Network slicing in optical fronthaul based Cloud Radio Access Network to support future 5G services

By

Huma Fida

Fall 2018-MS(IT-10)-00000276557

Supervisor

Dr. Arsalan Ahmad

Department of Electrical Engineering

A thesis submitted in partial fulfillment of the requirements for the degree of

Masters of Science in Information Technology (MS IT)

In

School of Electrical Engineering and Computer Science,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(April 2022)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Network slicing in optical fronthaul based Cloud Radio Access Network to support future 5G services" written by HUMA ABBASI, (Registration No 00000276557), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Advisor: Dr. Arsalan Ahmad

Date: 14-Oct-2021

Signature (HOD): ______________________

Date: ______________________

Signature (Dean/Principal): ______________________

Date: ______________________

Approval

It is certified that the contents and form of the thesis entitled "Network slicing in optical fronthaul based Cloud Radio Access Network to support future 5G services" submitted by HUMA ABBASI have been found satisfactory for the requirement of the degree

Advisor : Dr. Arsalan Ahmad

Signature: ____________

Date: 14-Oct-2021

Committee Member 1:Dr. Asad Waqar Malik

Signature: ____________

Date: 08-Oct-2021

Committee Member 2:Prof. Hasan Ali Khattak

Signature: ____________

Date: 09-Oct-2021

Committee Member 3:Dr. Umair Hashmi

Signature: ____________

Date: 09-Oct-2021

Dedication

I dedicate this thesis to my family and friends. A special feeling of gratitude to my loving parents, Mr and Mrs Fida Abbasi,my in-Laws Mr and Mrs Aqeel Ahmed Butt whose words of encouragement and push for tenacity ring in my ears, my biggest support,my husband Hashim Ahmed Butt who encouraged me to pursue my dreams and finish my dissertation and always told me that i can do it .my supervisor and all the teachers of SEECS who taught and supported me.

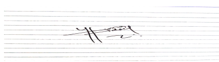
And as Hellen keller says,

"Alone we can do so little; together we can do so much."

Certificate of Originality

I hereby declare that this submission titled "Network slicing in optical fronthaul based Cloud Radio Access Network to support future 5G services" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name: HUMA ABBASI

Student Signature:  ________________

Acknowledgment

I would like to thank Allah Almighty who gave me the courage, determination, strength, motivation and above all the ability to complete this research-based Master's thesis. It is my strong belief that nothing could have been possible without His guidance and blessings.

I would like to thank my parents for their love, care and their financial and moral support.

I would sincerely express my gratitude to my first supervisor Dr. Arsalan Ahmad from SEECS, NUST, Pakistan, who helped with full determination to complete this research work. I have been extremely fortunate to work under his worthy supervision. I believe that his guidance in choosing the topic, timely feedback, and constructive comments on every step kept me going throughout this journey. Most importantly, he has set an unforgettable example for me to pursue my career in the road of good researchers.

I would like to pay a bundle of thanks to my co-supervisors Dr. Umair hashmi from NUST University, who guided me throughout my thesis and who gave me valuable suggestions and directions to move on in this research work.

I would also like to pay my gratitude to my co-supervisors Dr. Asad Waqar and Dr. Hassan Ali Khattak from NUST University,Islamabad,for giving me their valuable suggestion throughout my thesis. It was a huge pleasure for me to complete the second phase of my thesis work under the supervision of them.

Last but not the least, I would like to acknowledge the support and wonderful collaboration of Dr Ashfaq Ahmed from Khalifa University ,Abu Dhabi, for their valuable guidance and for his patient support through out my work.

Table of Contents

List of Abbreviations

Abbreviations

5G	fifth generation mobile access
BBU	Baseband Unit
RRH	Remote radio Head
CS	cell site
CO	central office
CU	Central Unit
CPRI	Common Public Radio Interface
DL	Downlink
ILP	Integer Linear Programming
RU	Radio Unit
UL	Uplink
WDM	Wavelength Division Multiplexing

List of Figures

List of Tables

Abstract

A huge traffic growth is expected in future due to an increase in the number of devices connection to the mobile access network. To overcome this challenge fifth generation (5G) mobile access will make radical changes to the current network. 5G will adopt advanced techniques such as coordinated multipoint (CoMP), small cell deployment, massive MIMO and centralized radio access network (C-RAN). Though C-RAN provides cost saving and energy efficiency as it involves the consolidation of BBUs at a common site but it does not consider scenarios such as uRLLC and mMTC . To address this challenge We propose a new CF-RAN architecture which takes advantage of the fog computing paradigm. CF-RAN uses network function virtualization where baseband processing can be migrated between the BBU hotel and local fog node that offers same functionalities as the BBU. In this manner, the fronthaul capacity issue could be resolved and low latency applications can be performed more efficiently by activating fog node and bringing functions closer to the user. We will device techniques that will be able to support different network slices i.e URRLC, mMTC and emBB on the same infrastructure using CF-RAN architecture and see its impact on the overall performance of the network.

CHAPTER 1

Introduction

To cater the huge demand of mobile networks, continuous improvement in network architecture needs to be considered. The mobile network demands have increased tremendously resulted in enhanced network architecture which will handle all types of services,5G seems to be an absolute promising solution for all types of services. To provide mobile network high reliability, low latency, and fast data transmission its high time to redesign network architecture as traditional architecture is not sufficient to handle all types of requirements. In 3G the base station was divided into two segments RRH and BBU, all the processing functions must be done on BBU, and all the radio functions were done in RRH, RRH placed closer to antenna and CS, to connect RRH to BBU fronthaul link was introduced [1]. The traffic from RRH to BBU was sent over CPRI or OBSAI protocol using fronthaul link.However, in 4G network the concept of centralization was introduced to increase the capacity in terms of cost and energy, all BBU are separated from RRH and placed at a one central location called BBU pool and all RRH are placed closer to antenna [2]. The notion of shared processing was introduced by BBU-pool virtualization; after which, now feasible to share available resources across many locations and distribute extra processing efforts when needed in different regions this architecture is called CRAN [3]. However, Despite CRAN may greatly reduce network costs, it may face into scalability issues if all baseband processing is centralized in a single cloud location, since satisfying CPRI bandwidth and latency requirements becomes problematic [4]. Moreover, CRAN doesn't handle all types of traffic it is "one size fits all" all approach making it impossible to satisfy the vast range of QoS requirements.it doesn't handle other scenarios such as massive machine-type communications massive machine type communication (mMTC) and ultrareliable low-latency communications (uRRLC) [5]. Furthermore, a fully centralized design places extremely high capacity demands on the MFH connections [6]. The mMTC services support tons of low cost devices with relatively relaxed data rates [7]. Therefore, persistent efforts are essentially required to improve the network architecture to cope with the huge traffic demand and to support these heterogeneous use cases. Cloud radio access network (RAN) (C-RAN) is perceived as a revolution in the cellular communication. In C-RAN architecture the baseband units (BBUs) are segregated from respective remote radio heads (RRHs), and are relocated to a centralized location in the form of BBU hotels [8]. The RRHs are kept at the cell sites (CSs). Traditionally, C-RAN is inherited in the fifth generation (5G) networks to meet the increasing traffic demand and the delays [9]. The number of functions performed locally at the CS and centrally at the BBU hotel depend on the employed functional split option [10]. This centralization has revealed encouraging performance in terms of cost saving, energy efficiency, and resource allocation. On the contrary, it is overburdened due to enormous fronthaul traffic, transported between the

RRH and the BBU hotel [11]. The fronhaul traffic is transported using common public radio interface (CPRI) protocol [12], using split option 8 defined by third generation partner project (3GPP) [13], which is exceedingly data-demanding. Moreover, the C-RAN one size fits all approach makes it impossible to satisfy the requirements of wide range of services, including ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and mMTC effectively [14].

To efficiently accommodate all request types and to address the challenge of huge fronthaul bandwidth demand, we propose a novel cloud fog RAN (CF-RAN) over wavelength division multiplexing (WDM) architecture in this work. In CF-RAN over WDM, the baseband processing is replicated at the fog node through network function virtualization (NFV) [15]. These fog nodes are deployed nearer to the RRH to support the delay-sensitive URLLC applications, and to offer same functionalities as the BBUs [16]. After the fog node activation, the traffic between the RRH and the fog node is termed as fronthaul, whereas traffic between the fog node and the core central office (CO), and between the BBU hotel and the core CO is the termed as the backhaul traffic. For the case of eMBB and mMTC requests, the fronthaul is carried between the RRH and the BBU hotel, as they have relatively less stringent fronthaul latency requirements. The backhaul traffic is transported from the BBU hotel to the core CO. For URLLC requests, the fronthaul is carried between the RRH and the fog node, whereas backhaul traffic is transported from the fog node to core CO. The fronthaul traffic is carried through the enhanced CPRI (eCPRI) using split option 7 [17], which has relaxed data rate requirements as compared to split option 8 and still provides a high level of functions centralization. Accordingly, the fronthaul capacity and low latency issues for the URLLC services are efficiently improved by activating fog nodes and bringing cloud functions closer to the user. Moreover, CF-RAN over WDM is a cost economical solution as lesser number of BBUs hotels are activated as compared to 2-layer architecture. The processing of URLLC traffic is done at the fog nodes, which is cost-effective as the nodes are equipped with limited processing capabilities compared to the BBU hotels.

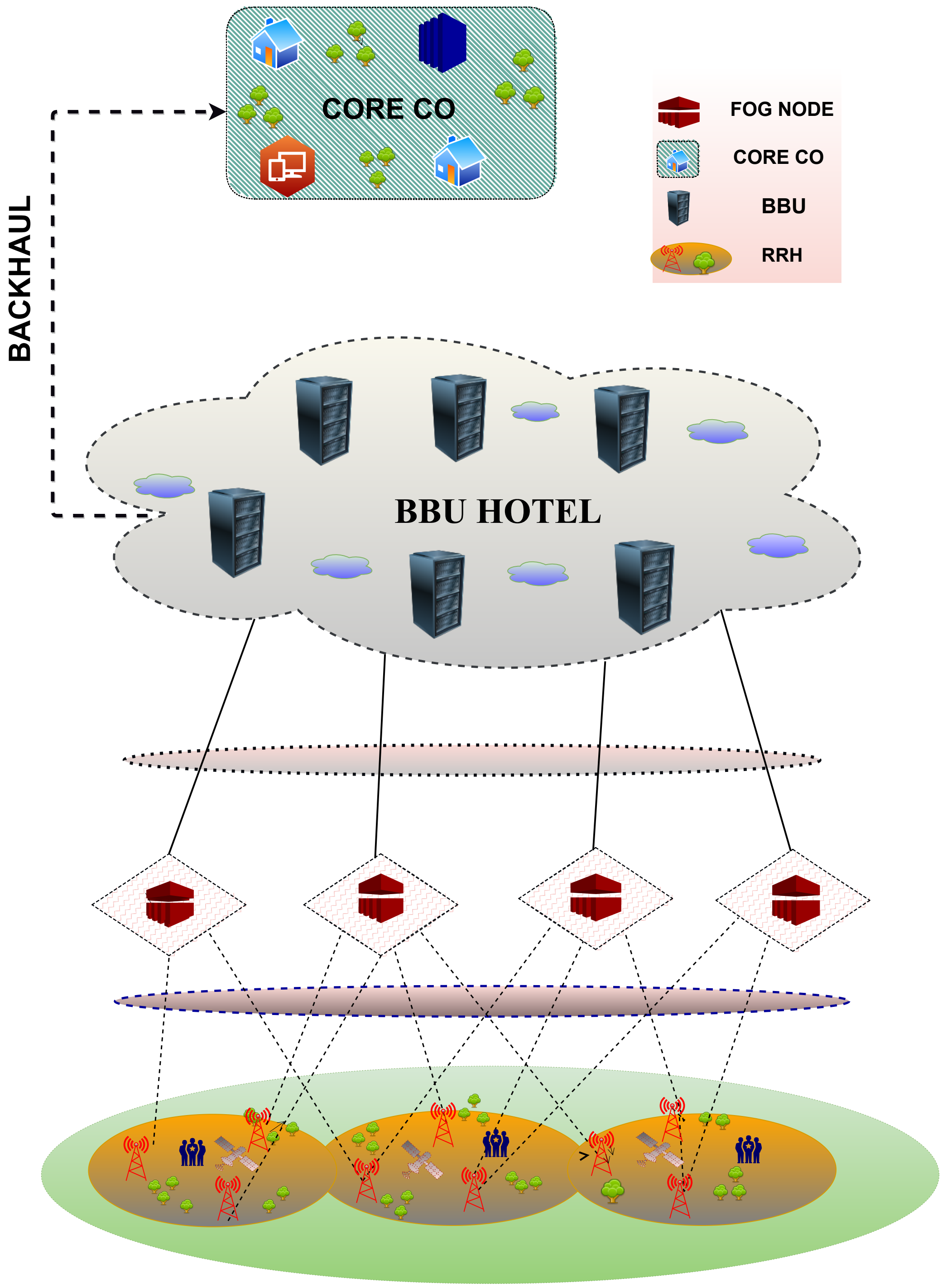


Figure 1.1: The proposed cloud fog radio access network (RAN) (CF-RAN) over wavelength division multiplexing (WDM) architecture.

