Passive Optical Network

Dynamic Bandwidth Allocation Algorithms

for Converged Fronthaul

A Thesis Submitted to the University of Kent for the degree of

Doctor of Philosophy in Electronic Engineering

By

Samuel O. Edeagu

June 2021

To my father,

S.N. Edeagu

The bones shall rise again...

– Ezekiel 37:1-5

Abstract

The research presented in this thesis is focused on the use of passive optical network (PON) dynamic bandwidth allocation (DBA) algorithms for converged fronthaul. This work adapts and extends an existing model for use with a status reporting DBA that incorporates a colorless grant. This thesis also proposes a hybrid DBA that operates with the status reporting DBA and cooperative DBA.

The models for the status reporting DBA and the proposed hybrid DBA are implemented in the OMNeT++ network simulator. The performance of both DBAs is evaluated using four simulation experiments in two deployment scenarios for a 10 Gbit/s symmetric PON (XGS-PON) system with an emphasis on the average optical network unit (ONU) upstream queueing delay and the proportion of frames meeting the latency requirement for fronthaul traffic. This is achieved by allocating bandwidth to the different types of traffic in a converged network based on their priority and latency requirement.

The key findings from the simulation experiments show that, when the delay distribution of frames is analysed, a status reporting DBA that incorporates a colorless grant is suitable for meeting the latency requirement in a converged network, provided an adequate allocation of bandwidth is made available for fronthaul traffic. More than 99% of the fronthaul traffic at 80% traffic load in all scenarios meet the fronthaul latency requirement in the status reporting DBA. Furthermore, when the two DBAs are compared, the hybrid DBA provides zero frame loss for fronthaul traffic in all scenarios which is not achievable using the status reporting DBA.

Acknowledgements

I would like to first express my gratitude to my supervisor Professor Nathan Gomes, for his support and guidance during my PhD programme.

My thanks also go to my research collaborators, Dr. Rizwan Aslam Butt and Prof. Dr. Sevia M. Idrus of Universiti Teknologi Malaysia (UTM) and to Dr. Jung-Bo Wang of Southeast University, China for fruitful discussions during his two-year fellowship at the University of Kent.

I am grateful to the Federal Government of Nigeria for funding three years of my PhD programme as a recipient of the National Information Technology Development Fund Scholarship for Doctoral Studies.

Finally, to my family and friends – you have my deepest thanks!

List of Abbreviations

Abbreviation	Description
3GPP	3rd Generation Partnership Project
4G	4th Generation (Mobile Communications)
5G	5th Generation (Mobile Communications)
5GC	5G Core Network
ACK	Acknowledgement
ADSL	Asymmetric Digital Subscriber Line
AON	Active Optical Network
APON	Asynchronous Transfer Mode Passive Optical Network
ARQ	Automatic Repeat reQuest
ATM	Asynchronous Transfer Mode
AWG	Arrayed Waveguide Grating
BBU	Baseband Unit
BPON	Broadband Passive Optical Network
BS	Base Station
BSC	Base Station Controller
BTS	Base Transceiver Station
BW	Bandwidth
BWmap	Bandwidth Map
C&M	Control and Management
CapEx, CAPEX	Capital Expenditure
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CEx	Coexistence Element
CN	Core Network
CO	Central Office
CO-DBA	Cooperative Dynamic Bandwidth Allocation
CoMP	Coordinated Multi-Point
CP	Control Plane
CPRI	Common Public Radio Interface
C-RAN	Centralised Radio Access Network
CU	Central Unit
CUPS	Control and User Plane Separation
CWDM	Coarse Wavelength Division Multiplexing
DBA	Dynamic Bandwidth Allocation
DBRu	Dynamic Bandwidth Report upstream
DC	Dual Connectivity
DCI	Downlink Control Information
DL	Downlink
DSL	Digital Subscriber Line

DU	Distributed Unit
DWDM	Dense Wavelength Division Multiplexing
E2E	End-to-End
eCPRI	Evolved (or Enhanced) Common Public Radio Interface
eICIC	enhanced Inter-Cell Interference Coordination
eMBB	Enhanced Mobile Broadband
eNB, eNodeB	LTE base station
EPC	Evolved Packet Core
EPON	Ethernet Passive Optical Network
eRE	eCPRI Radio Equipment
eREC	eCPRI Radio Equipment Control
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved UTRAN
F5G	Fifth Generation Fixed Network
FAPI	Functional Application Platform Interface
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FMC	Fixed Mobile Convergence
FSAN	Full-Service Access Network
FTTB	Fibre to the Building
FTTC	Fibre to the Curb
FTTCab	Fibre to the Cabinet
FTTH	Fibre to the Home
FTTN	Fibre to the Node
FTTP	Fibre to the Premises
FTTx	Fibre to the x
GEM	GPON Encapsulation Method
gNB, gNodeB	5G NR base station
GPON	Gigabit-capable Passive Optical Network
GPRS	General Packet Radio Services
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Retransmit reQuest
HEC	Header Error Correction
HLS	High Layer Split
HSPA	High Speed Packet Access
IACG	Immediate Allocation with Colorless Grant
ICIC	Inter-Cell Interference Coordination
IEEE	Institute of Electrical and Electronics Engineers

