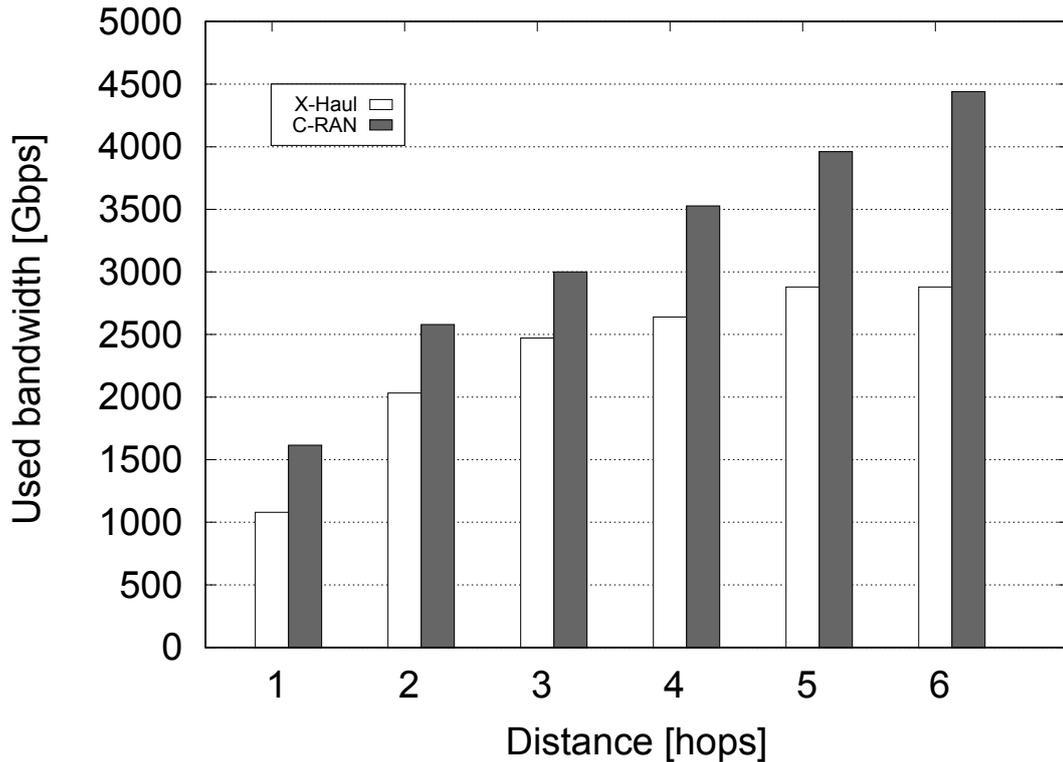| Distance [hops]	Xhaul	C-RAN
1	11%	33%
2	30%	31%
3	33%	15 %
4	21%	15%
5	0%	16%
6	0%	17%

Figure 6.10: Comparison of the total used bandwidth as a function of the distance in hops for Xhaul and C-RAN with DPP protection.

three different network sizes, 38, 20 and 14 nodes. In this set of results, the link bandwidth is fixed at 100 Gbps and the maximum allowed distance is 3 hops. Since in this work the number of RU per cell is fixed to 10 and also the objective is to provide protection, then having 24 PUs per node is the minimum which can satisfy the required constraint on the cost of activation of all nodes. By increasing the value of PU in each node, the algorithm has more degrees of freedom to optimally choose the nodes needed to be activated. This trend is true for all network sizes considered.

In order to show the effect of limitation of bandwidth, in figure 6.14 the value of available PUs in a node is fixed to 500 and the maximum allowed distance is 3 hops. Despite having fairly relaxed constraint over PU and distance, having strict requirements on the bandwidth produce a higher number of active nodes. This is due to the link congestion which results in longer function chains and eventually activation of more nodes. By relaxing the bandwidth constraint, instead, the algorithm can benefit from more centralization and fewer nodes needed to be activated.

The two plots in figures 6.15 and 6.16 and present the effect of function chaining by showing the variation of bandwidth usage in Xhaul compared to C-RAN for both DPP and SPP. In both figures, the percentage of links using a certain amount of bandwidth is plotted. The maximum allowed bandwidth per each link is bounded to 100 Gbps. As indicated the main benefit of function splitting is a reduction of bandwidth usage. This is coming from the execution of baseband functions in different entities. Since in C-RAN architecture this technique does not apply (always using option 8), as a result, no uniform bandwidth usage can be seen. This fact is shown in both plots by having the majority of links occupied with a certain amount
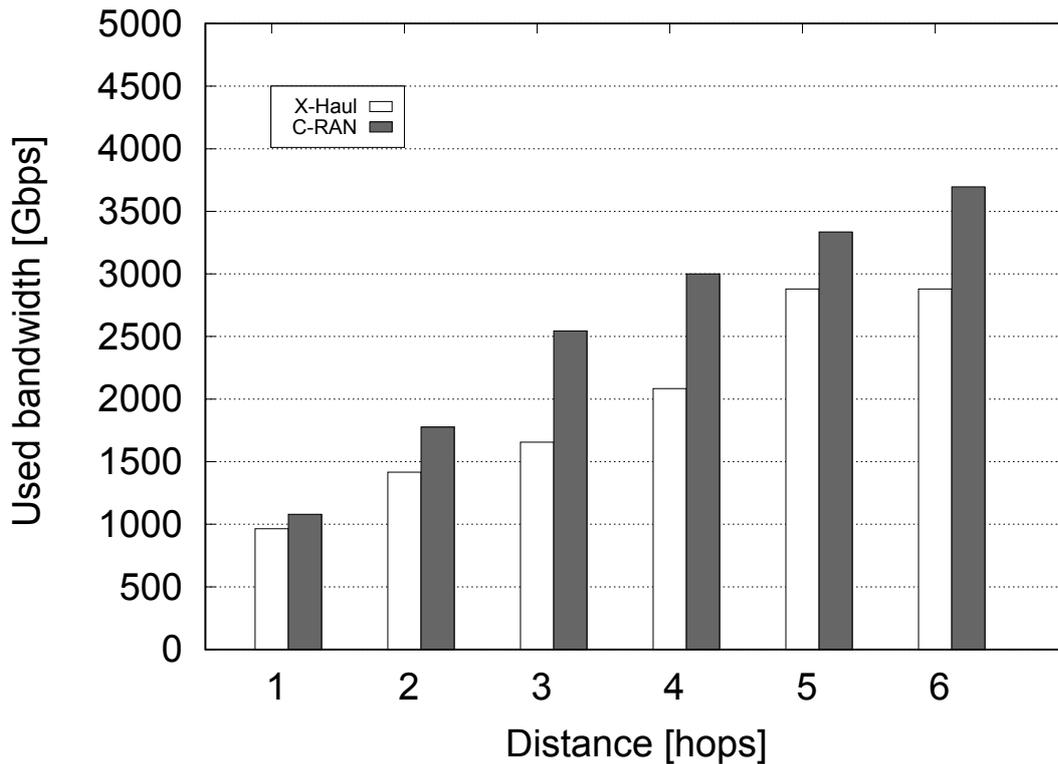
Figure 6.11: Comparison of the total used bandwidth as a function of the distance in hops for Xhaul and C-RAN with SPP protection.

of bandwidth. In the Xhaul architecture instead, thanks to the different levels of function split, a more distributed trend is evidencing.