CHAPTER 2

Literature Review

This chapter focuses on the previous work based on functional split,Network slicing, C-RAN architecture and WDM architecture techniques.

A significant amount of study has been done in recent years on C-RAN. A comprehensive survey on C-RAN has been provided in [18]. BBU placement problem in C-RAN is investigated in [19, 20]. An integer linear programming (ILP) based mathematical framework is proposed to evaluate the impact of fronthaul latency on the BBU centralization. To address the limitations of the C-RAN architecture in terms of scalibity, a CF-RAN architecture over time and wavelength-division multiplexing (TWDM)-passive optical networks (PON) is proposed and investigated in [21]. The proposed architecture's effectiveness is evaluated in terms of power consumption. The migration of cloud resources between the virtual BBUs and fog nodes in CF-RAN over virtualized PON (VPON) is studied in [22]. In [23] authors provide a comprehensive survey on network slicing. In [24] The efficacy of C-RAN being a promising network design. , which enables the virtualization techniques has been proposed. Besides network slicing, it is depicted that the packet based switching on fronthaul can significantly improve the performance, but at the same time C-RAN has strict latency requirements. To deal with the low-latency constraints of URLLC, the possibility of deploying cloud resource at the edge node closer to the user to support mobile fronthaul (MFH) is investigated in [25].To meet the low-latency requirements of URLLC, [25] investigates the feasibility of putting cloud resources closer to the user at the edge node to support mobile fronthaul (MFH).The considered simulation scenario in this work is too compact and unrealistic. From the literature survey, it can be seen that enough research has been conducted on network slicing in 5G networks. The researchers have mainly focused on network slicing in traditional distributed RAN (DRAN) and C-RAN. There is no study that we are aware of that considers the network slicing concept in optical MFH.

2.1 Centralized Radio Access Network

To overcome the expected network congestion 5G mobile access will adopt advanced methods such as (CoMP) coordinated multipoint, massive MIMO, small cell deployment and Centralized Radio Access Network (C-RAN) instead of traditional RAN. C-RAN involves the idea of physically separating the two components of eNodeB namely Remote Radio Head (RRH) and Baseband Unit (BBU) which were previously co-located at the cell site (CS) cabinet

2.1.1 C-RAN Architecture

CRAN architecture as shown in 2.1 , consists of three main elements i)Remote Radio head (RRH) ii) BBU pool iii) Core central office and the links connecting

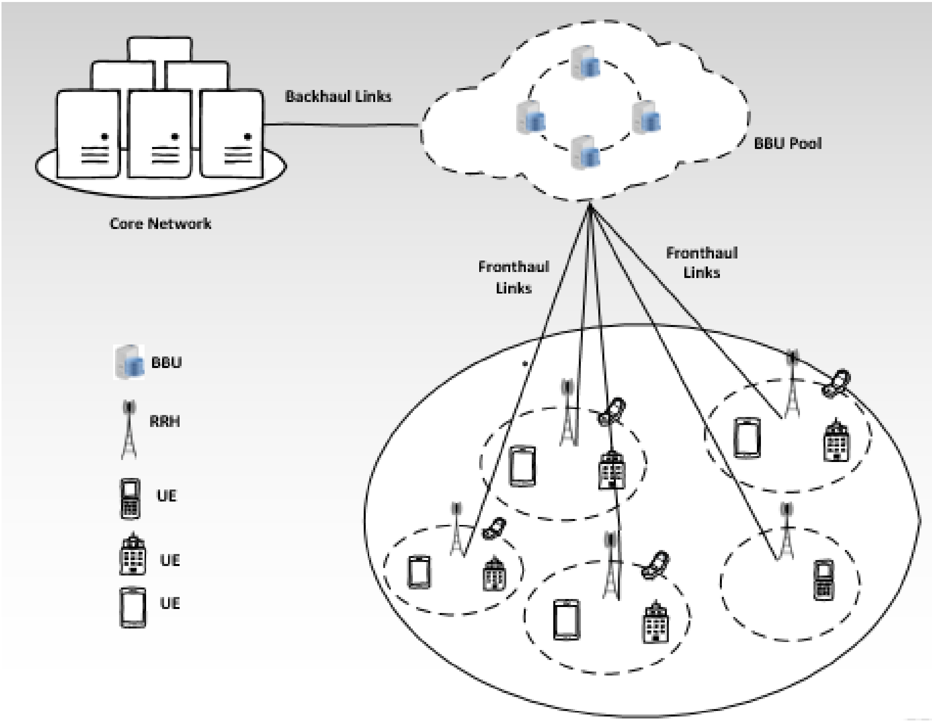


Figure 2.1: Centralized radio access network architecture.

them [26]. It can be seen that in C-RAN the RRHs remain at the CS and are attached to the antenna while the BBUs are separated from the CS and are kept at a centralized location (central office (CO)) [27]. This centralization of BBUs at a common site is known as BBU hotelling [28] where BBUs of different RRHs are kept together in the form of BBU pools as shown in Fig. 2.1. The fronthaul links are the ones that connect the RRH at CS to the BBU hotels. while those which connect the BBU hotel to the core CO are the backhaul links. BBU hotelling promises to be a great solution in term of cost, energy efficiency and overall performance of the network [29]. BBU hotelling helps in reducing the complexity of the base station as only limited functions are performed by the RRH located at CS.

2.2 Problem Statement

Due to the growing number of devices linked to the mobile access network, massive traffic increase is expected in the future..In the recent years, lot of research work has been carried out on C-RAN. A comprehensive survey on C-RAN has been provided in [18].C-RAN is perceived as a revolution in the cellular communication. In C-RAN architecture the BBUs are segregated from respective RRHs, and are relocated to a centralized location in the form of BBU hotels [8]. The RRHs are kept at the CSs. Traditionaly, CRAN is inherited in 5g networks in order to deal with increased traffic load and delays. [9]. The amount of functions,in numbers, executed locally at the CS and centrally at the BBU hotel is determined by the functional split option used. [10]. Cost savings, energy efficiency, and resource allocation have all improved as a result of this centralization.. It is, on the contrary, overburdened as a result of heavy fronthaul traffic between the RRH and the BBU hotel. [11]. The fronthaul traffic is transported using CPRI protocol [12], using split option 8 defined by 3GPP [13], which is exceedingly data-demanding. Moreover, the CRAN one size fits all approach makes it impossible to satisfy the requirements of wide range of services, including URLLC, eMBB, and mMTC effectively [14].

2.3 Proposed Solution

The major goal of this study is to lower the number of active BBus and to reduce fronthaul traffic capacity issue while lowering the network's total cost. We propose novel CF-RAN over WDM network architecture to meet the latency requirements of URLLC services.A mathematical model based on Integer Linear Programming is proposed (ILP). The framework considers the practical network constraints To solve the given mathematical framework, a low-complexity greedy heuristic algorithm is proposed..The performance of the proposed architecture is certified through Montecarlo simulation. The simulations verify the gains of CF-RAN over WDM, as compared to the traditional 2-layer centralization approach. We propose a greedy heuristic approach with the goal of reducing network cost and latency while providing maximum network connectivity. The algorithm also performs the grooming routing and wavelength assignment for the incoming requests while ensuring slice isolation.Finally, we compare the outcomes to the typical 2-layer technique for various network circumstances and traffic occurrences. The major contributions in this work are:

1. Using the CFRAN over WDM architecture, we presented the concept of network slicing, which divides a physical network into multiple configurable slices.

2. We present a mathematical model based on integer linear programming (ILP). The framework takes into account the practical network constraints.

3. To solve the proposed mathematical framework, a low complexity greedy heuristic algorithm is developed.

4. Montecarlo simulation is used to validate the proposed architecture's performance. The simulations demonstrate that the proposed CF-RAN over WDM outperforms the conventional two-layer centralization approach.

2.4 Thesis Layout

Chapter 2 describes the literature review related to some previous work. We discuss the concept of functional split and various techniques for the BBU placement problem. Chapter 3 describes the network architecture and various solutions for BBU placement in a WDM network. In Chapter 4 the proposed techniques are described. Chapter 5 discusses the results obtained through both techniques and compare the performance of heuristic with the ILP. Chapter 6 highlights the conclusion and future work.

CHAPTER 3

Network Architecture

3.1 CF-RAN Architecture

The proposed hierarchical CF-RAN over WDM architecture is shown in Fig. 1.1. In this architecture, the nodes are divided in stages with CSs at stage 1, fog nodes at second stage, access and main COs at 3rd stage, and core CO at 4th stage. An optical fiber links the nodes. Each fiber contains multiple wavelengths, and wavelengths are assigned to each slice to provide isolation for different slice types. Only the traffic of the same type can be groomed into shared light paths (optical channels spanning one or multiple physical links) [30].In CF-RAN over WDM, the baseband processing is replicated at the fog node through NFV [15]. These fog nodes are deployed nearer to the RRH to support the delay-sensitive URLLC applications, and to offer same functionalities as the BBUs [16].

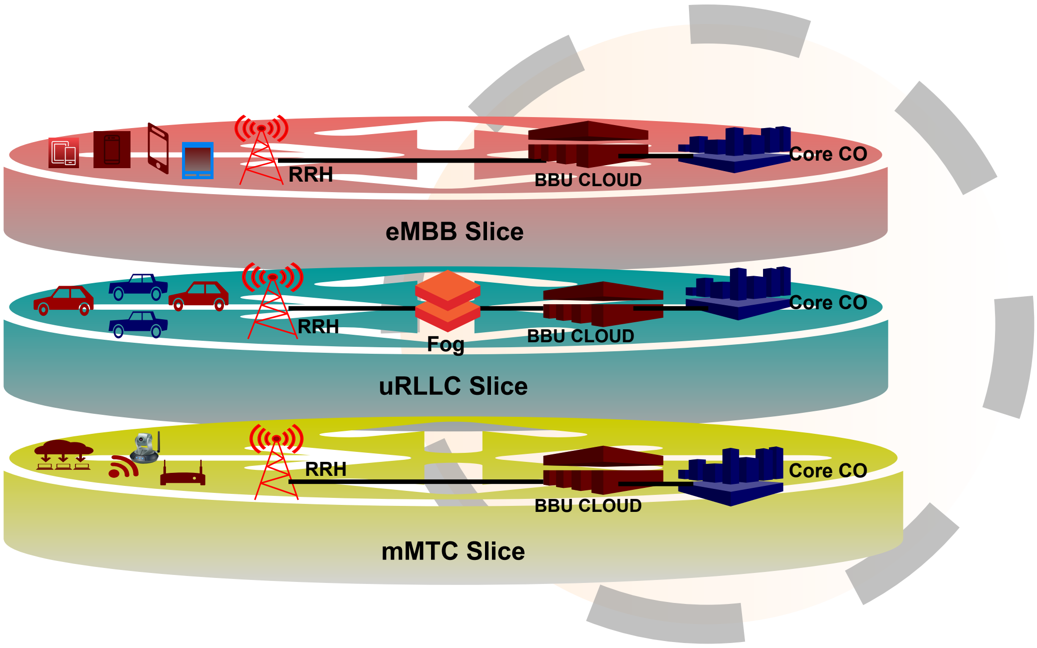


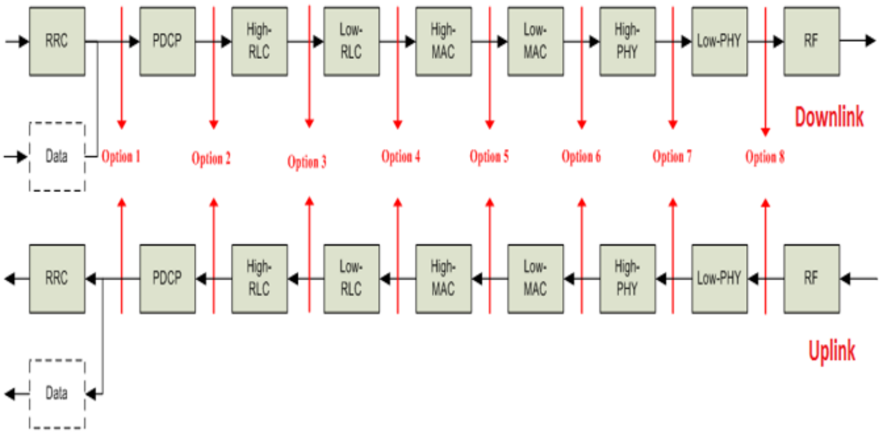
Figure 3.1: Network slicing considering three traffic types.