IFFT	Inverse Fast Fourier Transform
IP	Internet Protocol
IQ	In-Phase Quadrature
ITU	International Telecommunication Union
ITU-R	ITU Radiocommunication Sector
ITU-T	ITU Telecommunication Standardization Sector
KPI	Key Performance Indicator
LAN	Local Area Network
L1	Layer 1
L2	Layer 2
L3	Layer 3
LLC	Logical Link Control
LLS	Low Layer Split
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
MAC	Media Access Control
MEC	Multi-Access Edge Computing
MEF	Metro Ethernet Forum
MIMO	Multiple Input, Multiple Output
mMTC	Massive Machine Type Communication
MSC	Mobile Switching Centre
NAS	Non-Access Stratum
NB, Node B	3G UMTS base station
nFAPI	Network Functional Application Platform Interface
NGC	Next Generation Core
NGFI	Next Generation Fronthaul Interface
NGMN	Next Generation Mobile Networks Alliance
NG-PON	Next Generation Passive Optical Network
NG-PON2	Next Generation Passive Optical Network 2
NG-RAN	Next Generation Radio Access Network
NR	New Radio
NRT	Non-Real Time
OAM	Operation, Administration and Maintenance
OAN	Optical Access Network
OBSAI	Open Base Station Architecture Initiative
ODN	Optical Distribution Network
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OFDM-PON	Orthogonal Frequency Division Multiplexing Passive Optical Network
OLT	Optical Line Terminal
OMCI	ONT Management and Control Interface
ONT	Optical Network Terminal

ONU	Optical Network Unit
OpEx, OPEX	Operational Expenditure
Open RAN, O-RAN	Open Radio Access Network
ORI	Open Radio Equipment Interface
OTN	Optical Transport Network
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PDU	Protocol Data Unit
PHY	Physical Layer
PLOAM	Physical Layer Operations, Administration and Maintenance
PON	Passive Optical Network
PtMP, P2MP	Point to Multipoint
PtP, P2P	Point to Point
PRACH	Physical Random Access Channel
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RE	Resource Element
REC	Radio Equipment Control
RF	Radio Frequency
RLC	Radio Link Control
RN	Remote Node
RNC	Radio Network Controller
RNL	Radio Network Layer
RRC	Radio Resource Control
RoE	Radio over Ethernet
RoF	Radio over Fibre
RRC	Radio Resource Control
RRH	Remote Radio Head
RRU	Remote Radio Unit
RT	Real Time
RTT	Round Trip Time
RU	Radio Unit
SDAP	Service Data Adaptation Protocol
SISO	Single Input, Single Output
SCF	Small Cell Forum
SLA	Service Level Agreement
SR-DBA	Status Reporting Dynamic Bandwidth Allocation
TCO	Total Cost of Ownership

T-CONT	Transmission Container
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TDM-PON	Time Division Multiplexing Passive Optical Network
TM-DBA	Traffic Monitoring Dynamic Bandwidth Allocation
TNL	Transport Network Layer
TSN	Time Sensitive Networking
TTI	Transmission Time Interval
TWDM	Time and Wavelength Division Multiplexing
TWDM-PON	Time and Wavelength Division Multiplexing Passive Optical Network
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UP	User Plane
UMTS	Universal Mobile Telecommunication System
URLLC	Ultra Reliable Low Latency Communication
UTRAN	UMTS Terrestrial Radio Access Network
VDSL	Very High Speed Digital Subscriber Line
vRAN	Virtualised Radio Access Network
WCDMA	Wideband Code Division Multiple Access
WDM	Wavelength Division Multiplexing
WDM-PON	Wavelength Division Multiplexing Passive Optical Network
XGEM	XG-PON/XGS-PON Encapsulation Method
XG-PON	10-Gigabit-capable Passive Optical Network
XGS-PON	10-Gigabit-capable symmetric Passive Optical Network

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background and Motivation

According to a recent report by Cisco, global mobile data traffic is expected to increase tremendously by the year 2023, with an estimated 5.3 billion users (see Figure *1.1*) and 13.1 billion mobile devices connected to the Internet, as shown in Figure *1.2* [1]. Ericsson forecasts global mobile data traffic to exceed 300 exabytes (EB) per month in 2026, up from 58 EB per month at the end of 2020 [2]. The current and future applications driving this high demand in mobile data traffic include industrial Internet of Things (IoT) solutions, connected home applications, 4K video streaming, real time gaming, autonomous driving cars, monitoring sensor networks, augmented reality (AR), virtual reality (VR), among others [1], [2]. The next generation of mobile networks are expected to satisfy this high and heterogeneous demand in traffic through new converged network architectures.

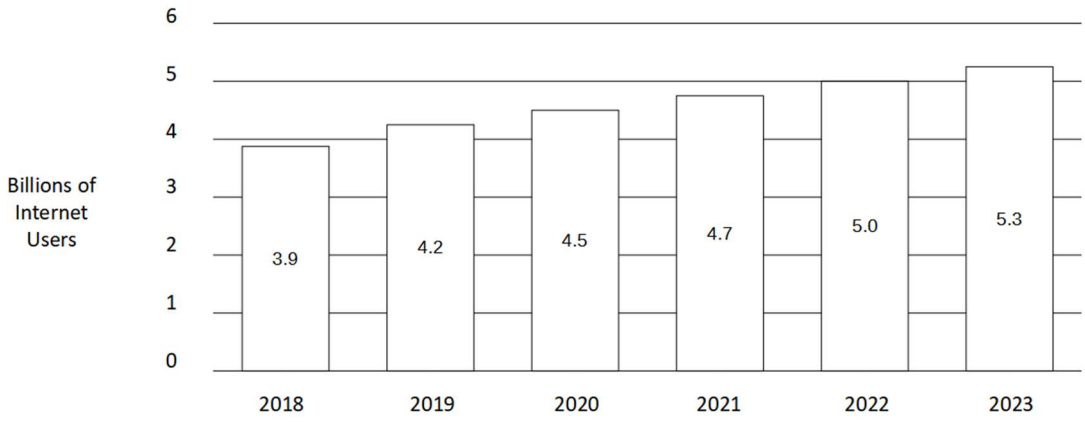


Figure 1.1: Global Internet user growth [1].