6.5 Conclusion

This chapter has described a novel approach to location algorithm in the 5G access network, based on function chaining as defined by the Xhaul architecture. The algorithm can assign $L1$, $L2$, $L3$, *core* and *service* functionalities to nodes according to distance and processing constraints while adapting to aggregate traffic variation during the day. Furthermore, two main different algorithms, DPP and SPP are proposed and compared to assign Xhaul functionalities to optical aggregation network nodes in relation to available resources. The results show the benefits of each approach in relation to assigned constraints. In particular for services with tight latency limitations, like uRLLC, the DPP technique, combined with suitable function chaining, leads to the best multiplexing gain in terms of bandwidth. Instead, for services with more relaxed latency constraints, like eMBB, SPP combined with function chaining allows obtaining the highest multiplexing gains.
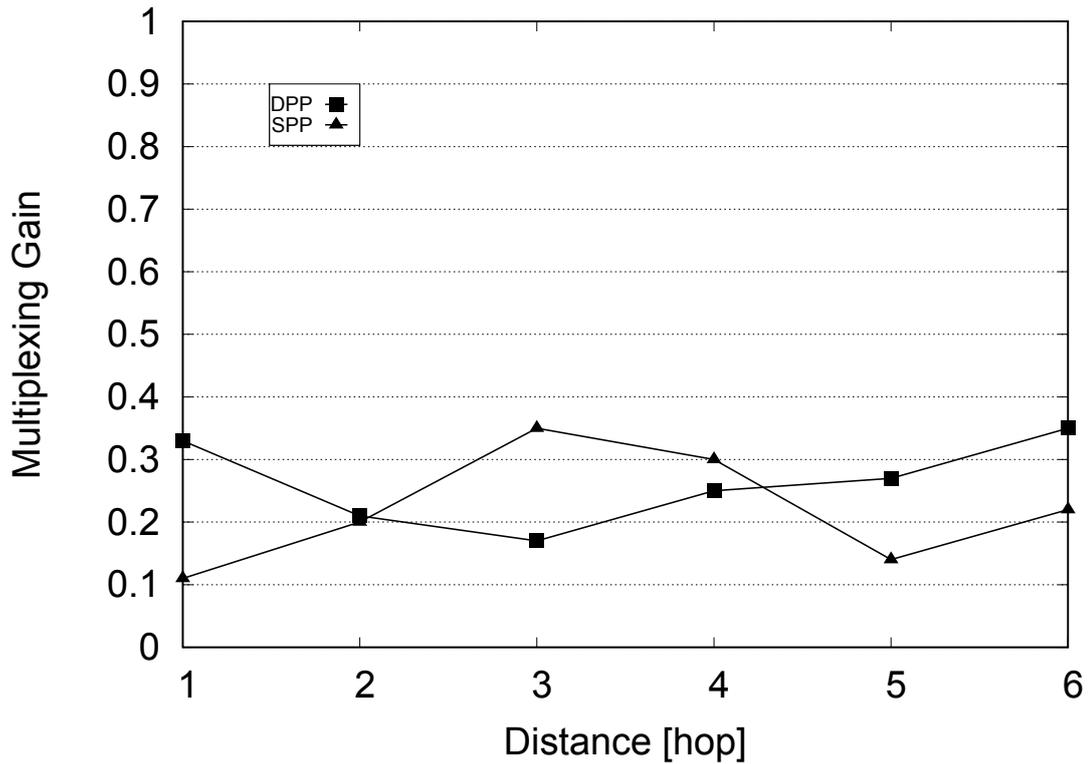
Figure 6.12: Comparison of the used bandwidth multiplexing gain in the case of DPP and SPP for Xhaul and C-RAN architectures as a function of the distance in hops.
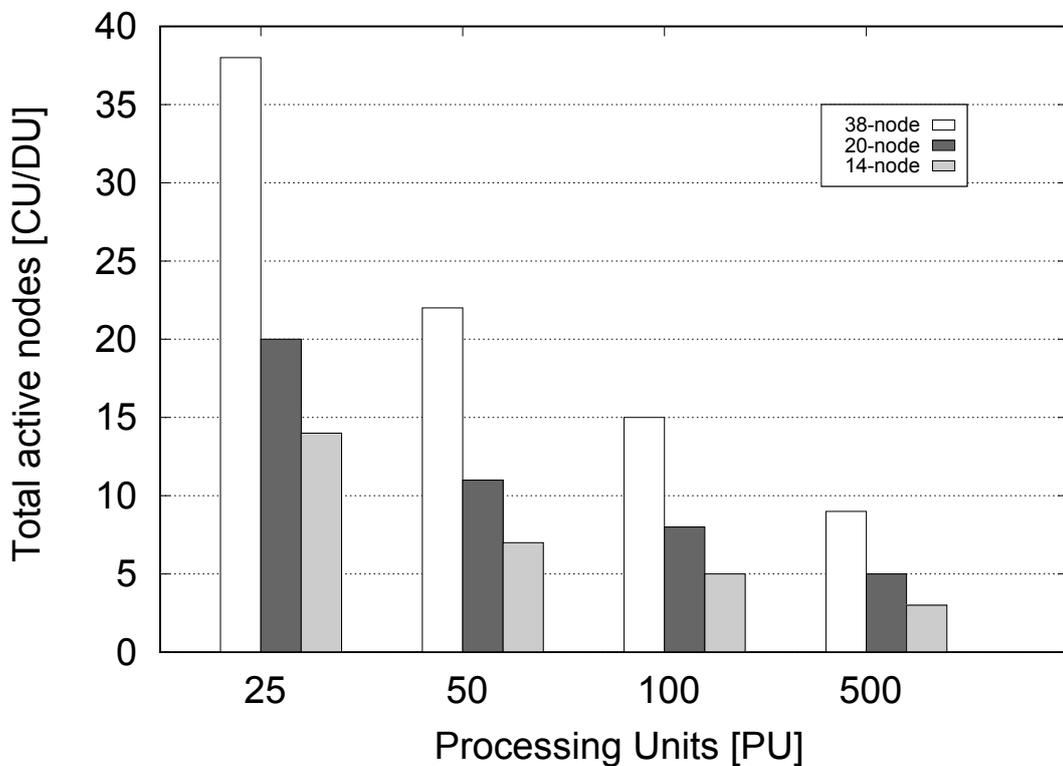
Figure 6.13: Comparison of the total number of active nodes as a function of the number of PUs for 3 network sizes in 3 hops distance with 100 Gbps bandwidth.
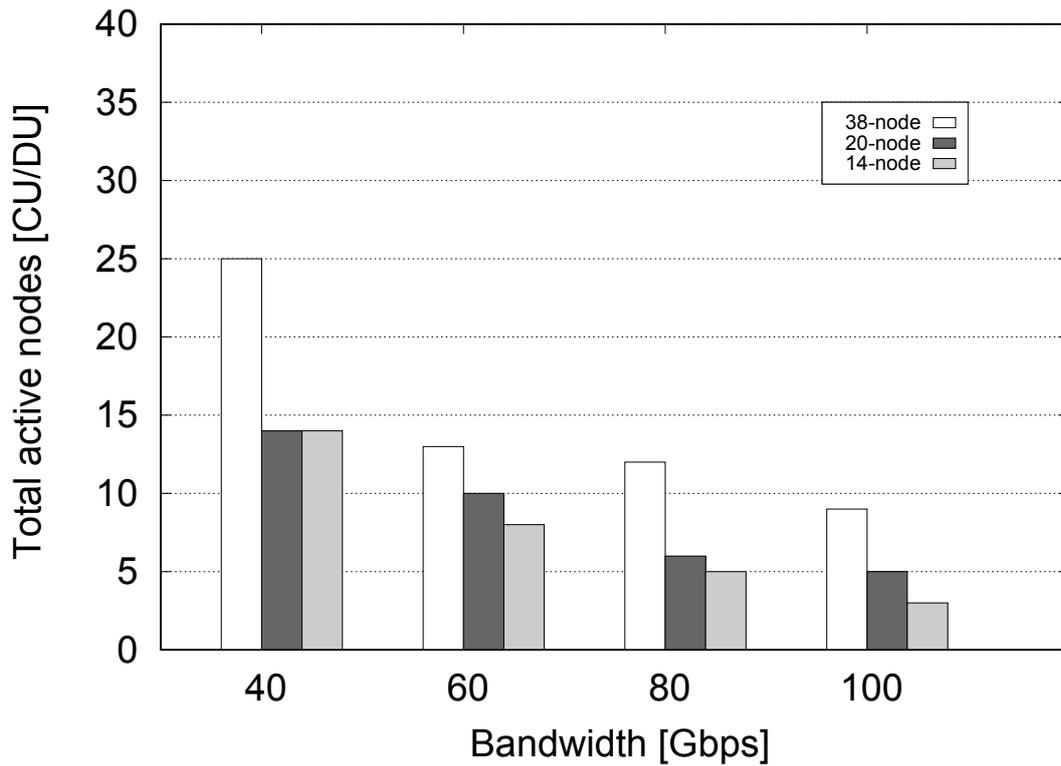
Figure 6.14: Comparison of the total number of active nodes as a function of the different bandwidth constraints for 3 network sizes in 3 hops distance with 500 PUs.
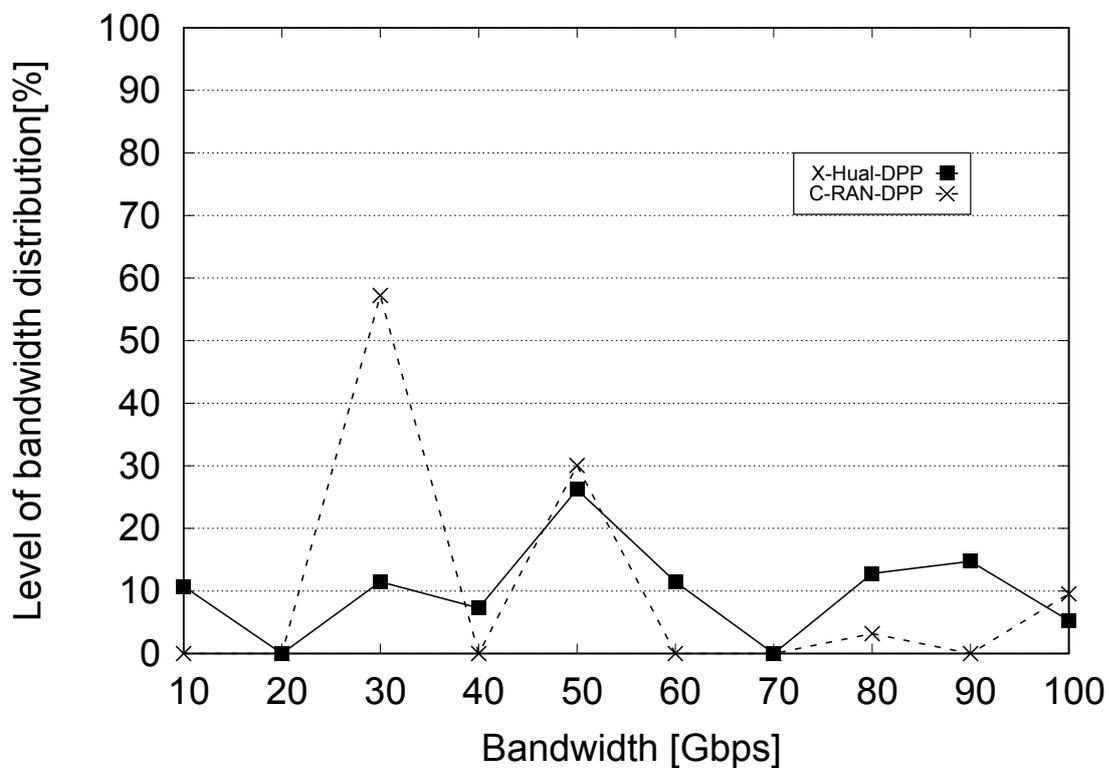
Figure 6.15: The distribution of bandwidth usage in C-RAN for the 38 node network with 3 hops distance and 100 Gbps bandwidth.
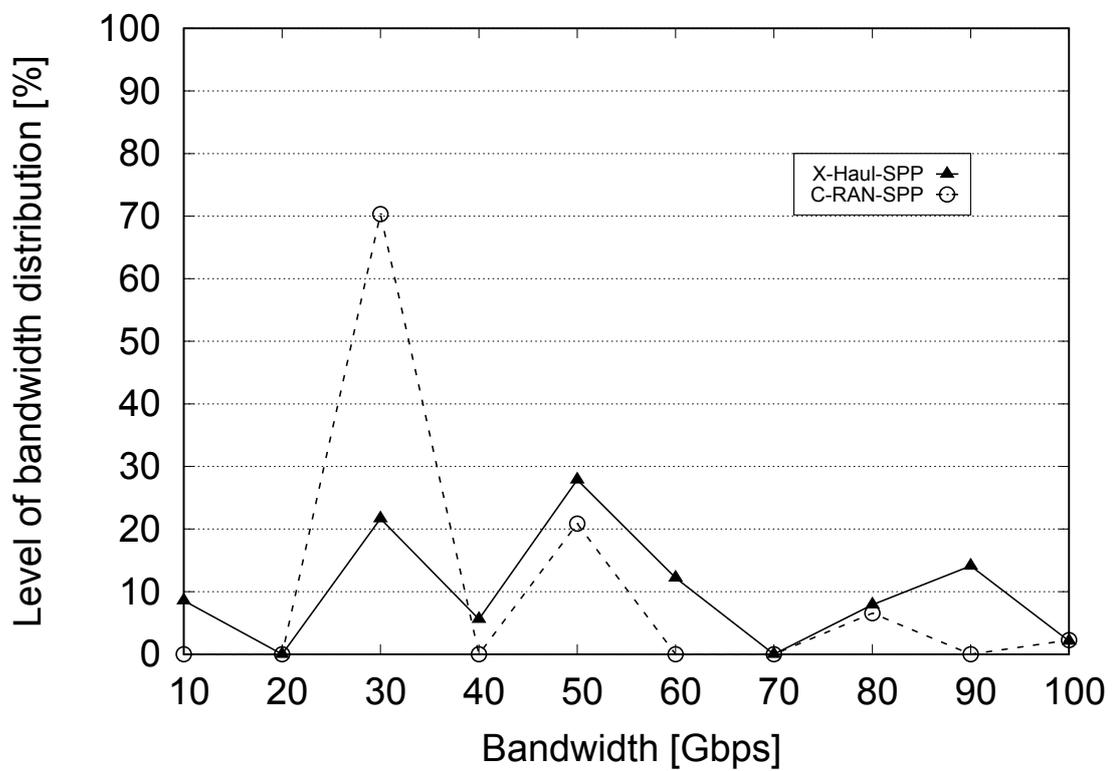
Figure 6.16: The distribution of bandwidth usage in Xhaul for the 38 node network with 3 hops distance and 100 Gbps bandwidth.