3.2 Functional Split

Physical separation is involved between the RRHs and the BBUs, in C-RAN. The question arises that which functions should be centralized and performed the central unit (CU) and which ones locally at the CS. The concept of functional split has been introduced to overcome this issue [31]. A total of 8 split option have been defined by 3GPP [32] and are shown in Fig. 3.2. These split points are discussed in detail in [33]. The split option being employed decides the functions that need to be performed at the CS by RRH and those that are needed to be performed at the CU. The split option 8 is used in C-RAN to carry fronthaul traffic that utilizes the CPRI interface. Because of CPRI's high data rate requirements, its deployment is not practicable [34]. To address the fronthaul capacity issue, it may be imperative to investigate a lower-level split point [35]. Split option 7 is the most extensively deployed split option that utilizes the eCPRI protocol for fronthaul traffic transfer [36]. The eCPRI protocol has relatively low bandwidth requirements; also, it provides effective use of available bandwidth than its predecessor, CPRI. Additionally, since it is packet-based, it may be framed within Ethernet [37]. We consider fronthaul and backhaul traffic in our proposed CF-RAN over WDM architecture. Split option 7, the eCPRI split, is used to transport fronthaul traffic. The type 1 traffic's fronthaul is carried between the RRH and the fog nodes. reconfigurable optical add-drop multiplexer (ROADM) [38] is installed on the fog nodes. The fog nodes can use the ROADM to add or remove wavelengths from the transport fiber [39]. It enables type 2 and type 3 requests to pass through fog nodes without being processed. As a result, fog nodes process only type 1 traffic. For type 2 and 3 requests, Fronthaul traffic is transported between RRH and the BBU hotel respectively, while BBU hotel and the core CO handle backhaul traffic. Backhaul is used to connect fog nodes to the core CO in the case of type 1 traffic. It passes through the BBU hotel node unprocessed.

3.2.1 CF-RAN-Split Option 7

Due to the high bitrate requirements of option 8 it is not practical to deploy that split. To address the fronthaul capacity issue lower-level split point should be considered [35]. The split option 7 is now the most widely deployed split option that uses eCPRI protocol for the transport of fronthaul traffic [36]. The eCPRI protocol has relatively relaxed bandwidth requirements, further, it uses the bandwidth more efficiently than its predecessor CPRI. Moreover, it is packet-based, thus it can be framed within the Ethernet [37]. In our proposed CF-RAN over WDM



source:*H. Yu, F. Musumeci, J. Zhang, M. Tornatore, and Y. Ji, "Isolation-aware5G RAN slice mapping over WDM metro-aggregation networks," Journalof Lightwave Technology, vol. 38, no. 6, pp. 1125–1137, 2020.*

Figure 3.2: 3GPP defined split option
[40]

architecture, two types of traffic are being considered: fronthaul and backhaul. The fronthaul traffic, first type of traffic, is transported by split option 7, which is the eCPRI split. The fronthaul for type 1 traffic is carried between the RRH and the fog nodes. The fog nodes are equipped with ROADM [41] which enables the them to add wavelengths or drop wavelengths from the transport fiber [39]. It allows the type 2 and type 3 requests to bypass fog nodes without being processed. Therefore, only type 1 traffic is processed at fog nodes. For type 2 and type 3 requests fronthaul traffic is transported between BBU and the RRH hotel, while the backhaul traffic is carried between the core BBU hotel CO. In the case of type 1 traffic, backhaul is carried between the fog nodes and the core CO. It traverses the BBU hotel node without being processed.

In moving from split option 8 to 1 the required fronthaul bitrate is relaxed in each case but that is achieved at the expense of lower centralization and more complex CS.

3.2.2 Network Slicing

In future 5th generation mobile network will grow explosively, because it provides a lot of services which requires high data rate transfer, low latency, high reliability, and high-capacity demand. As these services demand better capacity, reliability, and latency [42].In Traditional mobile network design, which does not handle all types of services and has a one-size-fits-all approach, must be upgraded [14]. As a result, network virtualization provides a long-term solution known as network slicing, which allows for service variety while providing each service with its own set of resources. Network slicing create several logical networks for a single physical infrastructure, Traditional mobile networks does not provide flexibility of to meet multi-service demands and providing users with various degrees of quality of service (QoS). Network function virtualization (NFV) is used to accomplish network slicing in a 5G network [43]. NFV is critical for network slicing. In [44] the author suggested a new architecture for network slicing,called the NextGen RRC architecture, to fix network slice selection problem and the problems in existing architecture. To gain the most fairness, throughput and QoS performance, the author [45] proposed a flexible network slicing approach for multiuser H-CRANs to handle BBU capacity allocation and power allocation. The authors of [46] applied and analyzed the concept of network slicing to a multi-cell RAN shared by several tenants along with radio resource management (RRM) functions that might find a use in supporting the concept of "splitting the radio resources among the RAN slices". In [6] the author suggested hierarchical edge cloud based, or HEC-based, MFH network-slicing framework to implement networking slicing on a MFH and allocate resources according to the slice demand by using INRM scheme. The concept of network slicing is introduced to support heterogeneous 5G services. Network slicing creates multiple logical networks over the same physical infrastructure using NFV and allocates resources to each slice based on its requirement [47]. We consider 3 types of slices, as shown in Fig. 3.1. The URLLC slice is referred to as type 1, eMBB as type 2, and mMTC as type 3. The existing network-slicing techniques are predominantly focused at dividing core or radio access networks. However, in the 5G optical MFH network, the potential of network slicing should be employed together, to get better resource utilization, in service provisioning, and more flexibility. To support diverse 5G services, the notion of network slicing is introduced. Using NFV, over the same physical infrastructure, network slicing establishes numerous logical networks and provides resources to each slice based on its requirements [**?**]. We consider three different types of slices, as seen in Fig. 3.1. Type 1 is the URLLC slice, type 2 is the eMBB slice, and type 3 is the mMTC slice. Existing network slicing techniques are mostly concerned with the segmentation of core or radio access networks. However, in the 5G optical MFH network, the potential for network slicing must be exploited collectively, so as to increase service provisioning flexibility and resource

3.3 Types of traffic in CF-RAN

Fig. 1.1 show traffic of two types, carried through the network in CF-RAN architecture the fronthaul and backhaul. These traffic types are discussed below.

3.3.1 TYPE 1- Fronthaul Traffic

This is the type of traffic that is introduced due to the separation between the RRH and the BBU and is carried at the CS. The fronthaul has strict latency and capacity requirements, which is unlike backhaul.

In CF-RAN after the fog node activation, the traffic between the RRH and the fog node is termed as fronthaul and is called TYPE 1 request.For URLLC requests, the fronthaul is carried between the RRH and the fog node.The fronthaul traffic is carried through the eCPRI using split option 7 [17], which has relaxed data rate requirements as compared to split option 8 and still provides a high level of functions centralization. Accordingly, the fronthaul capacity and low latency issues for the URLLC services are efficiently improved by activating fog nodes and bringing cloud functions closer to the user.

3.3.2 TYPE 2 & 3 - Fronthaul Traffic

For the case of eMBB and mMTC Type 2 and Type 3 requests, the fronthaul is carried between RRH and BBU, as they have relatively less stringent fronthaul latency requirements.

3.4 Backhaul Traffic

This traffic is traditional type, that is carried betwixt the Core and the CO/CS, which is Packet based and it is delay tolerant to some extent.

Type 1- Backhaul

This is generated by RRH, and is carried between the fog nodes and the Core CO. For URLLC requests, backhaul traffic is transported from the fog node to core CO.

Type 2 & 3 Backhaul

This is the type of traffic that is generated by Cs. For the case of embb and mmtc requests, The backhaul traffic is transported from the bbu hotel to the core co.

CHAPTER 4

System Model Problem formulation and Algorithm

We consider a WDM network topology that includes CS, fog, and CO nodes, and a single core CO. The set of nodes is denoted by $\mathbb{N} = \{\mathbb{N}_c \cup \mathbb{N}_f \cup \mathbb{N}_b\}$, where $\mathbb{N}_c$, $\mathbb{N}_f$, and $\mathbb{N}_b$ represent the set of CS nodes, fog nodes, and the set of COs, including the core CO, respectively. The nodes are connected by a maximum of K fibers per link. The request types are denoted as $t \in \{1,2,3\}$, where $t = 1$ represents URLLC requests, $t = 2$ represents eMBB requests, and $t = 3$ represents mMTC requests. The set of URLLC, eMBB, and mMTC fronthaul requests is denoted by $\mathbb{R}_f^{(1)}$, $\mathbb{R}_f^{(2)}$, and $\mathbb{R}_f^{(2)}$, respectively. If the path delay is less than D_t, where $t = 1$, a URLLC fronthaul request is routed towards the fog node via the set of wavelengths dedicated for type URLLC through the pre-computed path $p \in \mathbb{P}$. This set of wavelengths is denoted by the symbol $\mathbb{W}_t$, where $t = 1$. Total path delay is the sum of propagation delay l_p, interface delay due to deployed split point l_s, and electronic switch delay l_{el}. The fog nodes are equipped with baseband processing capabilities and can handle incoming requests. Only requests of the same type could be groomed into a shared lightpath. The fog node then generates a backhaul request, which is routed to the core CO via the virtual link $v \in \mathbb{V}$. All pathways that establish a virtual link among a source and a destination node are contained in the set $\mathbb{V}$ .If the RRH generates an eMBB or mMTC request, it is routed through path p to the BBU hotel without being processed at the fog node. If the fronthaul latency constraint is satisfied, the request is forwarded over wavelengths dedicated to eMBB or mMTC slice. The request is processed at the BBU hotel, and a backhaul request is initiated, which is carried to the core CO.

4.0.1 Inputs

- $\mathbb{N}$ is a set of nodes, such that $\mathbb{N} = \{\mathbb{N}_c \cup \mathbb{N}_f \cup \mathbb{N}_b\}$. Here $\mathbb{N}_c$ is a set containing the indices of CSs, $\mathbb{N}_f$ contains the indices of fog nodes, and $\mathbb{N}_b$ contains the indices of CO nodes, including the core CO, $\{o\}$ represents the index of core CO. Moreover, N, F, and B, denotes the total number of nodes, the total number of fog nodes, and the total number of BBUs hotels (equal to total number of COs), respectively. The total number of CSs are represented as N_c.

- $\mathbb{N}_{fb} = \{\mathbb{N}_c \cup \mathbb{N}_b\}$ is a set that contains the indices of all fog nodes and COs, including the core CO.

- $\mathbb{P}$ is a set of computed routed paths derived from the topology of all source and destination nodes. Furthermore, $\bar{\mathbb{P}}_{ij} \in \mathbb{P}$ is a set of all possible paths between nodes i and j. The set $\bar{\mathbb{P}}_{i*}$ contains all paths that begin with node i, and the set $\bar{\mathbb{P}}_{*i}$ contains all paths that end with node i. Similarly, $\bar{\mathbb{P}}_i$ is a set of paths that begin or end at node i.

- $\mathbb{V}$ is a set of virtual links. Furthermore, $\bar{\mathbb{V}}_i \in \mathbb{V}$ is a set of all virtual links that begin or end at node i. The set $\bar{\mathbb{V}}_{i*}$ contains all virtual links that begin at node i, and the set $\bar{\mathbb{V}}_{*i}$ contains all virtual links that end at node i.
- The set of physical links is represented by $\mathbb{E}$, with each element denoted by e.
- $\mathbb{T}$ is a set containing the types of requests, where each element of $\mathbb{T}$ is denoted by the symbol t.
- $\mathbb{P}_v$ is a set of paths that constitute a virtual link v.
- $\mathbb{P}_e$ is a set of paths that pass through link e, and P_e are the total number of paths along link e.
- $\mathbb{R}$ is a set of requests, with each element denoted by r. Furthermore, $\mathbb{R} = \{R^t_{fn}, R^t_{bn}\}$, where R^t_{fn} and R^t_{bn} are the sets of all fronthaul and backhaul requests of type t from the n^{th} node. The total number of request is given as R.
- $\mathbb{W}$ is a set of all wavelengths. An element in $\mathbb{W}$ is denoted by λ. Furthermore, the set of wavelengths dedicated to the t^{th} request type is denoted as $\mathbb{W}_t \in \mathbb{W}$. In addition, the capacity of each wavelength is denoted as C.
- The required capacities for fronthaul and backhaul requests of type t are C^t_f and C^t_b.
- The maximum number of fibers that can be placed on each link is K.

- D_t represents the maximum allowable fronthaul latency for a request of type t.

- The delay introduced by each electronic switch is denoted by the symbol l_{el}. Furthermore, the interface delay introduced by the deployed split option is denoted by the symbol l_s. It can also be expressed as length.