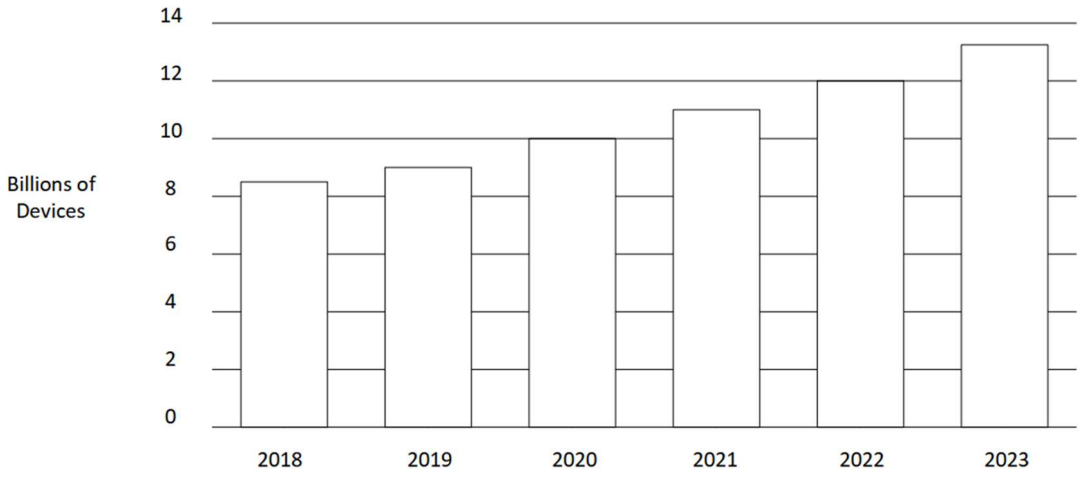


Figure 1.2: Global mobile device and connection growth [1].

Traditionally, the development of fixed and mobile networks has progressed independently. Fixed networks started from the first generation using the POTS (Plain Old Telephone Service) analogue technology to transmit voice traffic, the second generation with asymmetric digital subscriber line (ADSL) technologies, the third generation with VDSL (Very High Speed DSL) – a higher data rate version of DSL, to the fourth generation using G.fast – a DSL protocol standard, and GPON, a passive optical network (PON) technology. Fixed networks are now in their 5th generation, known as Fixed Generation Fixed Network (F5G) [3]. The key developments over the generations have been the shift from analogue to digital technologies and the increasing replacement of copper with optical fibre in the access network infrastructure, which has led to higher capacity, lower latency, higher reliability and longer distances for broadband access.

Starting in the 1970s, mobile networks have undergone four generations – 1G, 2G, 3G and 4G – and the fifth generation (5G) is currently being deployed worldwide. The

mobile network generations are discussed extensively in Chapter 2. The introduction of 5G mobile networks are targeted at enabling new requirements that must be met, such as high capacity (> 10 Gbit/s), low latency (< 1 ms) and high reliability (> 99.999% connection availability) [4].

The large-scale deployment of 5G mobile networks will rely heavily on optical fibre connections, especially in the access segment. The convergence of the fixed and mobile networks, also known as fixed mobile convergence (FMC), will provide broadband access for fixed residential and business users as well as serve as a transport network for mobile networks. Future passive optical networks are also expected to meet the same transport requirements as 5G in order to transmit both mobile and fixed access traffic.

The European FP7 project – COnvergence of fixed and Mobile BrOadband access/aggregation networks (COMBO) – defined two key aspects of a FMC network [5]. In a structurally converged network, the existing network infrastructure is shared and reused for 4G and 5G small cell deployments. In functionally converged networks, a set of generic network functions are implemented to realise similar goals in different network types (fixed, mobile, Wi-Fi).

Since 2017, the mobile standards body 3rd Generation Partnership Project (3GPP) and the Broadband Forum's Wireless-Wireline Convergence (WWC) Work Area have collaborated on FMC for 5G. The first set of convergence specifications describing the architecture, functions and interfaces for 5G FMC were published at the end of 2020 [6]–[8], [9]. The 5G wireless wireline convergence architecture includes a mediation

function, the access gateway function (AGF), that resides between existing fixed access networks (fibre, coaxial, twisted-pair) and the 5G core to create a truly converged deployment. The ultimate goal of the collaboration is to have a common core for both the fixed and mobile access networks [10].

Fixed broadband access networks generally use time division multiplexing passive optical networks (TDM-PON) to provide connectivity because of their ability to use one fibre cable to connect multiple users which leads to a cost advantage when compared to a point-to-point connection. The total bandwidth is shared between all users using a dynamic bandwidth allocation (DBA) scheme with one wavelength in the upstream and one in the downstream direction. The DBA method widely used in TDM-PON is the status reporting DBA (SR-DBA) [11] which adapts to the actual demands of the network traffic and allows for statistical multiplexing that increases bandwidth utilisation efficiency. However, the latency due to the DBA for upstream transmission in a TDM-PON is a major challenge in 5G networks. A new DBA method has been specified by the ITU Telecommunication Standardization Sector (ITU-T) [12], named the cooperative DBA (CO-DBA), which allows for cooperation between 5G network elements and PON transport elements when allocating bandwidth.

Much of the research on latency in converged networks has largely focused on PONs specified by the Institute of Electrical and Electronics Engineers (IEEE), a standards development organisation. A relatively few DBA algorithms based on the ITU-T standards have been studied for use in the converged networks. This will be discussed in Chapter 4.