Chapter 7

Statistical Multiplexing for Packet-based Fronthaul

7.1 Introduction

In the recent architectural proposal, the highly demanding traffic, generated by functional splits, is carried together with packet-based backhaul (BH) traffic over a common optical infrastructure on different wavelength channels of the same fiber links [77]. This solution can be designed to meet the latency requirements of fronthaul (FH) traffic but may limit the system scalability and lead to low resource utilization with consequently high deployment costs. Hence, techniques for improving the optical channel utilization while meeting the strict performance requirements of FH traffic are needed and investigated in this paper. Ethernet-based links can be adopted to implement FH and standardization bodies are recently very active on the definition of the requirements to support FH traffic on these widely deployed interfaces. Delay requirements are identified as quite challenging being the Ethernet not originally designed for delay sensitive applications, such as fronthauling [78].

In this chapter an integrated hybrid architecture is reviewed and extended with pre-emption to be applied to C-RAN optical transport network. An integrated Ethernet based interface, where FH and BH traffic is multiplexed to achieve high utilization of the wavelength resource, is proposed and evaluated to meet the delay requirement of FH traffic while offering throughput capability to BH traffic.

A converged fronthaul/backhaul scenario, as depicted in figure 7.1, is considered for evaluating the additional BH throughput that can be obtained as a consequence of the application of the mechanism to wavelength channels. In figure 7.1 a sample C-RAN topology is shown where Integrated Hybrid Nodes (IHNs) are interconnected by optical links, possibly forming a mesh or other topologies. IHN is assumed to be equipped with Ethernet interfaces, which ensure high backward compatibility and low cost. RRU is connected to IHNs which host baseband unit (BBU) functionalities. The set of BBUs available in an IHN forms the BBU hotel. These functionalities can be virtualized in a C-RAN and moved throughout the network to optimize the access service.

The management of the migration of virtual BBU functionalities is performed by the SDN control/management plane, not shown in the figure. As a consequence of different possible locations for BBU hotels in IHNs, FH and BH traffic can be present on each link of the C-RAN. FH traffic is assumed according to the Common Public
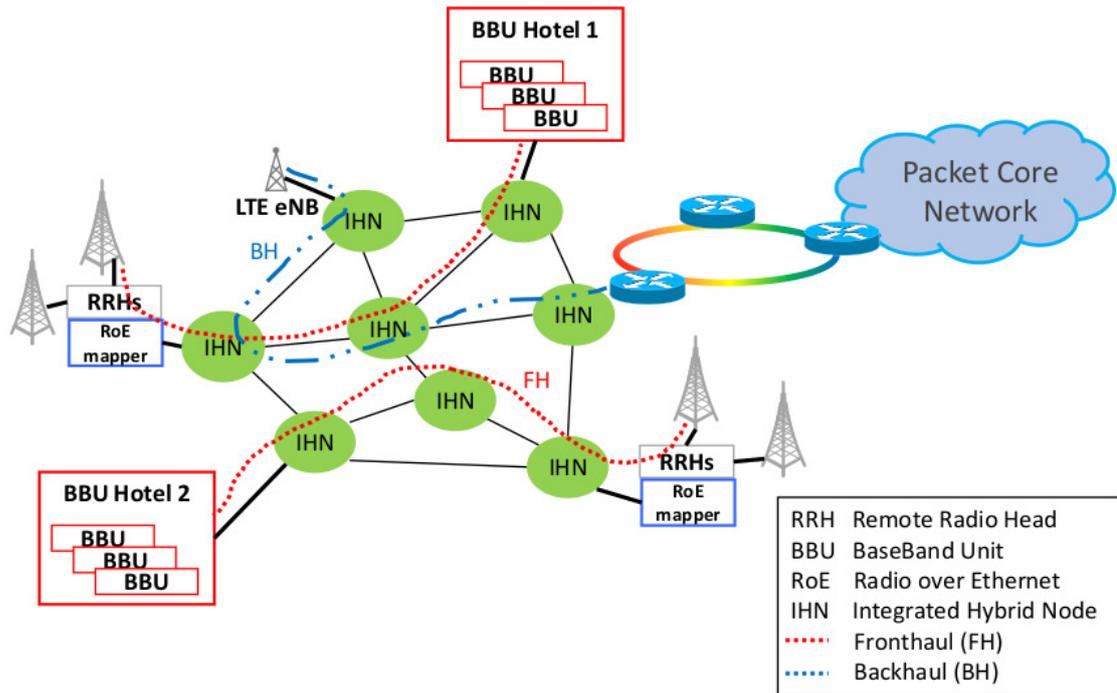
Figure 7.1: Converged fronthaul/backhaul scenario.

Radio Interface (CPRI) standard as encapsulated in Ethernet frames. The traffic characteristics of the CPRI traffic as generated by the different standardization options are taken into account and managed by a suitably extended integrated hybrid mechanism. As a consequence of the deterministic behavior of the encapsulated CPRI traffic, the benefits of introducing a segmentation policy, on the BH traffic is also considered. Segmentation is the act of carving up BH traffic into smaller pieces so they can be inserted into the possible FH traffic gaps.

7.2 Traffic Aggregation

The exponential increase in mobile users and enormous bandwidth requirements by mobile applications emphasize the need for introducing a solution to increase the throughput. Traffic aggregation is one of the well-established concepts in the networking context and in this section, the aggregation technique and how C-RAN can benefit from it will be explained in detail.

7.2.1 Integrated Hybrid Optical Network in C-RAN

Figure 7.2 shows a sample network topology implementing the C-RAN. The BBU serving an RRU can be activated in different hotels for resource optimization, service continuity or energy efficiency, thus possibly requiring dynamic association between RRUs and BBUs. This dynamic association is thought to be managed by a suitable SDN control/management plane. Moreover, traditional base stations (e.g., LTE eNB) may also be present in the same area, requiring connection to the core network. As a consequence, both FH and BH traffic need to be transferred on each optical network segment. A solution to deploy such a scenario can be to assign dedicated

Figure 7.2: IHN multiplexing scheme.

wavelength channels to each kind of traffic, either FH or BH so that FH links can be designed to meet strict delay requirements and BH traffic is statistically multiplexed on separated channels. FH delay requirement comes from the 3 milliseconds budget for a round 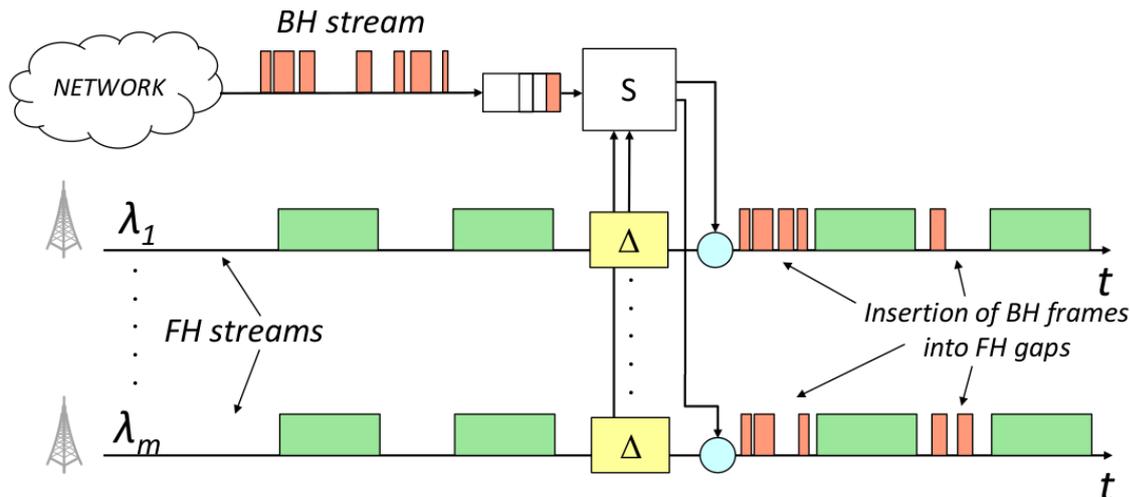trip time in Hybrid Automatic Repeat Request (HARQ). This budget includes the RRU and BBU processing which leaves only 100-200 microseconds for fronthaul one-way propagation delay. Also, CPRI has a one-way jitter requirement of few nanoseconds.