- The length of the path p is represented by l_p. It can also be expressed as propagation delay.

4.0.2 Outputs

The outputs of the proposed optimization problem are $\mathbf{b}$, $\bar{\mathbf{F}}$, $\mathbf{x}$, $\mathbf{Z}$, $\mathbf{Y}$, $\bar{\mathbf{Y}}$, $\mathbf{U}$, and $\mathbf{o}$, where

- $\mathbf{b}$ is a binary vector such that $\mathbf{b} \in \mathbb{B}^{1\times B}$, where B is the total number of BBU hotels located at COs. An element in $\mathbf{b}$ is given as b_i, with $b_i = 1$ if a BBU hotel is activated at node i, otherwise $b_i = 0$.

- $\bar{\mathbf{F}}$ denotes a binary matrix such that $\bar{\mathbf{F}} \in \mathbb{B}^{F\times N_c}$, where F and N_c denote the total number of fog nodes, and total number of CSs, respectively. An element in $\bar{\mathbf{F}}$ is represented as $\bar{f}_{f,n}$, with $\bar{f}_{f,n} = 1$ if the processing functions of fog node f are activated for CS n, and $\bar{f}_{f,n} = 0$ otherwise.

- $\mathbf{x}$ is a binary indicator for the activation of fog nodes, such that $\mathbf{x} \in \mathbb{B}^{1\times F}$. An element in $\mathbf{x}$ is denoted as x_f, where $x_f = 1$ if fog node f is activated, otherwise $x_f = 0$.

- $\mathbf{Z}$ is the binary matrix defined as $\mathbf{Z} \in \mathbb{B}^{B\times N_c}$, where B is the total number of BBU hotels, and N_c is the total number of CSs. A $\mathbf{Z}$ element is represented as $z_{i,n}$, with $z_{i,n} = 1$ if the processing functions of the CO i are activated for the RRH n, and $z_{i,n} = 0$ otherwise.

- $\mathbf{Y}$ is a binary matrix of the form $\mathbf{Y} \in \mathbb{B}^{R_b \times V}$, where R_b denotes the total number of backhaul requests and V denotes the total number of virtual links. $y_{r,v}$ represents an element of $\mathbf{Y}$, such that $y_{r,v} = 1$ if a backhaul request r is routed over the virtual link v, otherwise $y_{r,v} = 0$.

- $\bar{\mathbf{Y}}$ is a binary matrix of the form $\bar{\mathbf{Y}} \in \mathbb{B}^{R_f \times P}$, where R_f denotes the total number of fronthaul requests and P denotes the total number of paths. An element of $\bar{\mathbf{Y}}$ is denoted as $\bar{y}_{r,p}$, such that $\bar{y}_{r,p} = 1$, if a fronthaul request r is routed through path p, $\bar{y}_{r,p} = 0$ otherwise.

- $\mathbf{U} \in \mathbb{Z}^{P \times L}$ is a matrix indicating the number of lightpaths with a particular wavelength established along a particular path. The element $u_{p\lambda}$ in $\mathbf{U}$ denotes the number of lightpaths established on path p with wavelength λ.

- $\mathbf{o}$ is an integer vector of the form $\mathbf{o} \in \mathbb{Z}^{1 \times E}$, with E denoting the total number of links. $o_e = \{0, 1, \cdots, K\}$ is an element of $\mathbf{o}$, where K is the maximum number of fibers that can be allocated in a single link.

4.0.3 Constraints

This section discusses the practical network constraints associated with the placement of the BBU hotels in the 5G C-RAN and the implications of this placement.

Fog node association

The RRH at the CS must be connected to single fog node. Mathematically it is given as,

$$\sum_{f=1}^{F} \bar{f}_{f,n} = 1, \ \forall n \in \{1, \cdots, N_c\}. \tag{4.1}$$

BBU hotel association

An RRH must be connected to a single BBU hotel.

$$\sum_{i=1}^{B} z_{i,n} = 1, \ \forall n \in \{1, \cdots, N_c\}. \tag{4.2}$$

Fog nodes activation

A fog node's processing functions are activated only if it serves at least one RRH.

$$\bar{f}_{f,n} \leq x_f, \ \forall f \in \{1, \cdots, F\}, \ n \in \{1, \cdots, N_c\}. \tag{4.3}$$

BBU hotel activation

The processing functions of a BBU hotel at CO are activated only if it serves at-least one RRH.

$$b_i \leq \sum_{n=1}^{N_c} z_{i,n}, \ \forall i \in \{1, \cdots, B\}. \tag{4.4}$$

Fiber deployment

Each link cannot include more fibers than the allowed maximum K.

$$\sum_{p=1}^{P_e} u_{p\lambda} \leq o_e \leq K, \ \forall e \in \{1, \cdots, E\},$$
$$t \in \{1,2,3\}, \lambda \in \mathbb{W}_t. \tag{4.5}$$

The number of fibers on each link is determined by this constraint. The number of fibers on each link must be greater than the number of lightpaths with the same λ that pass through that link, i.e., ($u_{p,\lambda} \leq o_e$). Furthermore, the number of fibers on each link is limited by K.

Fronthaul capacity

The total amount of fronthaul requests routed via a path p should be less than the maximum capacity of the path.

$$\sum_{r \in \mathbb{R}_f^t} \sum_{p \in \mathbb{P}_v} C_t^f \bar{y}_{r,p} \leq \sum_{\lambda \in \mathbb{W}_t} \sum_{p \in \mathbb{P}_v} C u_{p\lambda}, \ \forall v \in \mathbb{V}, t \in \mathbb{T}. \tag{4.6}$$

This constraint assures that the aggregate capacity of all fronthaul requests routed through a path p is less than the path's maximum capacity. When a request is routed via path p belonging to virtual link v, $\bar{y}_{r,p}$ becomes active. The left side of the equation is equal to the sum of all fronthaul requests of type t carried over paths associated with virtual link v. The capacity of a fronthaul request of type t is denoted by C_t^f. If a path utilizes the wavelength λ, then $u_{p,\lambda}$ becomes active. The right hand side of the equation equals the maximum capacity of routes associated with virtual link v, where C is the maximum capacity of a lightpath.

Backhaul capacity

The aggregate of all backhaul requests routed over a virtual link should not exceed the capacity of the virtual link.

$$\sum_{r \in \mathbb{R}_t^b} C_t^b y_{r,v} \leq \sum_{\lambda \in \mathbb{W}_t} \sum_{p \in \mathbb{P}_v} C u_{p\lambda}, \forall v \in \mathbb{V}, t \in \mathbb{T}. \tag{4.7}$$

This constraint assures that the aggregate of all backhaul requests transported over a virtual link v is less than the virtual link's maximum capacity. When a backhaul request is routed via virtual link v, $\bar{y}_{r,v}$ is enabled. The left side of the equation equals the sum of all backhaul requests of type t that are carried over the virtual link v. The capacity of a backhaul request of type t is denoted by C_t^b. If a path associated with virtual link v uses the wavelength λ, then $u_{p,\lambda}$ becomes active. The right hand side of the equation equals the maximum capacity of routes associated with virtual link v, where C is the maximum capacity of a lightpath.

Constraint 8: Fronthaul delay

The delay in fiber propagation l_p, the delay in electronic switches l_{el}, and the interface delay of RRH for the split point l_s should all be less than or equal to the maximum allowable fronthaul latency D_t for request type t.

$$\sum_{p \in \mathbb{P}} (l_p + l_{el} + l_s) \bar{y}_{r,p} \leq D_t, \forall t \in \mathbb{T}, r \in \mathbb{R}_f^t. \tag{4.8}$$

Constraint 9: Backhaul requests routing

This constraint ensures that backhaul requests are routed via virtual links beginning at fog nodes (for type 1 requests) and BBU hotels (for types 2 and 3 requests) and ending at the core CO.

$$\sum_{v \in \mathbb{V}_{*i}} y_{r,v} - \sum_{v \in \mathbb{V}_{i*}} y_{r,v} = \begin{cases} 1 - z_{i,n}, & \text{if } i = o,\ t = \{2,3\} \\ -z_{i,n}, & \text{if } i \in \mathbb{N}_b \setminus o,\ t = \{2,3\} \\ 1, & \text{if } i = o,\ t = \{1\} \\ -\bar{f}_{i,n}, & i \in \mathbb{N}_f,\ t = \{1\} \end{cases} \tag{4.9}$$

$$\forall n \in \mathbb{N}_c, t \in \mathbb{T}, i \in \mathbb{N}_{fb}, r \in \mathbb{R}^t_{bn}.$$

If the BBU hotel of a node is placed in the core CO, the first term on the right hand side becomes 0, indicating that no backhaul traffic is sent. If the BBU hotel for a node is placed in an intermediate CO, backhaul requests are carried out through a virtual link launched from that CO ($-z_{i,n} = -1$) and ended at the core CO ($1 - z_{i,n} = 1$). Backhaul requests of type 1 are transmitted over a virtual link initiated at the fog node (i.e., $-\bar{f}_{i,n} = -1$) and terminating at the core CO (third term on the right hand side = 1).

Constraint 10: Fronthaul requests routing

This constraint ensures that fronthaul requests are routed via lightpaths that originate at RRHs in CSs and terminate at fog nodes (for type 1 requests) and BBU hotels in COs (for type 2 and type 3 requests)

$$\sum_{p \in \mathbb{P}_{*i}} \bar{y}_{r,p} - \sum_{p \in \mathbb{P}_{i*}} \bar{y}_{r,p} = \begin{cases} \bar{f}_{i,n}, & \text{if } i \neq \{n,o\},\ t = \{1\} \\ \bar{z}_{i,n}, & \text{if } i \neq n,\ t = \{2,3\} \\ -1, & \text{if } i = n \end{cases} \tag{4.10}$$

$$\forall n \in \mathbb{N}_c, t \in \mathbb{T}, i \in \mathbb{N}, r \in \mathbb{R}^t_{fn}.$$

Fronthaul requests are launched by RRHs at the CSs and terminate at the fog nodes for type 1 requests, and at COs where the BBUs hotels of the respective RRHs are activated for type 2 and 3 requests. When a type 1 request is carried over a lightpath from the request generating node n to the fog node i, $\bar{f}_{i,n} = 1$ indicates that the request has been terminated at i. Similarly, $z_{i,n} = 1$ indicates that request r has been terminated at the BBU location for node n. On the right hand side of the equation, the third term denotes the generation of requests by the CSs.

4.0.4 Objective Function

The primary aim of this research is to lower total network cost, which is proportional to the number of active BBU hotels and fog nodes, while still conforming to practical network constraints. The authors present a model for multi-objective optimization in which α and $(1-\alpha)$ are the weights of each objective, such that $\alpha \in (0,1)$. The objective function is described as follows:

$$\mathcal{Z} = \alpha \sum_{i \in \mathbb{N}_b} \mathbf{b}_i + (1-\alpha) \sum_{f \in \mathbb{N}_f} \mathbf{x}_f. \tag{4.11}$$

4.0.5 Optimization problem

Consequently, the optimization problem is stated as follows:

$$\text{OP1} : \min \mathcal{Z} \tag{4.12}$$

$$\text{s.t.}$$

$$\text{Constraints } (4.1) - (4.10).$$

The optimization problem in (4.12) is an NP-hard problem. It takes real effort to get an optimal solution for an NP-hard problem. There are solvers that fairly tackle NP-hard problems and produce near-optimal solutions. However, the technique's complexity scales with network size, finding it challenging to get useful results for larger networks in an acceptable timeframe from these solvers. Therefore, we present a low-complexity heuristic algorithm based on a greedy approach. For large-scale network scenarios, the proposed algorithm sufficiently solves the problem with significantly less complexity.