1.2 Thesis Goals

The main goals of this thesis are listed below:

The first goal of this thesis is to investigate the use of a passive optical network dynamic bandwidth allocation algorithm for converged fronthaul by adapting and extending the model presented in [13]. This model was used to study energy efficient cyclic sleep dynamic bandwidth allocation schemes for passive optical networks [14], [15].

The second goal of this thesis is to demonstrate the PON DBA using simulation experiments and to analyse its performance. The analysis focuses on the average ONU upstream queueing delay and the proportion of frames meeting the fronthaul latency requirement.

The third goal of this thesis is to propose a hybrid DBA that operates with a PON DBA and the cooperative DBA, which enables a mix of different traffic types to be transmitted on the same PON while satisfying the transport requirements for converged fronthaul. This will be based on a 250 μs latency requirement and 99.9% reliability requirement for fronthaul traffic.

The fourth goal of this thesis is to show that the hybrid DBA can be used to effectively reduce the latency, increase the proportion of frames meeting the fronthaul latency requirement and efficiently utilise the bandwidth in a converged fronthaul. The hybrid

DBA is validated by comparing it with the extended Immediate Allocation with Colorless Grant (IACG) DBA model on which it operates with.

1.3 Contributions

This section discusses the key contributions of the research work presented in this thesis.

Extension of PON DBA model for use in a converged network.

The original DBA model has been extended to account for the different types of traffic in a converged network using distinct traffic generators for each transmission container (T-CONT) type in the optical network units (ONUs) and assigning them different bandwidth service classes with priorities. The model was also extended to include frame counters at the ONUs to keep track of the number of frames exiting the ONU queue in the upstream direction. The ONU frame counters enable an accurate determination of the ONU queueing delay by using a fixed frame size of 1500 bytes. The original DBA model used the tri-modal distribution traffic generation model which generated frames of three different sizes – 64, 500, 1500 bytes. Due to the frame fragmentation that is allowed in ITU TDM-PONs, the original model took into account the queueing delay of all frame fragments of a particular frame leaving the ONU which led to inaccurate results. In addition, the extended model enabled traffic to be generated in different proportions for each T-CONT so as to investigate different simulation experiments.

Use of IACG DBA for meeting latency requirements in a converged network carried out for the first time.

Existing ITU TDM-PON DBA algorithms are largely based on the GIANT DBA which only grants bandwidth once during a service interval and does not have a colorless grant phase like the IACG DBA. The use of IACG DBA provides lower delay values due to it sending bandwidth grants in every downstream frame and the use of the colorless grant phase which assigns the unallocated bandwidth of the upstream frame at the end of the DBA cycle to each ONU equally. The delay distribution of the T-CONT frames were analysed to determine whether the IACG DBA algorithm is suitable for meeting the latency requirement in a converged network. Several simulation experiments in two deployment scenarios were investigated to determine the percentage of fronthaul traffic that met the latency requirement. This was achieved by making a trade-off between the different traffic types whereby fronthaul traffic is given a proportionally much higher bandwidth allocation, as the other types of traffic have less stringent latency requirements. The results showed that 99% of fronthaul traffic at 80% load in the two deployment scenarios met the latency requirement. The findings, which were published in the proceedings of the 25th International Conference on Optical Network Design and Modelling (ONDM 2021) [16], are presented in Chapter 4.

Proposal of a novel hybrid DBA that operates with the IACG DBA and Cooperative DBA.

The proposed hybrid DBA enables a mix of different traffic types to be transmitted on the same PON while satisfying the strict latency requirement for fronthaul traffic. The latency-sensitive fronthaul traffic is handled by the cooperative DBA while the latency-tolerant non-fronthaul traffic (midhaul, backhaul and fixed access traffic) is handled by

the IACG DBA, which is a status reporting DBA. By combining the cooperative DBA with the IACG DBA, the hybrid DBA reduces the idle period in a converged network. The idle period is the time during which the ONUs wait for their bandwidth allocation from the OLT. Reducing the idle period leads to a reduction in latency and increases the bandwidth utilisation of frames in a converged network. In addition, the hybrid DBA achieves upstream bandwidth allocation that takes into account the priorities of the T-CONT types to meet the latency requirements of different traffic types in a converged network. The hybrid DBA when compared with the IACG DBA was shown to provide zero frame loss for fronthaul traffic in all scenarios which is not achievable using the IACG DBA. These findings are presented in detail in Chapter 5 and have been submitted to a journal for publication.

1.4 Publications

The following have been published as a result of the research carried out and presented in this thesis.

[1] S.O. Edeagu, "Passive Optical Networks in 5G Mobile Fronthaul," 2nd PhD and ECI Meeting Report of Activities and Proceedings, European Network for High Performance Integrated Microwave Photonics (EUIMWP) COST Action CA16220, Thessaloniki, Greece, 9 September 2019, pp. 63 – 72.

[2] S.O. Edeagu**,** R.A. Butt, S.M. Idrus, and N.J. Gomes, "Performance of PON Dynamic Bandwidth Allocation Algorithm for Meeting xHaul Transport

Requirements," Proceedings of 25th International Conference on Optical Network Design and Modelling (ONDM 2021), Gothenburg, Sweden, 28 June – 1 July 2021, pp. 1 – 6.