The integrated hybrid multiplexing scheme Integrated Hybrid Optical Network (IHON) was first proposed to implement statistical multiplexing of the Guaranteed Service Traffic (GST) and Statistical Multiplexed (SM) in Ethernet packet-based nodes [79]. In IHON a small fixed delay (Δ) is added to guaranteed traffic (GST) so that statistically multiplexed (SM) traffic can be inserted in GST gaps, with minimum delay and zero PDV (figure 7.2), as it was experimentally proved in [79]. To minimize the delay of GST traffic, IHON can be extended to allow GST traffic pre-emption on SM traffic and the effectiveness of this mechanism was analyzed in [80].

Here, the integrated hybrid concept with pre-emption is applied to a network segment of a C-RAN where FH traffic, i.e. CPRI flow encapsulated in Ethernet frames [81], is identified as GST, with zero PDV, while BH traffic is dealt with pre-emption as SM traffic. During the transmission of an FH frame, incoming BH packets are stored in a buffer until an output channel is free. A scheduler (represented by the block S in figure 7.2) senses the input channels to detect FH frames and is in charge of deciding when to start and interrupt the transmission of BH packets on the output channels. IHON eliminates PDV of the FH traffic because the fixed delay Δ enables a time-window which gives sufficient time for processing and decision of BH packet preemption. This goes beyond, e.g., the *IEEE 802.1Qbu* pre-emption [82] recommended in the *IEEE 802.1CM* standard [83] for FH, where FH packets may experience anyway PDV corresponding to the service time of 155 byte.
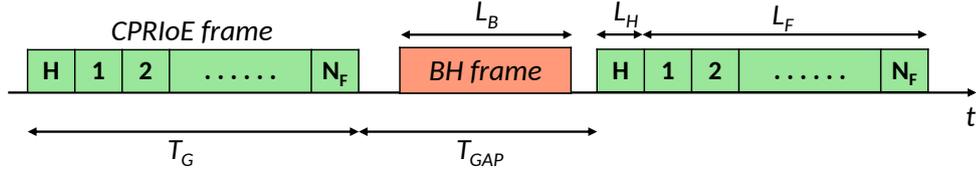
Figure 7.3: CPRIoE parameters applied to hybrid node.

Table 7.1: List of parameters used to describe CPRIoE and hybrid nodes.

Parameter	Description
N_F	Number of CPRI basic frames forming a CPRIoE payload.
L_F	Payload length for CPRIoE frame.
R_W	Output channel rate.
T_G	CPRIoE duration.
T_{GAP}	Gap duration.
Δ	Fixed delay to avoid collision.
ρ_B	Offered BH load per channel.
L_B	Average length of BH frames.
T_{guard}	Guard time.
T_{CPRI}	CPRI basic frame duration.
R_{CPRI}	CPRI flow generation rate.
L_H	Length of CPRIoE header.
m	Number of channels in the switch output interface.

7.2.2 Mapping of CPRI traffic in IHON

Common Public Radio Interface over Ethernet (CPRIoE) traffic characterization has been analyzed in [84] and [85]. A list of parameters used in this study is reported in table 7.1, while an example of the IHN output line is provided in figure 7.3. RRUs generate CPRI flows at different rates (R_{CPRI}) set by the standard [86]. Each flow is composed of CPRI basic frames with a fixed duration of $T_{CPRI} = 260\ ns$, equal for all CPRI options. A certain number of CPRI basic frames (N_F) are encapsulated in an Ethernet frame forming the CPRIoE payload of length:

$$L_F = N_F * R_{CPRI} * T_{CPRI} \tag{7.1}$$

CPRIoE frames are then sent by RRUs towards IHON switches, where they are delayed by Δ. Also, conventional BH traffic reaches the switches, loading the output channels with parameter ρ_B. In order to avoid collision between different frames on the output line, a guard time T_{guard} is applied during which the transmission of any data is not permitted. IHON switches have m output channels, each characterized by a rate R_W, and accommodates CPRIoE frames of duration:

$$T_G = \frac{L_H + L_F}{R_W} \tag{7.2}$$

where L_H is the header of CPRIoE frames assumed to be 44 byte [86].

Depending on N_F and R_{CPRI}, the gap duration T_{GAP} is selected according to:
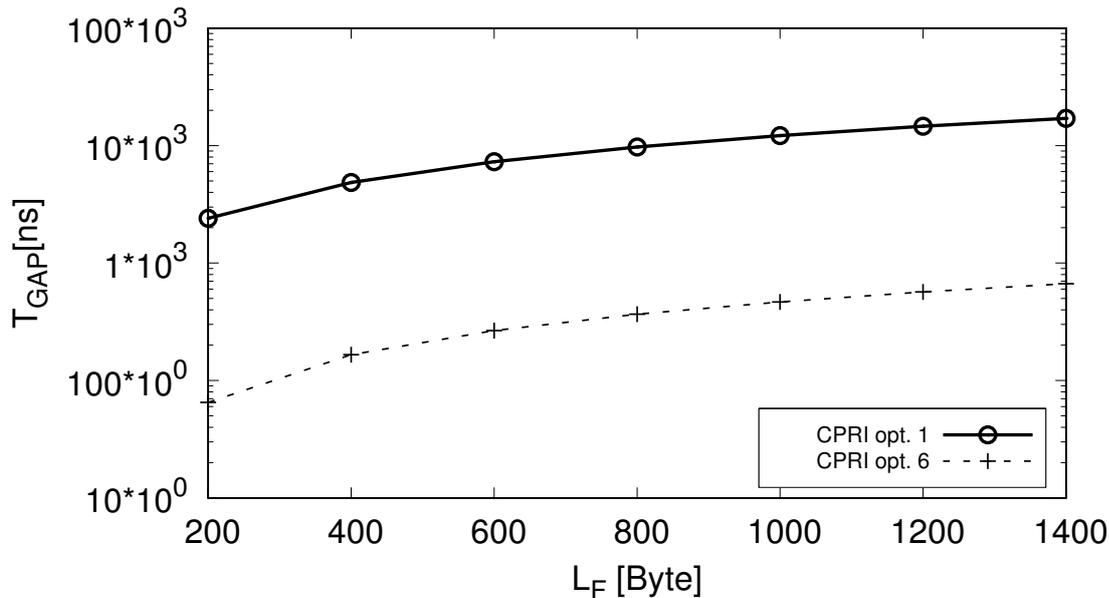
Figure 7.4: T_{GAP} as a function of different values of payload length L_F for CPRI option 1 and 6 on a 10 *Gbps* line.

$$T_{GAP} = \frac{L_F}{R_{CPRI}} - T_G \tag{7.3}$$

By looking at equation 7.3, it is possible to notice that, depending on R_{CPRI}, different values for T_{GAP} can be obtained for the same length of CPRIoE packets L_F. An example of T_{GAP} using CPRI option 1 and 6 for a line rate of $R_W = 10$ *Gbps* is depicted in figure 7.4.

This time gap is used in the hybrid multiplexing scheme to aggregate BH traffic on the same transport channel. To this end, two different policies are here considered:

- A BH packet is transmitted when a gap is available and it is possibly pre-empted upon arrival of a new GST burst, in case BH packet duration is longer than the gap itself. In the case of pre-emption, the BH packet is lost. This policy is indicated as P policy, with the insertion of an entire packet into the by-pass GST flow.

- A BH packet waits for a gap and in case the BH packet is longer than the gap it splits into segments that are transmitted by as many gaps as needed. This avoids the need for pre-emption but introduces some overhead to manage segmentation and additional functionalities. This policy is indicated as S policy, where packets are divided into N_S segments of the suitable size for their insertion into the GST flow.

7.3 Numerical Results

To evaluate the benefits introduced by the proposed mechanism, an event driven simulator in C++ language has been developed. One RRU generates a CPRI flow

Figure 7.5: BH success probability as a function of payload length L_F for different BH packet length L_B using CPRI option 1 and 6.