4.1 Proposed Heuristic Service Aware BBU Hotel and Fog Node Activation Algorithm (SB-FAA)

We describe a heuristic approach with low complexity for obtaining a near-optimal solution to the optimization problem (4.12). The proposed method is named service aware BBU hotel and fog node activation algorithm (SB-FAA) and is described in detail in Algo 1. The method minimizes the number of active BBU hotels and fog nodes for all requests and performs grooming routing and wave-

Algorithm 1: Service Aware BBU Hotel and Fog Node Activation Algorithm (SB-FAA)

```
Input: N ← no. of nodes, E all the edges, G(N,E), r ←request={s ← source node, t ←
   type of request}, f_r ← required fronthaul capacity for r, m_r ← required midhaul
   capacity for r, b_r ← required backhaul capacity for r;
Define: z_s ← binary variable that turns on when s is associated with a BBU hotel, P_xy ←
   is path between node x and y, D_t ← required fronthaul latency for type t slice request,
   C_xy ← calculates the residual capacity of path between node x and y, w ← set of active
   BBU hotels, x ← set of active fog nodes, ;
Initialization: w ← φ, x ← φ;
Execution;
if z_s == φ then
    L ← All reachable CO from s;
    for l ∈ L; do
        P_sl ← shortest path between s and l;
        Q ← Shortest Path length for s to l;
        if t = 2 or t = 3 then
            if Q ≤ D_t then
                if C[P_sl] >= f_r and C[P_lo] >= b_r then
                    z_s ← l;
                    if l not in w then
                        w ← l;
                    end
                    Route r on path P_so;
                end
            end
        end
    end
    if z_s == φ or capacity of path is not enough to carry fr and b_r then
        return request-blocked;
    end
end
if t == 1 then
    L ← All reachable COs from s;
    for l ∈ L do
        Q ← Path length between s and l;
        if Q ≤ D_t and l ∈ w then
            if C_sl ≥ f_r and C_lo ≥ b_r then
                Route r on path P_so;
            end
        end
    end
    if no node in L can satisfy r then
        for f ∈ F do
            if s is connected to f and C_sf ≥ f_r and C_fl ≥ m_r and C_fo ≥ b_r then
                if f is not active then
                    f̄_f ← 1;
                    x ← f;
                end
                Route r on path P_so;
            else
                Return request-blocked;
            end
        end
    end
end
Outputs: w and x;
```

length assignment (GRWA) [48]. Additionally, the algorithm aims to minimize blocked requests while enhancing connectivity. The request generating node s and its type t are used to characterize an incoming request r. If r represents an eMBB or mMTC request, the algorithm determines whether s is associated with an active BBU hotel. If such an association does not exist, the algorithm determines all COs that are reachable from node s. The shortest path is computed between node s and candidate CO, and the path length is calculated using the Dijkstra algorithm [49], (**step 1-7**). The path length is used to calculate the required latency, which includes interface delay l_s at RRH due to the employed split as well as delay due to electronic switches l_{el}. If the required latency is less than the maximum allowable fronthaul latency, the residual capacity of the path is evaluated. If the path has enough capacity to transport both C_t^f and C_t^b, then s is associated with the candidate CO. If candidate CO does not already have an active BBU hotel, it is activated and r is routed through the computed path (**steps 8-12**). If no node is capable of hosting BBU for the request-generating node s, the request is blocked (**steps 13-14**). The algorithm assesses whether s is associated with an active BBU hotel upon receipt of request r. If such an association exists, the algorithm determines location of BBU hotel for s using the $find$ function. The algorithm then determines the residual capacity in the links of the determined path. If the capacity is sufficient to transport both C_t^f and C_t^b, r will be transported via that path; otherwise, the request will be blocked (**steps 16–20**).

If the request r is of type 1, the algorithm determines whether it can be processed at an active CO. The distance between s and active COs is calculated. The path delay is calculated using the calculated path length. The path is used to route r if its latency is less than the maximum allowable fronthaul latency D_t and the capacity of the path's links is sufficient to carry r (**steps 22-27**). If no active CO satisfies r, the algorithm calculates the residual capacity in the link between s and the fog node f, as well as the link between f and o (**steps 28-31**). If sufficient capacity is available, the fog node that s is connected to is activated if it was not active already, and r is routed along the computed path (**steps 32-33**). r is blocked if f is unable to satisfy r (**steps 34-35**).

4.1.1 Computational Complexity

The complexity of the algorithm is frequently expressed in floating point operations (FLOPS). Due to the fact that FLOPS are machine-independent, they provide a fair method for conducting a precise and reliable complexity analysis.

The proposed algorithm can be divided into 2 major blocks, namely block 1 (**steps 1-20**) and block 2 (**steps 21-36**). Given the fact that the blocks are mutually exclusive, they can never run concurrently. The first block contains a *for* loop that executes up to B times, where B is total number of nodes equipped with BBU hotel. The shortest path is determined within the loop using Djisktra's algorithm, which has a worst-case complexity of N^2, where N is the total number of nodes. There are approximately ten assignments within the loop, and assuming one FLOPS for each assignment, there will be a total of ten FLOPS. The block from **steps 16-20** contains simple assignments and requires a total of 5 FLOPS. As a result, the complexity of block 1 in the worst-case scenario is $\approx (10N^2)B$. There is a *for* loop in block 2 from **steps 22-27**. The loop executes a maximum of B times. There are approximately eight assignments within the loop. As a result, the complexity of this *for* loop is approximately $8B$. At **steps 29-36**, the *for* loop executes F times, where F is the total number of fog nodes. There are seven assignments contained within the loop. As a result, this loop's complexity is $7F$. Block 2's total complexity will be $\approx 8B+7F$. Finally, the proposed algorithm's worst-case complexity is specified as $\approx R(\max\{(10N^2)B, 8B+7F\})$, where R is the total number of requests.

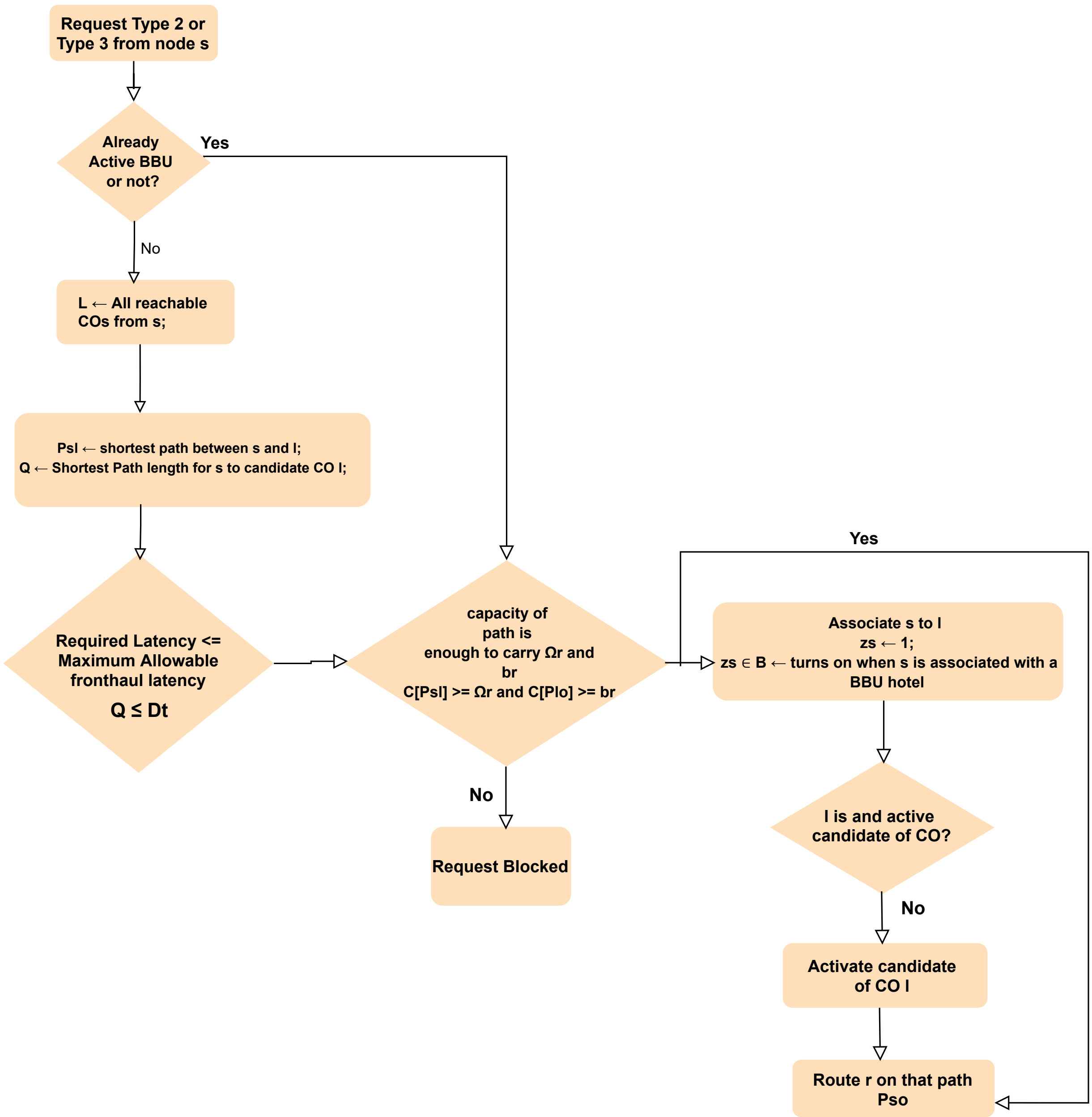


Figure 4.1: flow chart -SBFAA with Request type 2 & 3

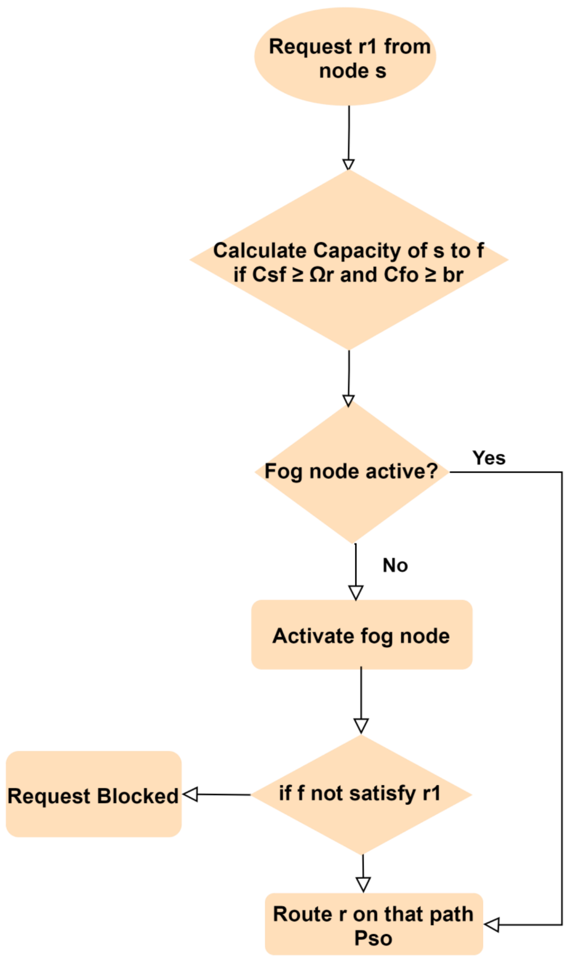
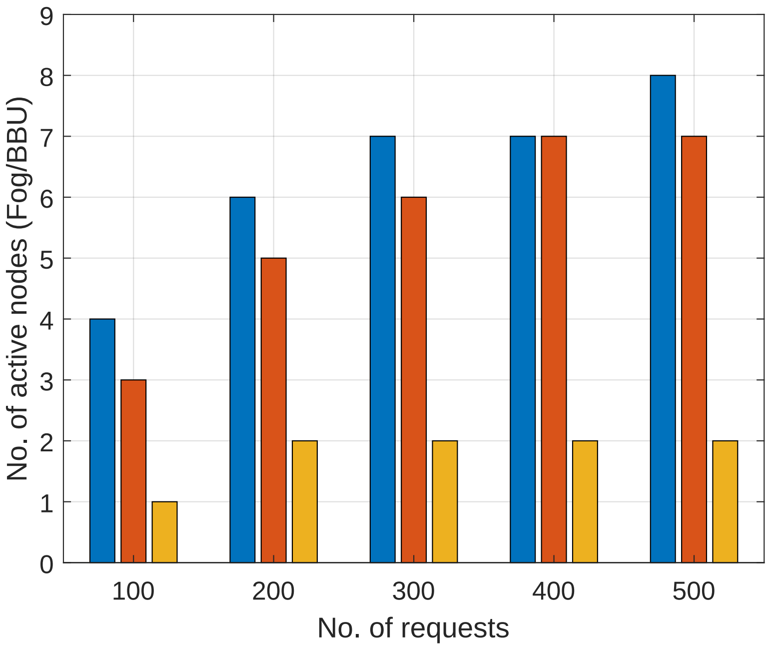
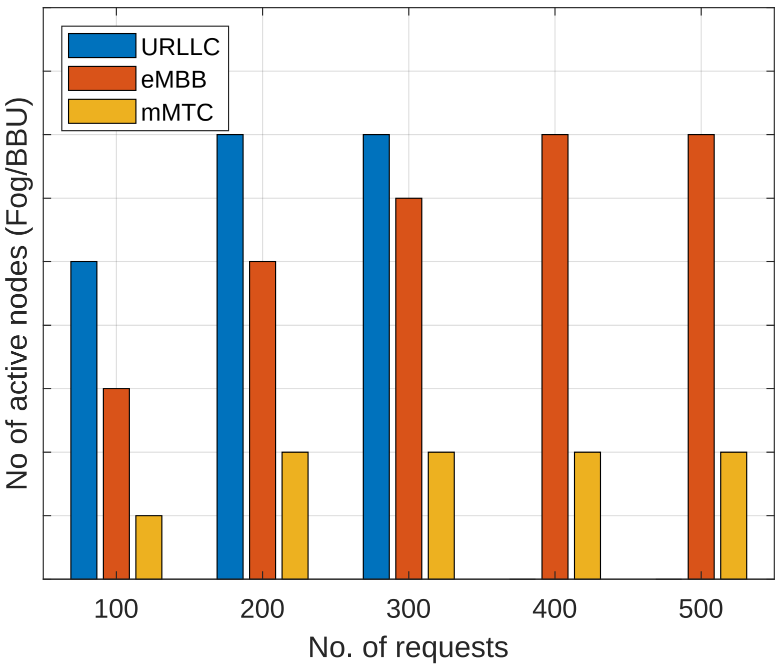


Figure 4.2: flow chart -SBFAA with Request type 1



(a) 3-layer architecture.