[3] S.O. Edeagu**,** R.A. Butt, S.M. Idrus, and N.J. Gomes, "A Hybrid Dynamic Bandwidth Allocation Scheme operating with IACG and Cooperative DBA for Converged Fronthaul," Optical Switching and Networking *(submitted).*

1.5 Thesis Outline

The thesis is divided into six chapters, including this one. The general contents of the chapters are given below:

Chapter 2 presents a review of the radio access network (RAN) architecture, starting with the various mobile network generations and architectures. The chapter focuses on the fifth generation (5G) of mobile networks and discusses the latency, synchronisation and reliability requirements for converged networks and the various RAN functional splits.

Chapter 3 describes an overview of passive optical networks (PONs), its architecture and standards. The various dynamic bandwidth allocation algorithms (DBAs) used in PONs are also discussed with a focus on ITU PON DBAs.

Chapter 4 presents the simulations of the ITU PON DBA introduced in Chapter 3. Different simulation scenarios are presented considering the various traffic types in a converged network. The simulations were conducted using the open source discrete-event network simulator called OMNeT++. The results obtained are analysed and discussed.

Chapter 5 presents a proposed hybrid DBA for a converged fronthaul. The hybrid DBA operates with the cooperative DBA and the IACG DBA, which is a status reporting DBA for PON.

Chapter 6 contains the conclusion of the research work presented in this thesis. In addition, the chapter outlines suggestions for possible future work.

2. Radio Access Network Architecture

2.1 Introduction

This chapter focuses on the radio access network architecture (RAN) of fifth generation (5G) mobile networks. In section 2.2, an overview of mobile network technologies is presented, followed by a discussion of new RAN architectures, the 5G transport network and its requirements. Finally, the 5G RAN architecture, the protocol stack and the different functional splits are described in section 2.3.

2.2 Mobile Network Architecture

A typical mobile network architecture is made up of a core network (CN) and a radio access network (RAN). The RAN connects a large number of user equipment (UE) to the core network via base stations using a transport network called the backhaul [17]. The RAN implements a radio access technology (RAT), such as GSM, UMTS, LTE, NR. Mobile networks are categorised into generations. A new generation usually occurs every ten years and started with the first generation in the early 1980s. This thesis is focused on the RAN segment, which is discussed in subsequent sections in this chapter.

First Generation (1G)

The first generation of mobile networks was based on analogue transmission techniques. There were several competing standards around the world. It was called the Advanced Mobile Phone Service (AMPS) in the United States, the Total Access Communication System (TACS) and later Enhanced TACS (ETACS) in the United Kingdom,

while the Nordic countries (Denmark, Finland, Iceland, Sweden, Norway) implemented the Nordic Mobile Telephone (NMT) system.

Second Generation (2G)

2G saw the advent of digital transmission technology and a more common standard, initiated by the European Telecommunications Standards Institute (ETSI). The Global System for Mobile Communications (GSM) was largely adopted in Europe, the Middle East, Africa and most of Asia. The United States opted for the IS-136 (Digital AMPS or D-AMPS) and IS-95 (cdmaOne). Provision for data was added through GPRS (General Packet Radio Services) and EDGE (Enhanced Data rates in GSM Environments). The GSM architecture [18], as shown in Figure *2.1*, includes base station controllers (BSC) which manage the radio resources for one or more base transceiver stations (BTS) located at various cell sites. The BTS connects to the BSC via the logical interface *Abis* while the BSC connects to the mobile switching centre (MSC) in the core network using the *A* logical interface.

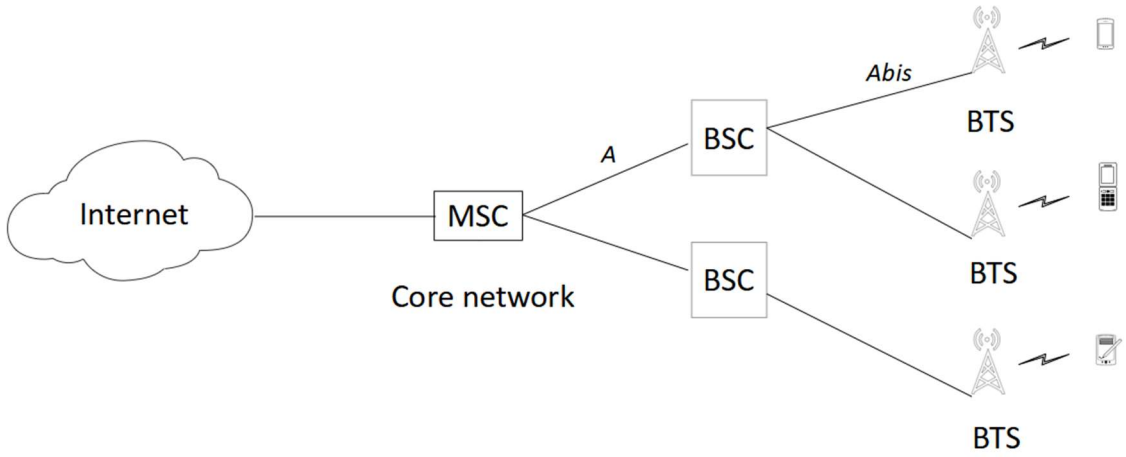


Figure 2.1: 2G GSM architecture.

Third Generation (3G)

3G, which was named IMT-2000 by the International Telecommunication Union (ITU), evolved as two variants. In Europe, it was called the Universal Terrestrial Mobile System (UMTS) and the American version was CDMA2000 – a code division multiple access (CDMA) version of IMT-2000. Further enhancements were introduced to UMTS known as the High Speed Packet Access (HSPA) and HSPA+. As shown in Figure *2.2*, the various base stations – Node B – are controlled by the radio network controller (RNC) through the logical interface *Iub* [18]. The RNC is connected to the core network using the logical interface *Iu* while communication between the RNCs occur over the logical interface *Iur*.