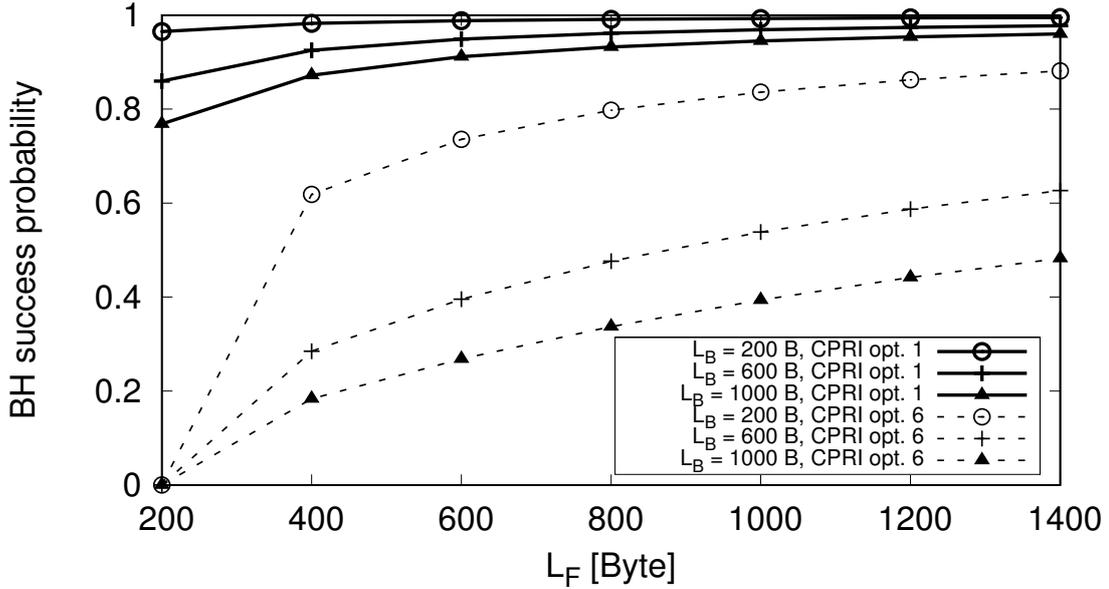according to two different options with rates $R_{CPRI} = 614.4\ Mbps$ (option 1) and $6.144\ Gbps$ (option 6). The IHON fixed delay $\Delta = 99.2\ ns$ is assumed, which corresponds to the smallest fragment ($124\ byte$) that can be preempted [87]. A time guard of 10 ns between frames is applied. A single output channel ($m = 1$) with rate $R_W = 10\ Gbps$ is considered. The number of CPRI basic frames in a guaranteed burst N_F is varied over the intervals $[1, 70]$ and $[1, 7]$, for CPRI option 1 and 6, respectively [88], so that the payload length L_F varies accordingly. A set of simulations varying the average BH packet length L_B is obtained with a load ρ_B such that a BH packet is always ready for transmission on the output channel. The length of BH frames is considered to be exponentially distributed with parameter L_B.

Figure 7.5 shows the success probability of the BH traffic, defined as the ratio between the packets not interrupted and the total packets in service, as a function of L_F, for both CPRI options, varying L_B. In both cases, the success probability increases with L_F, due to the resulting larger T_{gap}. Option 6 shows lower performance than option 1 due to the smaller size of the gap, especially when L_F is low, so suggesting to use larger N_F in this case. However, increasing N_F increases the encapsulation delay, which may impact the maximum reach of the FH connection.

Figure 7.6 reports the BH throughput, normalized to the output line rate ($10\ Gbps$), as a function of L_F for option 1 varying L_B. The figure also reports the maximum normalized capacity left by FH traffic. The value of throughput in the case of the P policy reaches 8.9 $Gbps$ only for high values of L_F with quite limited influence of L_B. The S policy, instead, is able to better exploit the available capacity for any value of L_F, except for the influence of the transmission guard times inserted. The same evaluation obtained for an option 6 in figure 7.7 shows a remarkable effect of the shorter gaps in the FH flow, that prevents also the F policy to fully exploit the available capacity for low values of L_F, due to the high numbers of segments needed
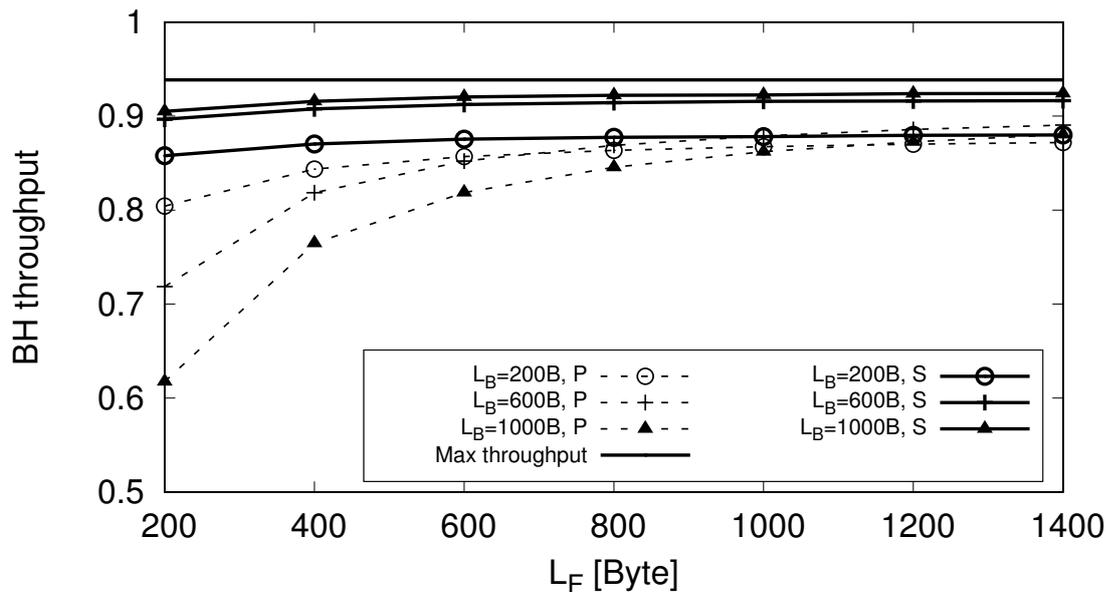
Figure 7.6: BH throughput, normalized to the output link capacity, as a function of payload length L_F for different BH packet length L_B using CPRI option 1. Solid lines for the case with segmentation (S), dashed lines for the no-segmentation case(P).

and related inserted transmission guard time.

Figure 7.8 reports the overhead introduced by the P and S policies calculated as the ratio of the number of bytes for Ethernet headers and the total number of bytes transmitted as BH traffic for an option 1. The same evaluation is presented in figure 7.9 for an option of 6. The effect of the S policy is more evident with option 6 where, due to the smaller gaps in the FH flow, multiple segments are typically required to transmit each BH packet. In any case, the additional overhead is quite limited when increasing L_F.

It is interesting to analyze the average number of segments to transmit BH packets in option 1 and option 6, as shown in figure 7.10 for the S policy. Option 1 allows the transmission of a packet as a single segment in most cases for any L_F. In option 6, instead, reasonable values of L_F seem to be not less than 1000 bytes which give an average number of segments less than 3 for any L_B, with a resulting overhead around 10%, which is reasonable as well. However, working with high L_F increases the encapsulation delay, which in the worst case is $18.3\mu s$ for CPRI option 1 and $1.83\mu s$ for CPRI option 6.