(b) 2-layer architecture.

(c) Number of active nodes for ILP.

CHAPTER 5

Numerical Evaluation and Results Analysis

The Results obtained by implementing the proposed techniques are discussed in this chapter. The optimization software AMPL+CPLEX was used to do the simulations for ILP. The heuristic scenarios were simulated out with Pycharm 3.8 and the Networkx package, using the Python programming language. We obtained results for various network scenarios and traffic instances and averaged our results accordingly.

5.0.1 Simulation parameters

To solve the proposed ILP model in Section 4, we assume a network consisting of 30 nodes seperated into 13 macro CSs, 7 COs including a core CO, and 10 fog nodes. These nodes are evenly distributed in a geographical area of 50 km^2. Moreover, the nodes are connected to each other via optical fiber links, with a maximum of 1 fiber per link, i.e., $K = 1$. Each fiber link supports 8 wavelengths

Table 5.1: Latency requirement of front/backhaul for each slice.

Request	Type	Fronthaul Latency	Backhaul Latency
URLLC	1	$\leq 50\mu s$	$\leq 500\mu s$
eMBB	2	$\leq 100\mu s$	$\leq 20ms$
mMTC	3	$\leq 250\mu s$	$\leq 20ms$

at 10 Gbps each. In order to provide wavelength isolation to each slice type, 4 wavelengths are dedicated for eMBB requests, and 2 each for URLLC and mMTC requests. More wavelengths are dedicated to the eMBB requests as they seek higher capacity. We generate a total of 500 traffic requests for each slice type. These requests are uniformly distributed among all the CSs. While we consider split option 7 for transport of fronthaul traffic, the required fronthaul capacity for each URLLC, eMBB, and mMTC request is 120 Mbps, 720 Mbps, and 80 Mbps, respectively [25]. The required backhaul for URLLC and mMTC requests is uniformly distributed between 10-20 Mbps, whereas for eMBB request required backhaul is 240 Mbps. The required fronthaul and backhaul latencies for each request type is shown in Table 5.2. The fronthaul latency for URLLC, eMBB, and mMTC requests are less than 50μs, 100μs, and 250μs, respectively. The delay due to an electronic switch is 20μs, while the interface delay at RRH for split option 7 is 25μs [50].

5.0.2 Simulation Results

The proposed three-layer CF-RAN over WDM architecture is evaluated in terms of active nodes, fronthaul delay, and total number of requests served. The optimization goal is to minimize the number of active BBU hotels and fog nodes in

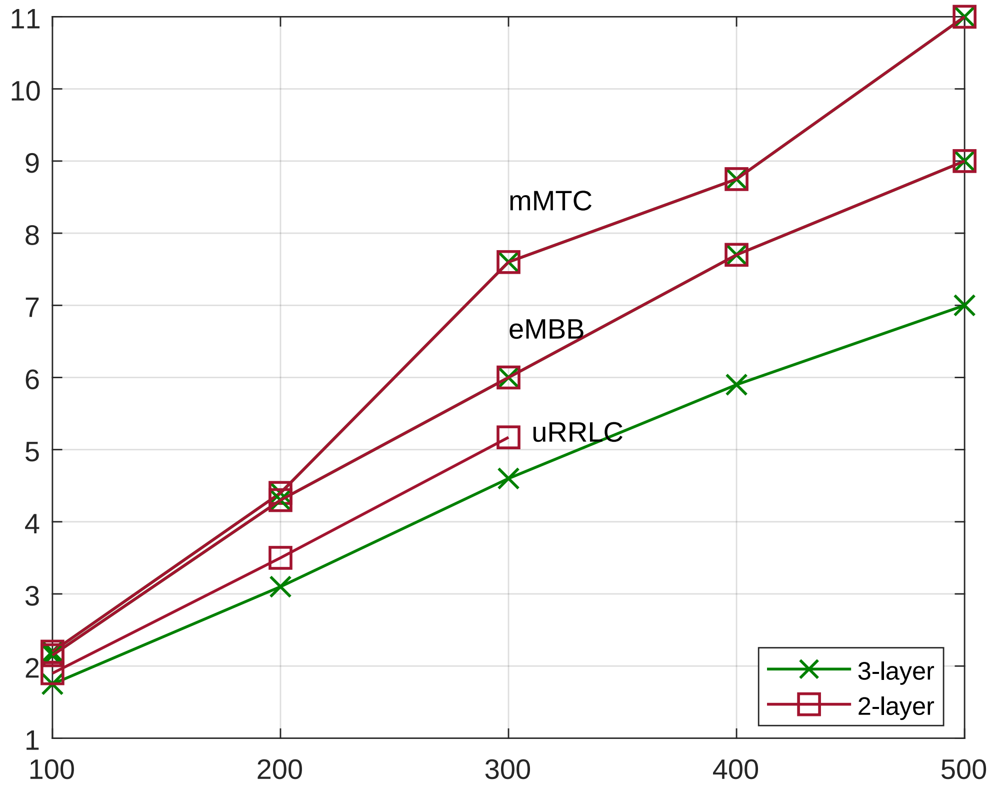


Figure 5.1: Comparison between 2-Layer and 3-Layer in terms of total fronthaul delay for ILP.

the network while adhering to practical network constraints.

In terms of active nodes, fronthaul delay, and total number of requests served, the proposed three-layer cfran over wdm architecture is assessed for three slice types, as shown in Figs. 4.3a and 4.3c. When can be observed, as traffic grows, both approaches result in an increase in the total number of active nodes. The proposed architecture behaves identically to the two-layer architecture in terms of eMBB and mMTC requests, as both approaches serve these requests via BBU hotels. The URLLC slice type results in a significant reduction in activated nodes because the same amount of traffic can be routed through fewer activated fog nodes. As traffic increases, the 2-layer approach no longer provides a feasible solution. On the other hand, CF-RAN over WDM can provide a viable solution for all generated requests without activating all available fog nodes.

The comparison of the proposed three-layer design to the two-layer architecture for three slice types in terms of total fronthaul delay is shown in Fig. 5.1. As illustrated in Fig. 5.1, the total fronthaul delay for each slice type lowers with CF-RAN over WDM architecture in comparison to the 2-layer architecture. The fronthaul delay is nearly identical for the mMTC and eMBB slice types. The 3-layer technique achieves a significant improvement for the URLLC slice type. Additionally, the two-layer architecture is unable to provide a feasible solution under huge traffic loads. To comply with the delay constraints, URLLC requests are always executed at nearby fog nodes. Due to the large distance between the CSs and the BBU nodes, significant fronthaul delay is induced, making it hard to execute URLLC requests at the BBU nodes.

5.0.3 Performance Evaluation of Proposed SB-FAA

To analyze the effectivness of the heuristic approach, we consider a network that consists of 140 nodes seperated into 70 Macro CSs, 30 COs including a core CO, and 40 fog nodes. These nodes are evenly distributed over a 200 km^2 region. Furthermore, the nodes are linked by optical fiber links, with a maximum of one fiber per link. Each fiber link is able to support 16 wavelengths at 40 Gbps each.

To provide wavelength isolation for each slice type, eight wavelengths are reserved for eMBB requests and four for URLLC and mMTC requests, respectively. A Poisson distribution is used to create 5000 traffic requests at random.

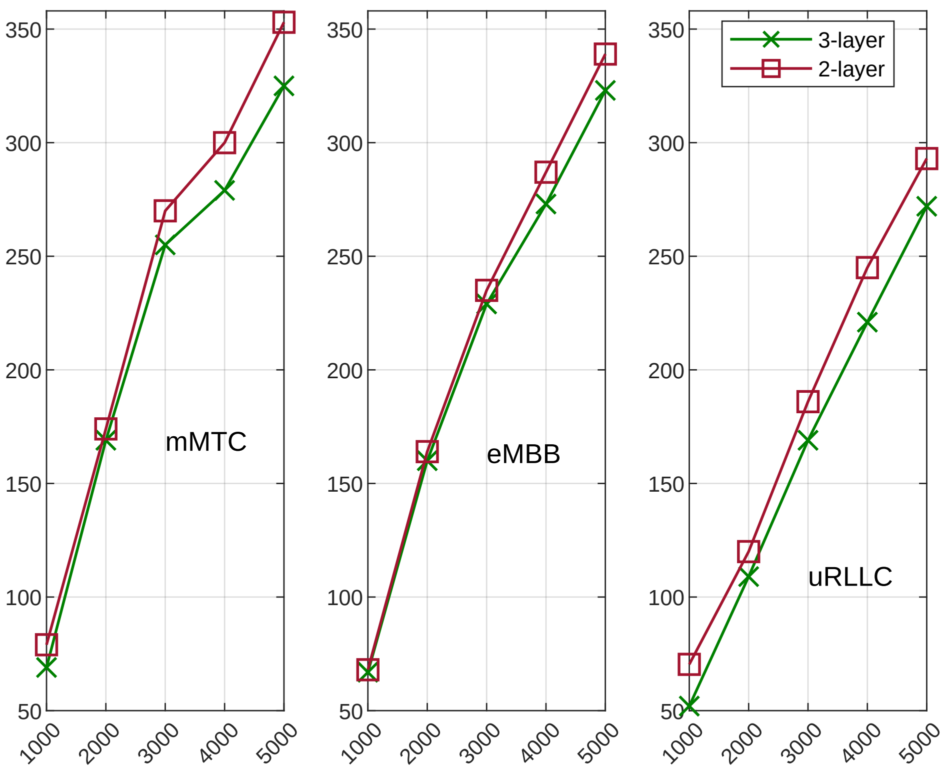


Figure 5.2: Comparison between 2-layer and 3-layer architecture in terms of total fronthaul delay for heuristic algorithm.

These 5000 requests are distributed evenly among all RRHs. Each CO contains a BBU hotel that is activated upon the receipt of a request. Similarly, each fog node is able to process baseband traffic and is triggered in response to the receipt of a type 1 request. Montecarlo simulation is performed to average the results of random simulations.