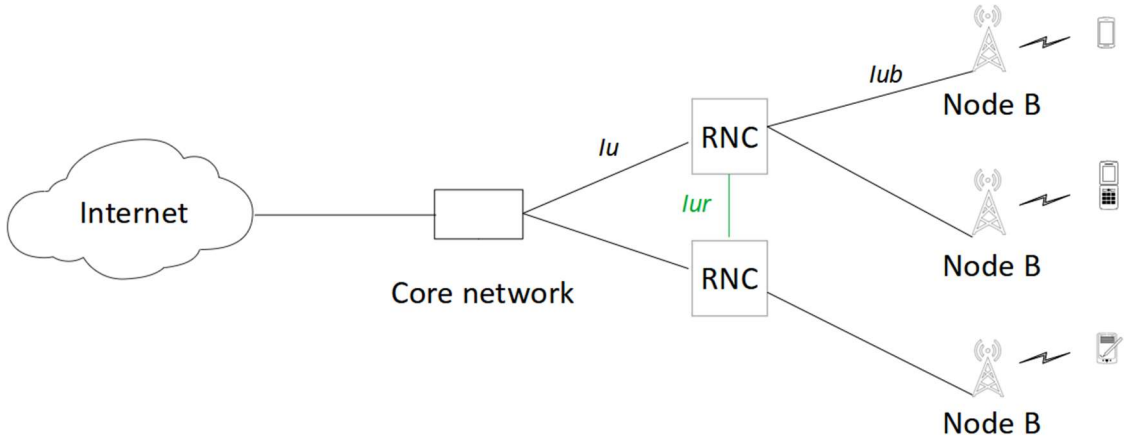


Figure 2.2: 3G UMTS architecture.

Fourth Generation (4G)

4G or IMT-Advanced, saw the convergence of the two variants in 3G into the Long Term Evolution of UMTS (LTE) standard used by the 3rd Generation Partnership Project (3GPP) standards organisation. Release 8 and 9 of the 3GPP specification series define LTE while Releases 10 – 12 specify LTE-Advanced. Release 13 is considered a bridge between 4G and 5G, and is called LTE Advanced Pro or 4.5G. As shown in Figure *2.3,* the base station, eNodeB or eNB, is connected to the core network, Evolved Packet Core (EPC), using the

logical interface, *S1*. The *X2* logical interface provides connections between the eNodeBs. LTE is based on a flat, all-IP (Internet Protocol) architecture, meaning there is only one element type (eNodeB) for the radio network, and one element type (EPC) for the core network [19].

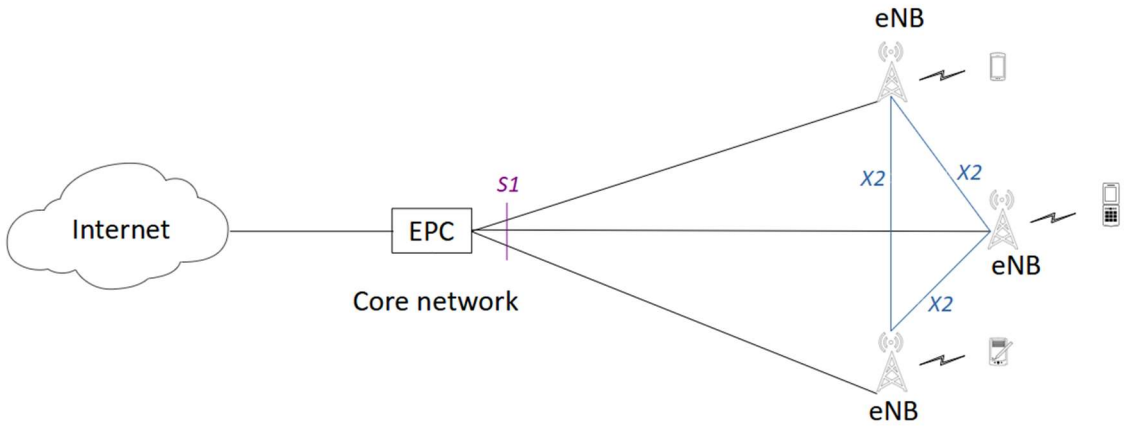


Figure 2.3: 4G LTE architecture.

Fifth Generation (5G)

In 2015, the ITU through its Radiocommunication Sector (ITU-R) [4] chose IMT-2020 as the formal name for the next generation of mobile broadband communication systems beyond IMT-Advanced (4G). The technical foundation of 5G started with the 3GPP Release 15 standard first published in 2018 [20] and completed in 2019. Further releases have since been published with Release 16 completed in 2020 and ongoing work with Release 17 and Release 18, also known as 5G Advanced. The various radio technology candidates for IMT-2020 were finalised by the ITU-R in 2021 [21]. In 5G, a new radio interface – New Radio (NR) – was introduced [22], where the use of more bandwidth at higher frequency ranges (e.g., around 4 GHz [23], [24] and around 30 GHz [25]) is available and a higher number of antenna ports and Multiple Input Multiple Output

(MIMO) layers are supported. The 5G NR architecture is discussed in section 2.3. There are several factors that drive the need for new mobile generations such as overcoming deficiencies and shortcomings of previous mobile generations, industry-driven standards and the need to benefit from licensing standard essential patents, research into new mobile technologies in university and industry labs, government policies and incentives.

Next generation mobile networks (5G and beyond) are expected to meet diverse latency (delay), bandwidth and reliability requirements. These requirements are needed in order to support new use cases, defined generally by the enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (URLLC) and massive machine type communications (mMTC) service types [4].