7.4   Conclusion

This section has explored the feasibility of the integrated hybrid network concept with pre-emption applied to C-RAN for transporting FH traffic with timing transparency combined with pre-emptive BH traffic within the same optical Ethernet channel. Performance evaluations have been presented for different CPRI options, finding the amount of BH traffic taking advantage of unused FH capacity. Remarkable BH throughput is shown especially for CPRI option 1. Scheduling BH packets only when gaps in the FH traffic of suitable size for the BH packets are present, is
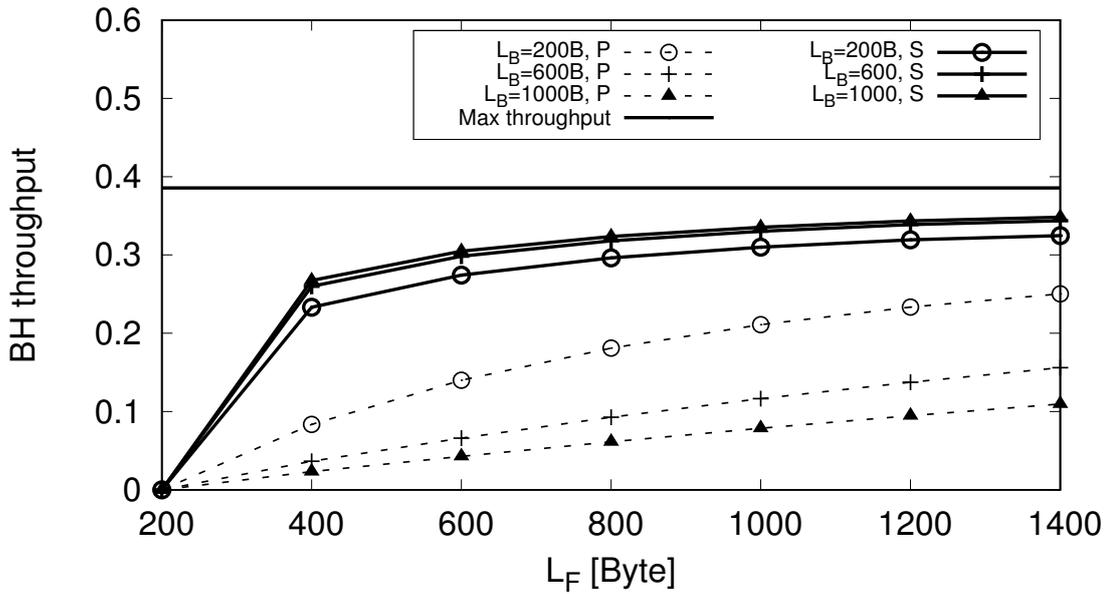
Figure 7.7: BH throughput, normalized to the output link capacity, as a function of payload length L_F for different BH packet length L_B using CPRI option 6. Solid lines for the case with segmentation (S), dashed lines for the no-segmentation case (P).
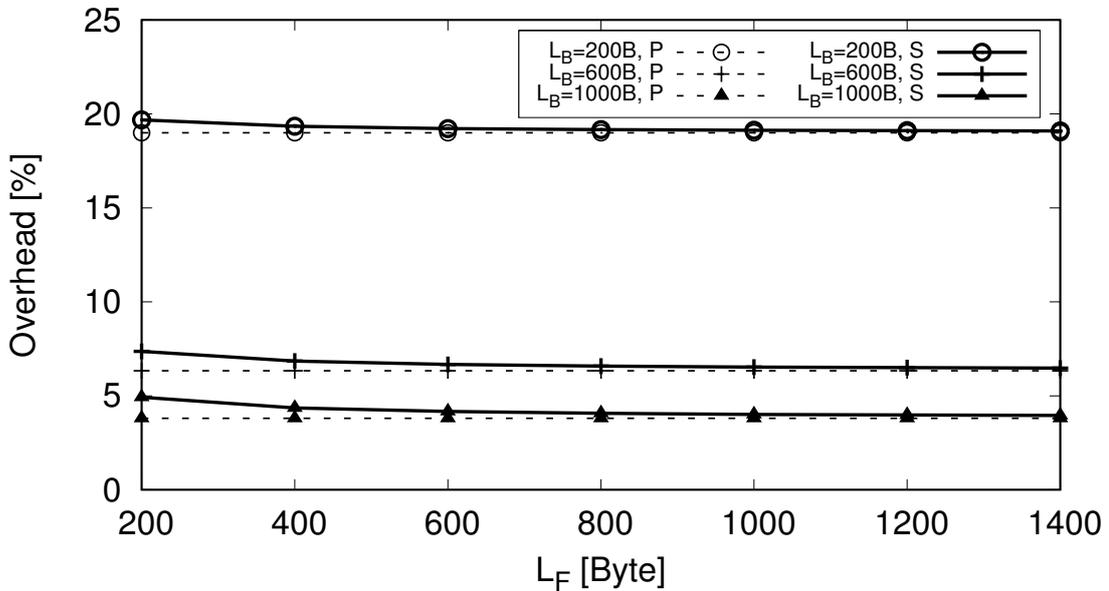
Figure 7.8: Overhead for BH packets as a function of payload length L_F for different BH packet length L_B using CPRI option 1. Solid lines for the case with segmentation (S), dashed lines for the no-segmentation case (P).
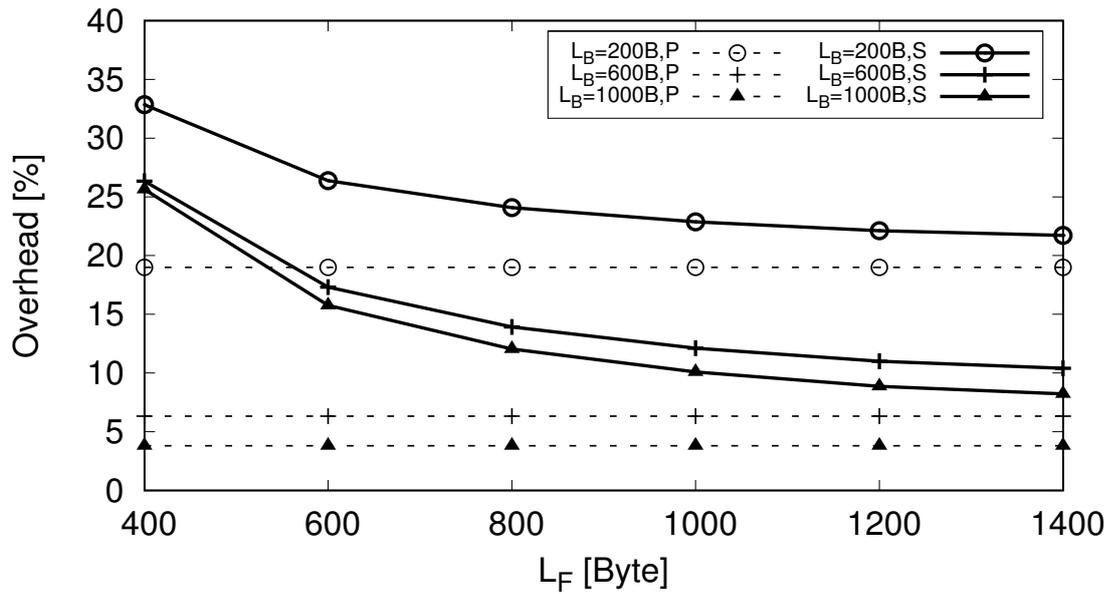
Figure 7.9: Overhead for BH packets as a function of payload length L_F for different BH packet length L_B using CPRI option 6. Solid lines for the case with segmentation (S), dashed lines for the no-segmentation case (P).
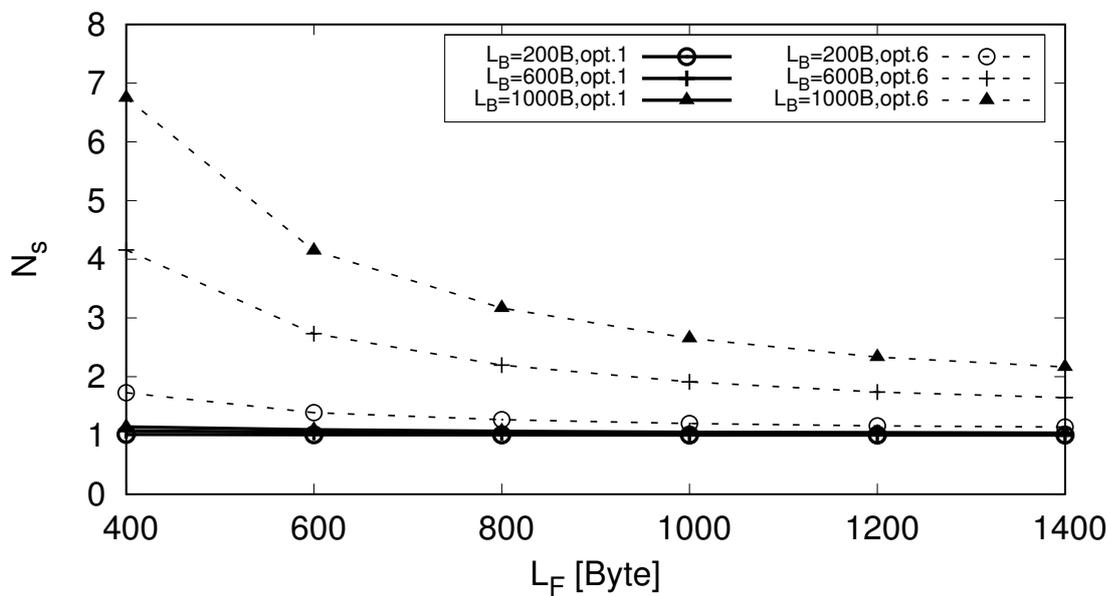
Figure 7.10: The average number of segments (N_S) required to send a BH packet as a function of payload length L_F for different BH packet length L_B using CPRI option 1 and 6.

an IHON characteristic. CPRI option 6 limits BH throughput because the smaller packet gaps in the FH traffic makes fitting of the BH traffic more difficult. Further investigations will include the introduction of controlled traffic mechanisms for adapting the BH traffic for better fitting the unused bandwidth in an FH multi-channel configuration.

Chapter 8

Concluding Remarks

This thesis illustrates the main outcomes of the research activities conducted throughout the three years of the Ph.D. program. Novel strategies for reliable deployment of C-RAN for 5G radio access networks have been proposed.

Chapter 3 presents a centralized solution based on the Facility Location Problem (FLP) for BBU hotel placement in C-RAN to achieve protection in the fronthaul optical network segment against single BBU hotel failure. Different solutions have been proposed and compared in terms of relevant cost parameters, namely the number of BBU hotels, ports and wavelengths. Additional costs with respect to solutions without protection are evaluated showing the effectiveness of the proposed algorithms to maintain additional low costs.

In chapter 4 the results obtained by the distributed algorithm are sub-optimal with respect to the centralized approach. However, the distributed algorithms can off-load the SDN orchestrator and smoothly adapt to changes in the network topology. Also, the proposed distributed algorithm is effective with dense networks showing better scalability than centralized optimal solutions.

Chapter 6 has described a novel approach to location algorithm in the 5G access network, based on function chaining as defined by the Xhaul architecture. The algorithm can assign different split layers, $L1$, $L2$, $L3$, $Core$ and $Service$ functionalities to nodes according to distance and processing constraints while adapting to aggregate traffic variation during the day. In the same line of the research two protection algorithms also has been presented namely, Dedicated Path Protection (DPP) and Shared Path Protection (SPP) to provide services even in case of single DU/CU or link failure. Furthermore, the shared methodology can achieve significant savings in terms of network resources.

In Chapter 7, the problem of sharing optical transport network resources is addressed. Here, it is proposed to use Ethernet encapsulation of CPRI frames to multiplex backhaul and fronthaul traffic over the same wavelengths, thus enabling high resource utilization.