Figure 5.2 analyzes the effectiveness of the proposed 3-layer architecture to the 2-layer architecture for three slice types in terms of fronthaul delay. The total fronthaul delay for each slice type decreases with the implementation of CF-RAN

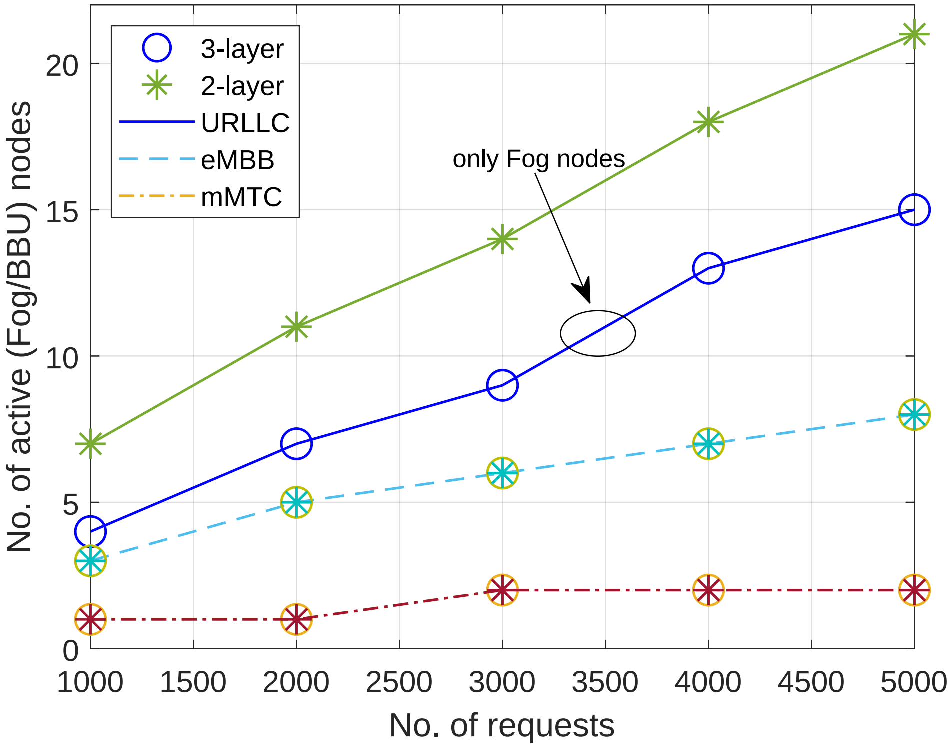


Figure 5.3: Comparison between 2-layer and 3-layer architecture in terms of number of active nodes for heuristic algorithm.

over WDM, as shown in Fig. 5.2, when compared to the 2-layer architecture. As with ILP, the most significant improvement in terms of fronthaul delay can be observed for the URLLC slice type, as type 1 requests are catered at fog nodes. 5.3 ddepicts the total number of active nodes with respect to total requests for both 3-layer and 2-layer architectures. As illustrated in Fig. 5.3,the number of active nodes increase for all cases with the increase in number of requests for both 2-layer and 3-layer architecture. Additionally,In case of eMBB and mMTC slices the number of active BBU hotels are same for both approaches. In both architectures the eMBB and mMTC requests are satisfied by BBU hotels, so there

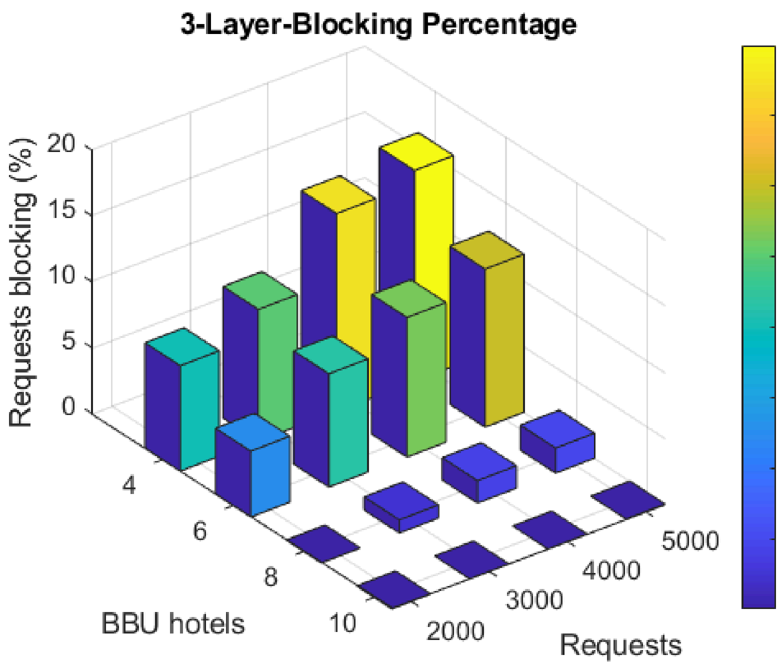


Figure 5.4: Request blockage rate for 3-Layer approach.

is no difference in required number of BBU hotels. For uRLLC requests, there exist a significant difference in the number of active BBU hotels. The reason is that in our proposed architecture uRLLC requests are mostly catered by fog nodes resulting in a notable decrease in the number of active BBU hotels.unless the request generating node is connected to an already-active BBU hotel. The request generating node must be connected to the active BBU node through a path that satisfies the required latency constraint in this case.

Figure 5.4 shows the request blockage percentage for the proposed 3-layer CFRAN over WDM architecture as the number of active BBU hotels and traffic grows. When there are fewer BBU hotels active in the network, request blocking

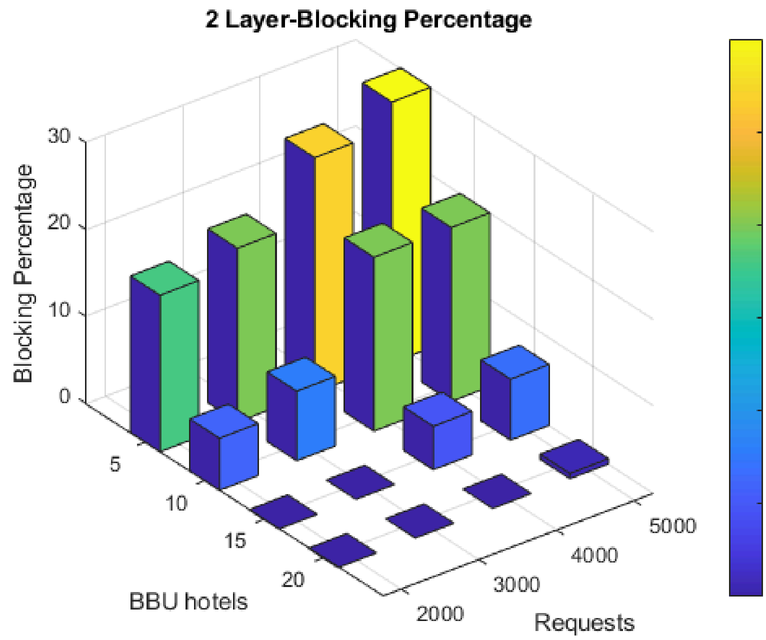


Figure 5.5: Request blockage rate for 2-Layer approach.

is increased, especially for heavier traffic. The percentage of requests blocked decreases as the number of active BBU hotels grows, and no requests are blocked as the number of active BBU hotels grows.

Figure 5.5 shows the request blockage percentage for the proposed 2-layer CRAN architecture as the number of active bbu hotels and traffic grows. Request blockage rate ibcreases as the number of requests increases in 2-layer archtecture.As the traffic is further increased, the 2-layer approach fails to provide a feasible solution. On the contrary, CF-RAN over WDM can provide feasible solution for all generated requests without activating all available fog nodes.

Another important metric for validating the proposed architecture is the percentage of served requests. Figures 5.6 and 5.7 compare the proposed 3-layer

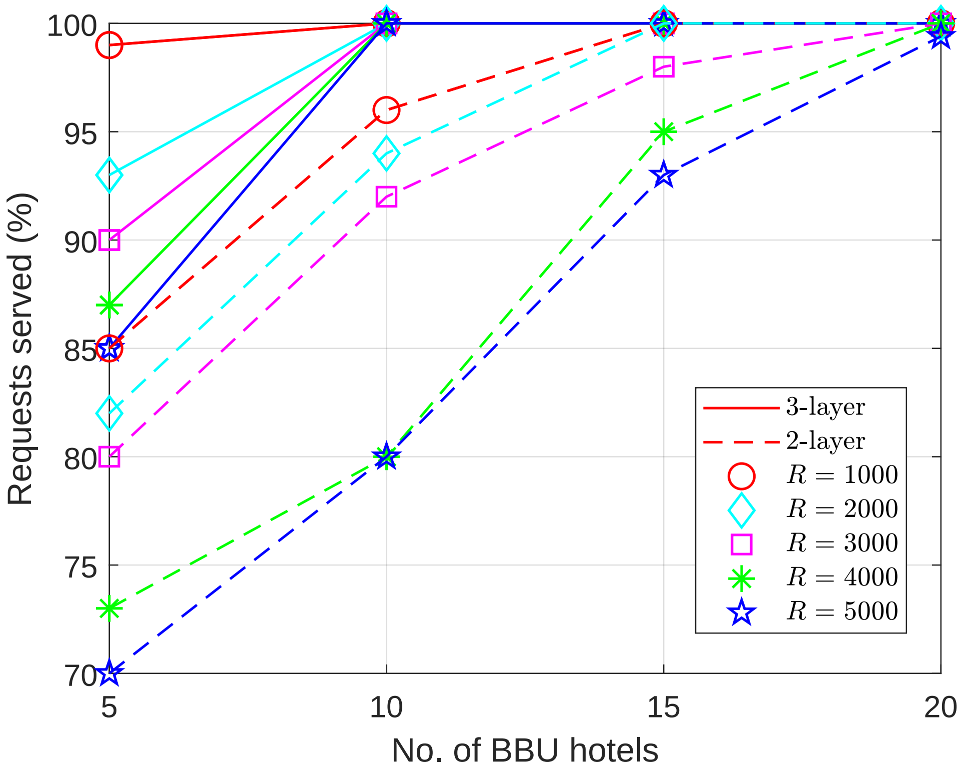


Figure 5.6: Comparison of 3- and 2-layer approaches in terms of served requests Vs. the total number of BBU hotels.

architecture to the 2-layer approach in terms of the percentage of requests served. In both cases, we set the same number of active BBU hotels represented by B in Fig. 5.7. As can be seen in both figures, an increase in the number of requests reduces the percentage of requests served when the number of active BBU hotels is lower in both cases. The request serving percentage increases in both cases as we increase the number of active BBU hotels. However, the proposed CF-RAN over WDM provides greater network connectivity than the 2-layer approach with significantly fewer active BBU hotels.

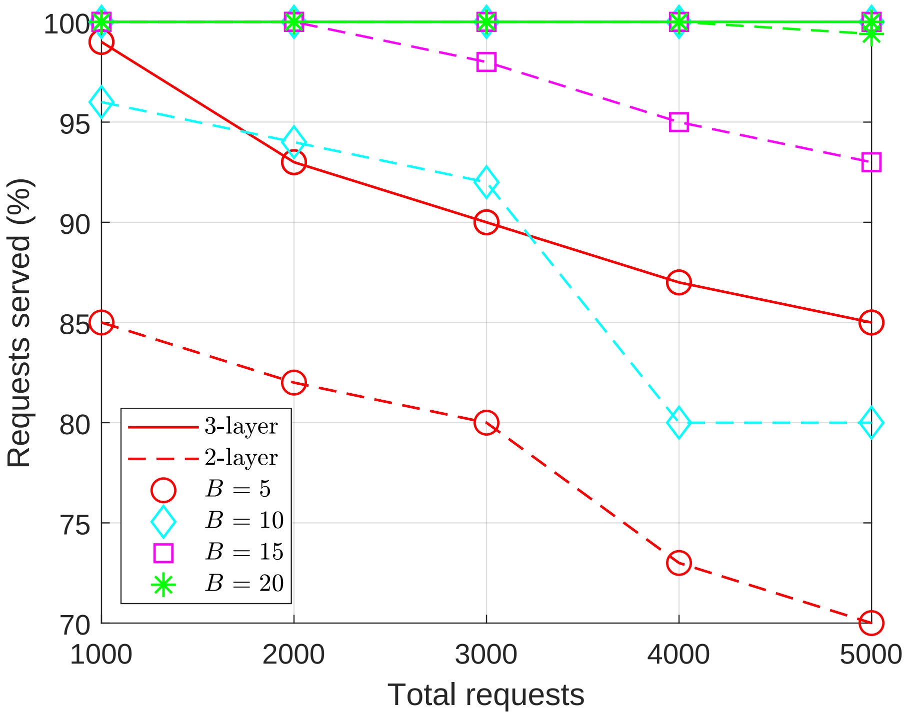


Figure 5.7: Comparison of 3- and 2-layer approaches in terms of served requests Vs. the total number of requests.

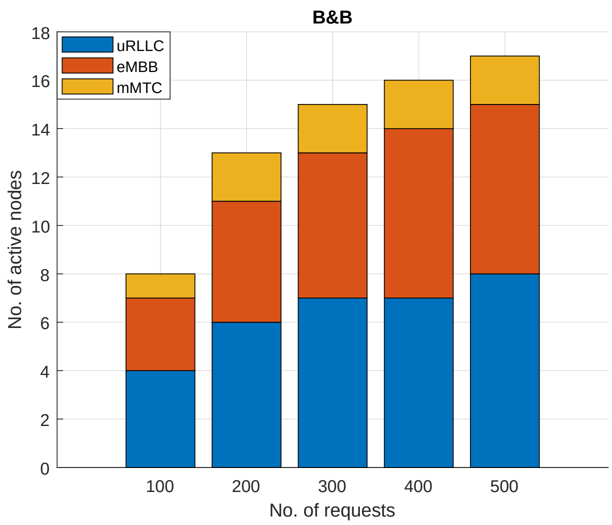


Figure 5.8: Comparison between the ILP and heuristic algorithm..