The eMBB use case addresses scenarios which require extremely high data rates and low latency. They also offer wide area coverage and hotspot, i.e., an area with a high number of users (e.g., stadia, shopping malls). The URLLC use case requires low latency and high reliability. mMTC, also known as massive IoT, is characterised by a very large number of connected devices which are capable of staying inactive for long periods of time. Table *2.1* lists examples of the three main 5G use cases and their requirements [26].

Table 2.1: Examples of 5G use cases and requirements [26].

Use cases	Example	Max End-to-end Latency	Data Rate	Reliability
eMBB	Augmented and virtual reality (AR/VR) environments	5 – 10 ms	0.1 – 10 Gbit/s	99.99 %
eMBB	Interactive gaming services	10 ms	0.1 – 1 Gbit/s	99.99 %
URLLC	Remote control for process automation	50 ms	100 Mbit/s	99.9999%
URLLC	Wireless road-side infrastructure backhaul	30 ms	10 Mbit/s	99.999%
mMTC	Medical monitoring	100 ms	< 1 Mbit/s	99.9999%

Addressing these new requirements led to the introduction of new RAN architectures such as the centralised RAN (C-RAN) [27] and the virtualised RAN (vRAN).

Legacy mobile networks are largely based on the **distributed radio access network (D-RAN)** architecture in which the base stations are connected to the core network by backhaul links [28]. In a D-RAN, the radio frequency (RF) and baseband processing are located at the cell site, as shown in Figure *2.4.*

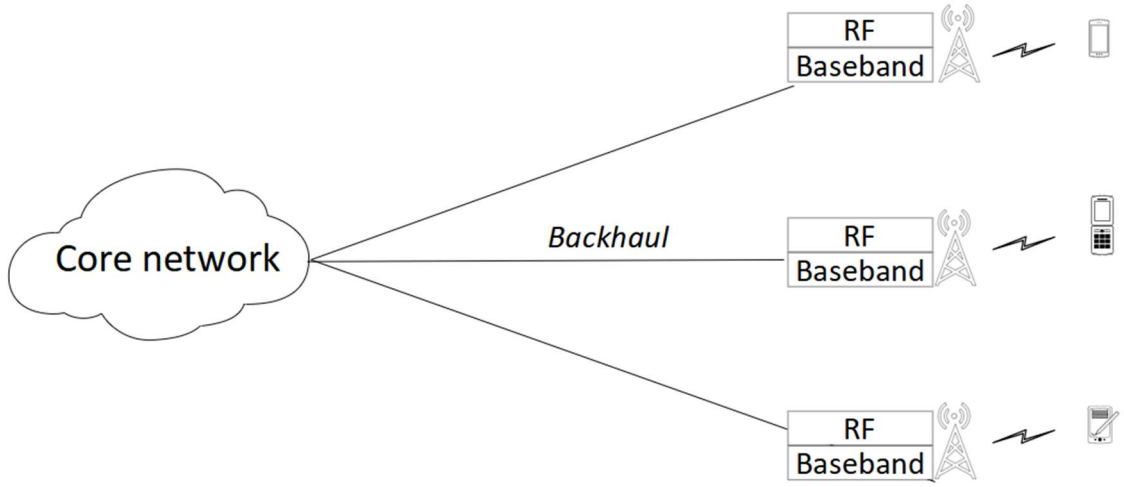


Figure 2.4: Distributed RAN (D-RAN) architecture.

The **Centralised RAN** was first proposed by the China Mobile Research Institute in 2009 [27]. The C-RAN involves separating the RF related functions of the base station (Remote Radio Head: RRH) from the baseband processing functions (Baseband Unit: BBU), as shown in Figure *2.5*. The BBU is centralised in a single location and serves a large number of RRHs [28], [29]. In general, there is one BBU for each radio access technology (GSM, UMTS, LTE, LTE-Advanced, NR). The transport link, known as the fronthaul network, connects the RRH to the BBU [30]. The protocol used to communicate between the BBU and the RRH – the Common Protocol Radio Interface (CPRI) – requires a significantly higher data transmission rate than the payload that it is carrying. The fronthaul is usually implemented using optical fibre because it provides large bandwidth and high data rates over long distances. The various options using optical fibre include dark fibre, wavelength division multiplexing (WDM), optical transport network (OTN) and passive optical network (PON) [27]. Other transport options include Ethernet (Ethernet over dark fibre, FlexE and G.mtn), free space optics and microwave and millimetre wave radio transmission [31]. The most widely used mobile network standards deployed as C-RAN is 4G (LTE and LTE-Advanced).

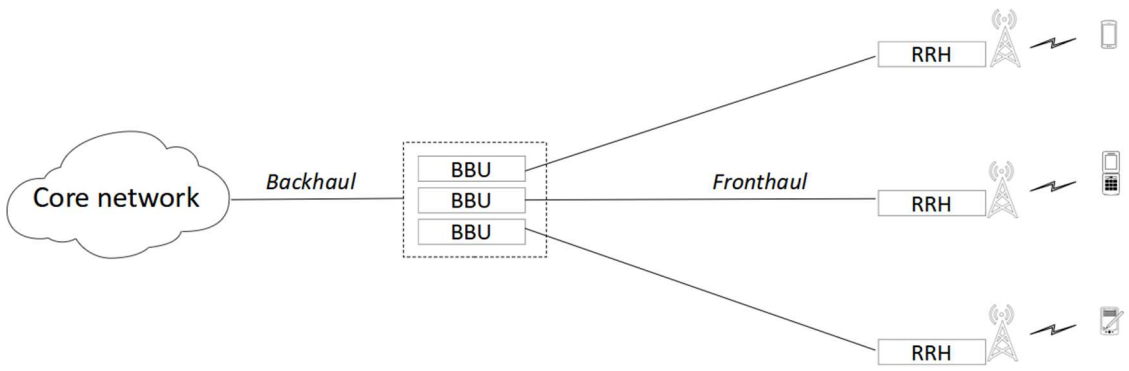


Figure 2.5: Centralised RAN (C-RAN) architecture.