Glossary

5G The fifth generation of mobile technology. 3, 9, 20

mMTC massive Machine Type Communication. 3, 20, 97

eMBB enhanced Mobile Broadband. 3, 20, 97, 119

uRLLC ultra Reliable Low Latency Communication. 3, 20, 97

C-RAN Cloud/Centralized Radio Access Network. 3, 14, 22, 25, 36, 68, 86, 90

CAPEX Capital Expenditure. 3

OPEX Operational Expenditure. 3

Xhaul front/mid/backhaul. 3, 97

SDN Software-Defined Networking. 3, 14, 36, 37, 68, 83, 90

NFV Network Function Virtualization. 3, 37

PU Processing Units. 3, 98, 105, 107, 116

BBU Baseband Unit. 3, 22, 26, 32, 100, 125

CU Central Unit. 3, 98, 100

DU Digital Unit. 3, 26, 37, 98, 100

FDA Fixed Distance Algorithm. 9, 39, 40

VDA Variable Distance Algorithm. 9, 39, 50

LTE Long-Term Evolution. 20, 25, 33, 97, 100

AR Augmented Reality. 20

VR Virtual Reality. 20

V2X Vehicle-to-Everything. 20, 21

E2E end-to-end. 21

3GPP 3rd Generation Partnership Project. 21, 33, 98, 101

QoS Quality of Service. 21, 22

TX Transmit. 21

RX Receive. 21

MIMO Multiple Input Multiple Output. 21, 25, 27, 33

EPS Evolved Packet System. 21

RAN Radio Access Network. 21, 25, 97

E-UTRAN Evolved Universal Terrestrial Radio Access Network. 21

EPC Evolved Packet Core. 21, 26

IoT Internet of Things. 21

D-RAN Distributed Radio Access Network. 21, 25, 29, 34

BS Base Station. 21, 25, 26, 97, 100

RRU Remote Radio Unit. 21, 26, 32

DAC Digital-to-Analog. 22

ADC Analog-to-Digital. 22

CoMP Coordinated Multi-point. 22, 34, 99, 103

NGFI Next Generation Fronthaul Interface. 22, 101, 105

WDM Wavelength Division Multiplexing. 22, 36

APS Automatic Protection Switching. 23

DSP Demand-wise Shared Protection. 23

SBPP Shared Backup Path Protection. 23

ML Machine Learning. 23, 36, 82, 83

DPP Dedicated Path Protection. 24, 107

SPP Shared Path Protection. 24, 107

OFDM Orthogonal Frequency Division Multiplexing. 25

RRH Remote Radio Head. 26

RU Radio Unit. 26, 98, 100

PA Power Amplifier. 26

RF Radio Frequency. 26, 31

eNodeB E-UTRAN Node B, also known as Evolved Node B. 26

S-GW Serving Gateway. 27

MME Mobility Management Entity. 27

RRM Radio Resource Management. 27

ICIC Inter-Cell Interference Coordination. 27

HetSNets Heterogeneous and Small cell Networks. 27

IQ In-phase and Quadrature. 28, 29, 100

PON Passive Optical Network. 28

UE User Equipment. 29

BSC Base Station Controllers. 29

RNC Radio Network Controllers. 29

AC Alternating Current. 29

DC Direct Current. 29

CPRI Common Public Radio Interface. 29, 32, 36, 56, 98, 100, 126

GPP General Purpose Processors. 30

CPU Centralized Processing Unit. 30

OBSAI Open Base Station Standard Initiative. 31

IP Internet Protocol. 31

ATM Asynchronous Transfer Mode. 31

REC Radio Equipment Control. 32

RAE Radio Access Equipment. 32

ETSI European Telecommunications Standards Institute. 32

ISG Industry Specification Group. 32

ORI Open Radio Interface. 32

HARQ Hybrid Automatic Repeat Request. 33, 102, 105, 127

eCPRI Ethernet CPRI. 33, 98, 100

RAT Radio Access Technology. 34

vRAN Virtualized RAN. 37

FLP Facility Location Problem. 39, 50, 53

ILP Integer Linear Programming. 50

FD Fixed Distance. 57

VD Variable Distance. 57

DFL Distributed Facility Location. 68

DC Data Center. 69, 73

TTL Time To Live. 74

ML-DFL ML distributed facility location algorithm. 83

TDS Training Data Sets. 83

RL Reinforcement Learning. 83

MEC Multi-access Edge Computing. 86

RSU Road Side Units. 87

QoE Quality of Experience. 97

PoP points of presence. 98

NR new access network. 99

FFT Fast Fourier Transformation. 101

PDU Protocol Data Unit. 102

DL Down Link. 102

SDU Service Data Unit. 102

UL Up Link. 102

ARQ Automatic Repeat Request. 102

BW Bandwidth. 107, 110, 116

BH backhaul. 125

FH fronthaul. 125

IHN Integrated Hybrid Nodes. 125

IHON Integrated Hybrid Optical Network. 127

GST Guaranteed Service Traffic. 127

SM Statistical Multiplexed. 127

CPRIoE Common Public Radio Interface over Ethernet. 128

Bibliography

[1]   NGMN Alliance. "5G white paper". In: *Next generation mobile networks, white paper* 1 (2015).

[2]   "Cisco Vision: 5G Thriving Indoors". In: (2017).

[3]   Karen Campbell et al. "The 5G economy: How 5G technology will contribute to the global economy". In: *IHS Economics and IHS Technology* 4 (2017), p. 16.

[4]   NGMN Alliance. *NGMN 5G white paper, white paper, 2015.* 2016.

[5]   Carsten Bockelmann et al. "Towards massive connectivity support for scalable mMTC communications in 5G networks". In: *IEEE access* 6 (2018), pp. 28969–28992.

[6]   Petar Popovski et al. "Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC)". In: *IEEE Transactions on Communications* (2019).

[7]   Müge Erel-Özçevik and Berk Canberk. "Road to 5G Reduced-Latency: A Software Defined Handover Model for eMBB Services". In: *IEEE Transactions on Vehicular Technology* 68.8 (2019), pp. 8133–8144.

[8]   Sandip Gangakhedkar et al. "Use cases, requirements and challenges of 5G communication for industrial automation". In: *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE. 2018, pp. 1–6.

[9]   Aleksandar Damnjanovic et al. "A survey on 3GPP heterogeneous networks". In: *IEEE Wireless communications* 18.3 (2011), pp. 10–21.

[10] Konstantinos Liolis et al. "Use cases and scenarios of 5G integrated satellite-terrestrial networks for enhanced mobile broadband: The SaT5G approach". In: *International Journal of Satellite Communications and Networking* 37.2 (2019), pp. 91–112.

[11] Amitava Ghosh et al. "LTE-advanced: next-generation wireless broadband technology". In: *IEEE wireless communications* 17.3 (2010), pp. 10–22.

[12] Mona Jaber et al. "A joint backhaul and RAN perspective on the benefits of centralised RAN functions". In: *2016 IEEE International Conference on Communications Workshops (ICC)*. IEEE. 2016, pp. 226–231.

[13] Jun Wu et al. "Cloud radio access network (C-RAN): a primer". In: *IEEE Network* 29.1 (2015), pp. 35–41.

[14] Andreas Maeder et al. "Towards a flexible functional split for cloud-RAN networks". In: *2014 European Conference on Networks and Communications (EuCNC)*. IEEE. 2014, pp. 1–5.

[15] Jens Bartelt et al. "5G transport network requirements for the next generation fronthaul interface". In: *EURASIP Journal on Wireless Communications and Networking* 2017.1 (2017), p. 89.

[16] Gero Schollmeier and Christian Winkler. "Providing sustainable QoS in next-generation networks". In: *IEEE Communications Magazine* 42.6 (2004), pp. 102–107.

[17] Pankaj K Agarwal et al. "The resilience of WDM networks to probabilistic geographical failures". In: *IEEE/ACM Transactions on Networking* 21.5 (2013), pp. 1525–1538.

[18] Himanshi Saini and Amit Kumar Garg. "Protection and restoration schemes in optical networks: a comprehensive survey". In: *International Journal of Microwaves Applications* 2.1 (2013).

[19] Nakamura Takaharu. "LTE and LTE-advanced: Radio technology aspects for mobile communications". In: *2011 XXXth URSI General Assembly and Scientific Symposium*. IEEE. 2011, pp. 1–4.

[20] Aleksandra Checko et al. "Cloud RAN for mobile networks—A technology overview". In: *IEEE Communications surveys & tutorials* 17.1 (2014), pp. 405–426.

[21] Cisco Visual Networking Index. "Global mobile data traffic forecast update, 2016–2021 white paper". In: *Cisco: San Jose, CA, USA* (2017).

[22] Insoo Hwang, Bongyong Song, and Samir S Soliman. "A holistic view on hyper-dense heterogeneous and small cell networks". In: *IEEE Communications Magazine* 51.6 (2013), pp. 20–27.

[23] DW Bliss and KW Forsythe. "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 1. IEEE. 2003, pp. 54–59.

[24] Thomas L Marzetta. "Massive MIMO: an introduction". In: *Bell Labs Technical Journal* 20 (2015), pp. 11–22.

[25] China Mobile. "C-RAN: the road towards green RAN". In: *White paper, ver* 2 (2011), pp. 1–10.

[26] Thomas Pfeiffer. "Next generation mobile fronthaul and midhaul architectures". In: *Journal of Optical Communications and Networking* 7.11 (2015), B38–B45.

[27] Matteo Fiorani et al. "On the design of 5G transport networks". In: *Photonic network communications* 30.3 (2015), pp. 403–415.

[28] I Chih-Lin et al. "Toward green and soft: A 5G perspective". In: *IEEE Communications Magazine* 52.2 (2014), pp. 66–73.

[29] Matteo Fiorani et al. "Joint design of radio and transport for green residential access networks". In: *IEEE Journal on Selected Areas in Communications* 34.4 (2016), pp. 812–822.

[30] CPRI Specification. "V4. 2". In: *Common Public Radio Interface (CPRI)* (2010).

[31]   Navid Nikaein et al. "Closer to Cloud-RAN: RAN as a Service". In: *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM. 2015, pp. 193–195.

[32]   Open Base Station Architecture Initiative et al. *'Reference Point 3 Specification Version 4.0'*.

[33]   Open Radio equipment Interface. *Requirements for Open Radio equipment Interface (ORI)(Release 4), ETSI Std., Oct. 2014*.

[34]   Nicola Carapellese, Massimo Tornatore, and Achille Pattavina. "Energy-efficient baseband unit placement in a fixed/mobile converged WDM aggregation network". In: *IEEE Journal on Selected Areas in Communications* 32.8 (2014), pp. 1542–1551.

[35]   Wolfgang Kiess et al. "Protection scheme for passive optical networks shared between a fixed and a mobile operator". In: *OFC/NFOEC*. IEEE. 2012, pp. 1–3.