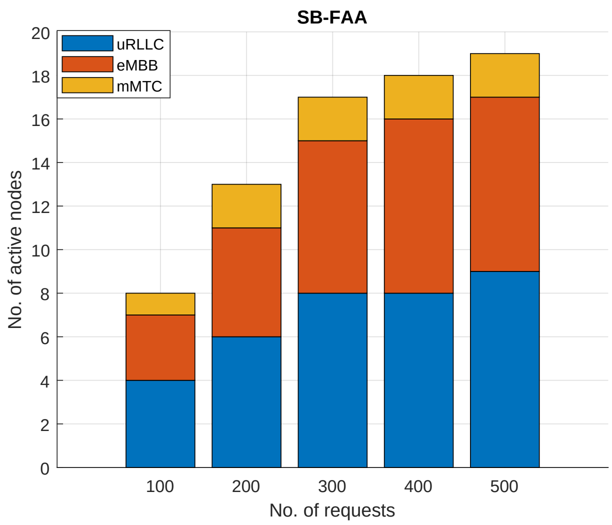


Figure 5.9: Comparison between the ILP and heuristic algorithm.

Table 5.2: Used values of parameters for Simulations

Parameter	Value
N	20
E	20
P	44
V	19
K	5
λ	8
C	10 Gbps
l_p	10-100μs
l_{EL}	20μs
L_D	0-200μs
R_C	200
R_M	200
R_F	50
c_r (fixed)	10-20 Gbps
c_r (mobile)	720 Mbps
c_r (fronthaul)	6 Gbps

5.0.4 Comparison between ILP and SB-FAA

We assume a network consisting of 30 nodes separated into 13 Macro CSs, 7 COs including a core CO, and 10 fog nodes. Once again the nodes are uniformly distributed in a 50 km^2 region. For a fair comparison, we consider same parameters and network scenario approaches. The comparison is provided relative to the number of active fog nodes and BBU hotels in the network for the 3-layer architecture. For both strategies as illustrated in Fig. 5.8,5.9, the number of active nodes increases with the increase in the number of requests. In case of URLLC and eMBB for higher traffic loads, the ILP slightly outperforms SB-FAA, whereas for mMTC traffic both techniques provide similar results. The near optimal performance of the proposed heuristic SB-FAA shows that CF-RAN over WDM outperforms traditional 2-layer approach irrespective of the used technique.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

We presented a novel CF-RAN over WDM architecture to address the stringent requirements of URLLC, eMBB and mMTC services. We jointly optimize the BBU hotel and fog node activation problem for CF-RAN over WDM architecture deployment. The fog nodes are deployed to cater the URLLC traffic seeking low latency, while the eMBB and mMTC traffic is processed at the BBU hotels. Further, dedicated wavelengths are assigned to each request type. We provided a mathematical approach based on ilp for lowering network costs by using the lowest number of active bbu hotels and fog nodes while satisfying the most requests. Furthermore, for realistic network scenarios, we present a low-complexity greedy-based heuristic approach. We compare the results obtained by deploying CF-RAN over WDM architecture with the traditional 2-layer centralization architecture. The simulation results reveal that the proposed architecture outperforms the traditional 2-layer architecture, in terms of BBU centralization, end-to-end (E2E) delay, and overall network connectivity. A significant improvement in terms of the number of active BBU hotels is observed as URLLC requests are processed at the fog nodes in the proposed model, thus provides benefit in terms of overall network deployment cost. Further, results show a notable improvement for all slice types in terms of E2E delay and network connectivity, as compared to the traditional 2-layer approach. We intend to continue our work and integrate the sdn controller at the bbu cloud into the CF-RAN architecture in the future.

6.2 Future Work

We intend to continue our work and integrate the sdn controller at the BBU cloud into the CF-RAN architecture in the future.

Bibliography

[1] R. K. Saha, S. Nanba, K. Nishimura, Y.-b. Kim, and K. Yamazaki, "RAN Architectural Evolution Framework toward 5G and Beyond Cellular-An Overview," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 592–593.

[2] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *IEEE network*, vol. 29, no. 1, pp. 35–41, 2015.

[3] L. M. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5G mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2018.

[4] R. I. Tinini, M. R. P. dos Santos, G. B. Figueiredo, and D. M. Batista, "5gpy: A simpy-based simulator for performance evaluations in 5g hybrid cloud-fog ran architectures," *Simulation Modelling Practice and Theory*, vol. 101, p. 102030, 2020.

[5] I. Chih-Lin, Y. Yuan, J. Huang, S. Ma, C. Cui, and R. Duan, "Rethink fronthaul for soft ran," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 82–88, 2015.

[6] C. Song, M. Zhang, Y. Zhan, D. Wang, L. Guan, W. Liu, L. Zhang, and S. Xu, "Hierarchical edge cloud enabling network slicing for 5g optical fronthaul," *Journal of Optical Communications and Networking*, vol. 11, no. 4, pp. B60–B70, 2019.

[7] N. H. Mahmood, M. Lauridsen, G. Berardinelli, D. Catania, and P. Mogensen, "Radio Resource Management Techniques for eMBB and mMTC services in 5G dense small cell scenarios," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[8] V. Q. Rodriguez and F. Guillemin, "Towards the deployment of a fully centralized Cloud-RAN architecture," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 2017, pp. 1055–1060.

[9] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *IEEE network*, vol. 29, no. 1, pp. 35–41, 2015.

[10] L. M. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2018.

[11] H. Yu, F. Musumeci, J. Zhang, Y. Xiao, M. Tornatore, and Y. Ji, "DU/CU placement for C-RAN over optical metro-aggregation networks," in *International IFIP Conference on Optical Network Design and Modeling*. Springer, 2019, pp. 82–93.

[12] Ericsson, "Common Public Radio Interface (CPRI); interface specification V7. 0," *Huawei Technologies Co. Ltd, NEC Corporation, Alcatel Lucent, and Nokia Networks*, 2015.

[13] "Study on new radio access technology: Radio access architecture and interfaces V14.0.0 (2017–03)," 2017.

[14] H. Yu, F. Musumeci, J. Zhang, M. Tornatore, and Y. Ji, "Isolation-aware 5G RAN slice mapping over WDM metro-aggregation networks," *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1125–1137, 2020.

[15] G. C. Valastro, D. Panno, and S. Riolo, "A SDN/NFV based C-RAN architecture for 5G mobile networks," in *2018 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, 2018, pp. 1–8.

[16] R. I. Tinini, D. M. Batista, G. B. Figueiredo, M. Tornatore, and B. Mukherjee, "Low-latency and energy-efficient BBU placement and VPON formation in virtualized cloud-fog RAN," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 4, pp. B37–B48, 2019.

[17] A. Ericsson *et al.*, "Common public radio interface (CPRI); interface specification v7. 0," *Huawei Technologies Co. Ltd, NEC Corporation, Alcatel Lucent, and Nokia Networks*, 2015.

[18] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks—a technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, 2014.

[19] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal BBU placement for 5G C-RAN deployment over WDM aggregation networks," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1963–1970, 2015.

[20] M. Ahsan, A. Ahmed, A. Al-Dweik, and A. Ahmad, "Functional split-aware optimal bbu placement for 5g cloud-ran over wdm access/aggregation network," *IEEE Systems Journal*, 2022.

[21] R. I. Tinini, L. C. Reis, D. M. Batista, G. B. Figueiredo, M. Tornatore, and B. Mukherjee, "Optimal placement of virtualized BBU processing in hybrid cloud-fog RAN over TWDM-PON," in *IEEE Global Commun. Conf.* IEEE, 2017, pp. 1–6.

[22] R. I. Tinini, D. M. Batista, G. B. Figueiredo, M. Tornatore, and B. Mukherjee, "Energy-efficient baseband processing via vBBU migration in virtualized cloud-fog RAN," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[23] A. Kaloxylos, "A survey and an analysis of network slicing in 5G networks," *IEEE Commun. Standards Mag.*, vol. 2, no. 1, pp. 60–65, 2018.

[24] L. M. Larsen, M. S. Berger, and H. L. Christiansen, "Fronthaul for cloud-ran enabling network slicing in 5g mobile networks," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

[25] C. Song, M. Zhang, Y. Zhan, D. Wang, L. Guan, W. Liu, L. Zhang, and S. Xu, "Hierarchical edge cloud enabling network slicing for 5G optical fronthaul," *IEEE J. Opt. Commun. Netw.*, vol. 11, no. 4, pp. B60–B70, 2019.

[26] I. Chih-Lin, J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li, "Recent progress on C-RAN centralization and cloudification," *IEEE Access*, vol. 2, pp. 1030–1039, 2014.

[27] S. Namba, T. Warabino, and S. Kaneko, "BBU-RRH switching schemes for centralized RAN," in *7th International Conference on Communications and Networking in China*, 2012, pp. 762–766.

[28] A. Saadani, M. El Tabach, A. Pizzinat, M. Nahas, P. Pagnoux, S. Purge, and Y. Bao, "Digital radio over fiber for LTE-advanced: Opportunities and challenges," in *2013 17th International Conference on Optical Networking Design and Modeling (ONDM)*, 2013, pp. 194–199.

[29] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul Networks," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[30] A. Ahmad, A. Bianco, and E. Bonetto, "Traffic grooming and energy-efficiency in flexible-grid networks," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 3264–3269.

[31] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wübben, J. Bartelt, R. Fritzsche, and P. Rost, "Towards a flexible functional split for cloud-RAN networks," in *2014 European Conference on Networks and Communications (EuCNC)*, 2014, pp. 1–5.

[32] 3GPP, "3GPP TR 38.801 V14.0.0 (2017-03): Study on new radio access technology: Radio access architecture and interfaces," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.801.

[33] C.-Y. Chang, N. Nikaein, R. Knopp, T. Spyropoulos, and S. S. Kumar, "FlexCRAN: A flexible functional split framework over ethernet fronthaul in Cloud-RAN," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–7.

[34] A. Checko, A. P. Avramova, M. S. Berger, and H. L. Christiansen, "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings," *Journal of Communications and Networks*, vol. 18, no. 2, pp. 162–172, 2016.

[35] G. O. Pérez, D. L. López, and J. A. Hernández, "5G new radio fronthaul network design for eCPRI-IEEE 802.1 CM and extreme latency percentiles," *IEEE Access*, vol. 7, pp. 82 218–82 230, 2019.

[36] J.-P. Elbers and J. Zou, "A Flexible X-haul Network for 5G and beyond," in *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, 2019, pp. 1–3.

[37] G. Otero Pérez, D. Larrabeiti López, and J. A. Hernández, "5G New Radio Fronthaul Network Design for eCPRI-IEEE 802.1CM and Extreme Latency Percentiles," *IEEE Access*, vol. 7, pp. 82 218–82 230, 2019.

[38] A. Ahmad, A. Bianco, E. Bonetto, M. Garrich, and J. R. Oliveira, "Switching node architectures in flexible-grid networks: A performance comparison," in *Intl. Conf. Opt. Netw. Des. Modeling*, 2014, pp. 49–54.

[39] T. A. Strasser and J. L. Wagener, “Wavelength-selective switches for ROADM applications,” *IEEE journal of selected topics in quantum electronics*, vol. 16, no. 5, pp. 1150–1157, 2010.

[40] H. Yu, F. Musumeci, J. Zhang, M. Tornatore, and Y. Ji, “Isolation-aware 5G RAN slice mapping over WDM metro-aggregation networks,” *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1125–1137, 2020.

[41] M. R. P. dos Santos, R. I. Tinini, G. B. Figueiredo, and D. M. Batista, “Data analysis and energy consumption prediction in a Cloud-Fog RAN environment,” in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*, 2019, pp. 1–6.

[42] M. Waqar, A. Kim, and P. K. Cho, “A transport scheme for reducing delays and jitter in ethernet-based 5g fronthaul networks,” *IEEE Access*, vol. 6, pp. 46 110–46 121, 2018.

[43] N. Alliance, “Description of network slicing concept,” *NGMN 5G P*, vol. 1, no. 1, 2016.

[44] T. Yoo, “Network slicing architecture for 5g network,” in *2016 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2016, pp. 1010–1014.

[45] Y. L. Lee, J. Loo, T. C. Chuah, and L.-C. Wang, “Dynamic network slicing for multitenant heterogeneous cloud radio access networks,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2146–2161, 2018.

[46] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "On radio access network slicing from a radio resource management perspective," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 166–174, 2017.

[47] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[48] A. Ahmad, A. Bianco, H. Chouman, V. Curri, G. Marchetto, and S. Tahir, "A transmission layer aware network design for fixed and flexible grid optical networks," in *Intl. Conf. Transparent Opt. Netw. (ICTON)*, 2015, pp. 1–4.

[49] D. B. Johnson, "A note on Dijkstra's shortest path algorithm," *J. ACM (JACM)*, vol. 20, no. 3, pp. 385–388, 1973.

[50] Y. Alfadhli, Y.-W. Chen, S. Liu, S. Shen, S. Yao, D. Guidotti, S. Mitani, and G.-K. Chang, "Latency performance analysis of low layers function split for URLLC applications in 5G networks," *Computer Networks*, vol. 162, p. 106865, 2019.