[36]   Ashraf Awadelakrim Widaa, Jan Markendahl, and Amirhossein Ghanbari. "Toward capacity-efficient, cost-efficient and power-efficient deployment strategy for indoor mobile broadband". In: (2013).

[37]   Marija Furdek, Nina Skorin-Kapov, and Lena Wosinska. "Attack-aware dedicated path protection in optical networks". In: *Journal of Lightwave Technology* 34.4 (2015), pp. 1050–1061.

[38]   M Farhan Habib et al. "Disaster survivability in optical communication networks". In: *Computer Communications* 36.6 (2013), pp. 630–644.

[39]   M Farhan Habib et al. "Design of disaster-resilient optical datacenter networks". In: *Journal of Lightwave Technology* 30.16 (2012), pp. 2563–2573.

[40]   Nick McKeown. "Software-defined networking". In: *INFOCOM keynote talk* 17.2 (2009), pp. 30–32.

[41]   Bo Han et al. "Network function virtualization: Challenges and opportunities for innovations". In: *IEEE Communications Magazine* 53.2 (2015), pp. 90–97.

[42]   Xinbo Wang et al. "Green virtual base station in optical-access-enabled cloud-RAN". In: *2015 IEEE International Conference on Communications (ICC)*. IEEE. 2015, pp. 5002–5006.

[43]   Gérard Cornuéjols, George Nemhauser, and Laurence Wolsey. *The Uncapicitated Facility Location Problem*. Tech. rep. Cornell University Operations Research and Industrial Engineering, 1983.

[44]   Shuqiang Zhang, Ming Xia, and Stefan Dahlfort. "Fiber routing, wavelength assignment and multiplexing for DWDM-centric converged metro/aggregation networks". In: *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*. IET. 2013, pp. 1–3.

[45]   Bahare Masood Khorsandi et al. "Survivable BBU Hotel placement in a C-RAN with an Optical WDM Transport". In: *DRCN 2017-Design of Reliable Communication Networks; 13th International Conference*. VDE. 2017, pp. 1–6.

[46]   Michal Pióro and Deep Medhi. *Routing, flow, and capacity design in communication and computer networks*. Elsevier, 2004.

[47] Luohao Tang et al. "Reliable facility location problem with facility protection". In: *PloS one* 11.9 (2016), e0161532.

[48] Christian Frank and Kay Römer. "Distributed facility location algorithms for flexible configuration of wireless sensor networks". In: *International Conference on Distributed Computing in Sensor Systems*. Springer. 2007, pp. 124–141.

[49] Paulo Valente Klaine et al. "A survey of machine learning techniques applied to self-organizing cellular networks". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2392–2431.

[50] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[51] Aleksandra Checko et al. "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings". In: *Journal of Communications and Networks* 18.2 (2016), pp. 162–172.

[52] Matteo Fiorani et al. "Transport abstraction models for an SDN-controlled centralized RAN". In: *IEEE Communications Letters* 19.8 (2015), pp. 1406–1409.

[53] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[54] Donald Michie, David J Spiegelhalter, CC Taylor, et al. "Machine learning". In: *Neural and Statistical Classification* 13 (1994).

[55] Matteo Fiorani et al. "Abstraction models for optical 5G transport networks". In: *IEEE/OSA Journal of Optical Communications and Networking* 8.9 (2016), pp. 656–665.

[56] Kourosh Naderi et al. "A reinforcement learning approach to synthesizing climbing movements". In: *2019 IEEE Conference on Games (CoG)*. IEEE. 2019, pp. 1–7.

[57] Arga Dwi Pambudi, Trihastuti Agustinah, and Rusdhianto Effendi. "Reinforcement Point and Fuzzy Input Design of Fuzzy Q-Learning for Mobile Robot Navigation System". In: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. IEEE. 2019, pp. 186–191.

[58] Sami Kekki et al. "MEC in 5G networks". In: *ETSI white paper* 28 (2018), pp. 1–28.

[59] Line MP Larsen, Aleksandra Checko, and Henrik L Christiansen. "A survey of the functional splits proposed for 5G mobile crosshaul networks". In: *IEEE Communications Surveys & Tutorials* 21.1 (2018), pp. 146–172.

[60] Federico Tonini et al. "Scalable Edge Computing Deployment for Reliable Service Provisioning in Vehicular Networks". In: *Journal of Sensor and Actuator Networks* 8.4 (2019), p. 51.

[61] Small Cell Forum. "Small cell virtualization: Functional splits and use cases". In: (2015).

[62] Michael Rüßmann et al. "Industry 4.0: The future of productivity and growth in manufacturing industries". In: *Boston Consulting Group* 9.1 (2015), pp. 54–89.

[63] GreenTouch Consortium et al. *Global study by GreenTouch consortium reveals how communications networks could reduce energy consumption by 90 percent by 2020*. 2013.

[64] Antonio De La Oliva et al. "Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks". In: *IEEE Wireless Communications* 22.5 (2015), pp. 32–40.

[65] 3GPP TR 38.801. "Study on new radio access technology: Radio access architecture and interfaces". In: (2017).

[66] Jay Kant Chaudhary et al. "C-RAN Employing xRAN Functional Split: Complexity Analysis for 5G NR Remote Radio Unit". In: *2019 European Conference on Networks and Communications (EuCNC)*. IEEE. 2019, pp. 580–585.

[67] Alexandros Kaloxylos. "A survey and an analysis of network slicing in 5G networks". In: *IEEE Communications Standards Magazine* 2.1 (2018), pp. 60–65.

[68] CPRI Consortium et al. *eCPRI Specification V1. 0*. 2017.

[69] Abdulrahman Alabbasi, Miguel Berg, and Cicek Cavdar. "Delay Constrained Hybrid CRAN: A Functional Split Optimization Framework". In: *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2018, pp. 1–7.

[70] I Chih-Lin et al. "Recent progress on C-RAN centralization and cloudification". In: *IEEE Access* 2 (2014), pp. 1030–1039.

[71] Bin Guo et al. "CPRI compression transport for LTE and LTE-A signal in C-RAN". In: *7th International Conference on Communications and Networking in China*. IEEE. 2012, pp. 843–849.

[72] AB Ericsson et al. *Common Public Radio Interface (CPRI) Specification V6. 0*. 2013.

[73] I Chih-Lin et al. "NGFI, the xHaul". In: *2015 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2015, pp. 1–6.

[74] NGMN Alliance. "Further study on critical C-RAN technologies". In: *Next Generation Mobile Networks* (2015).

[75] Bahare M Khorsandi and Carla Raffaelli. "BBU location algorithms for survivable 5G C-RAN over WDM". In: *Computer Networks* 144 (2018), pp. 53–63.

[76] Bahare M Khorsandi, Federico Tonini, and Carla Raffaelli. "Design methodologies and algorithms for survivable C-RAN". In: *2018 International Conference on Optical Network Design and Modeling (ONDM)*. IEEE. 2018, pp. 106–111.

[77] Elaine Wong et al. "Enhancing the survivability and power savings of 5G transport networks based on DWDM rings". In: *IEEE/OSA Journal of Optical Communications and Networking* 9.9 (2017), pp. D74–D85.

[78] Anna Pizzinat et al. "Things you should know about fronthaul". In: *Journal of Lightwave Technology* 33.5 (2015), pp. 1077–1083.

[79] Steinar Bjornstad, Dag Roar Hjelme, and Norvald Stol. "A packet-switched hybrid optical network with service guarantees". In: *IEEE Journal on Selected Areas in Communications* 24.8 (2006).

[80]   Raimena Veisllari et al. "Field-trial demonstration of cost efficient sub-wavelength service through integrated packet/circuit hybrid network". In: *Journal of Optical Communications and Networking* 7.3 (2015), A379–A387.

[81]   Raimena Veisllari, Steinar Bjomstad, and Jan Petter Braute. "Experimental Demonstration of 100 Gb/s Optical Packet Network for Mobile Fronthaul with Load-independent Ultra-low Latency". In: *2017 European Conference on Optical Communication (ECOC)*. IEEE. 2017, pp. 1–3.

[82]   Walter Cerroni and Carla Raffaelli. "Analytical model of quality of service scheduling for optical aggregation in data centers". In: *Photonic Network Communications* 28.3 (2014), pp. 264–275.

[83]   Luca Valcarenghi, Koteswararao Kondepu, and Piero Castoldi. "Analytical and experimental evaluation of CPRI over Ethernet dynamic rate reconfiguration". In: *2016 IEEE International Conference on Communications (ICC)*. IEEE. 2016, pp. 1–6.

[84]   Philippos Assimakopoulos et al. "A Converged Evolved Ethernet Fronthaul for the 5G Era". In: *IEEE Journal on Selected Areas in Communications* 36.11 (2018), pp. 2528–2537.

[85]   Raimena Veisllari, Steinar Bjornstad, and Jan P Braute. "Experimental Demonstration of 100 Gb/s Optical Network Transport and Aggregation for Ethernet Fronthaul with Low and Bounded Delay". In: *2018 Optical Fiber Communications Conference and Exposition (OFC)*. IEEE. 2018, pp. 1–3.

[86]   Divya Chitimalla et al. "5G fronthaul-latency and jitter studies of CPRI over Ethernet". In: *IEEE/OSA Journal of Optical Communications and Networking* 9.2 (2017), pp. 172–182.

[87]   Philippos Assimakopoulos, Mohamad Kenan Al-Hares, and Nathan J Gomes. "Switched ethernet fronthaul architecture for cloud-radio access networks". In: *IEEE/OSA Journal of Optical Communications and Networking* 8.12 (2016), B135–B146.

[88]   AB Ericsson et al. "CPRI Specification V6. 0: Common Public Radio Interface (CPRI); Interface Specification". In: *Section* 2 (2013), pp. 62–70.