The benefits of the C-RAN architecture to mobile operators and equipment vendors for future mobile networks include [32], [33]:

- Reduction of operational and capital expenditure (OPEX and CAPEX) – By implementing C-RAN, the number of BBUs at the cell sites are reduced leading to lower energy consumption and cost of base station deployment.

- The centralisation of BBUs makes coordinated and cooperative processing techniques such as coordinated multi-point (CoMP) and enhanced inter-cell interference coordination (eICIC) much easier to implement. This leads to increased throughput because of lower interference levels and decrease in the time needed to perform handovers.

- The flexible allocation of radio resources for different network traffic loads among the connected RRHs leads to a statistical multiplexing gain.

A **virtualised RAN (vRAN)**, is an implementation of the centralised RAN whereby the baseband processing in the BBU pool at the central location is virtualised, as shown in Figure *2.6*. The virtualised BBUs run on commercial-off-the-shelf (COTS) hardware and the computing resources in the BBU pool can dynamically be shared among the cells in a mobile network.

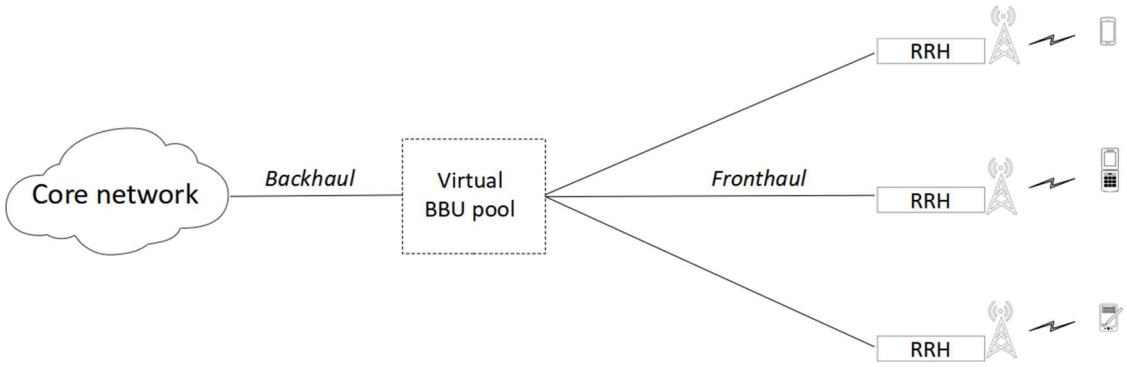


Figure 2.6: Virtualised RAN (vRAN) architecture.

The 5G transport network is made of three segments: fronthaul, midhaul and backhaul. In 5G, the BBU is now split into two network elements – the Distributed Unit (DU) and the Central Unit (CU), which form the 5G base station (gNB). The RRH has been replaced with the network element known as the Radio Unit (RU) and the core network is known as 5G Core (5GC). The transport network connects the various 5G network elements (5GC, CU, DU and RU). The **backhaul** connects the 5GC to the CU, the **midhaul** connects the CU with the DU and the **fronthaul** connects the DU with the RU [34], as illustrated in Figure *2.7*. Together, the three network segments are referred to as an **xHaul** network [35]. The key transport options that may be used within the 5G access network include dark fibre, passive and active WDM, Ethernet (Ethernet over dark fibre, Flex Ethernet and G.mtn), OTN, microwave and millimetre wave radio transmission, PON and free space optics [31]. The split architecture in 5G poses new challenges for the xHaul network since fronthaul, midhaul and backhaul traffic must meet different latency (delay) requirements in the same network. The maximum one-way delay requirement for fronthaul is 250 µs. The midhaul and backhaul latency requirements are similar, with a delay requirement of 1.5 ms – 10 ms for midhaul and 1 ms – 50 ms for backhaul [36],

[37]. The delay constraint for midhaul and backhaul are derived mainly from the target service's (eMBB, URLLC and mMTC) latency requirements, as listed in Table *2.1*.

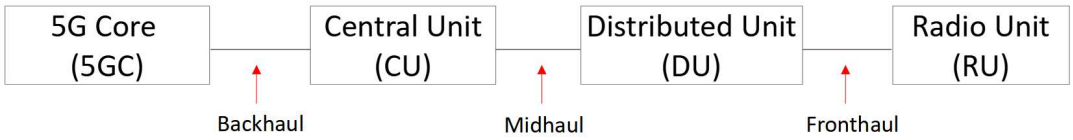


Figure 2.7: 5G transport network.

With the introduction of the fronthaul in the centralised RAN architecture, several interface specifications were defined for communicating between the BBU and RRH. The three main interfaces are the Common Public Radio Interface (CPRI) [38], the Open Base Station Architecture Initiative (OBSAI) [39] and the Open Radio Interface (ORI) [40]. CPRI was designed by the leading telecommunication equipment vendors (Ericsson, Huawei, NEC Corporation and Nokia) and has been adopted by most telecom operators.

The CPRI specification defines the structure of a CPRI frame used for the transfer of user plane data, control and management data as well as synchronisation data between RU and DU. CPRI is a constant bit rate interface which transmits the digitised in-phase (I) and quadrature-phase (Q) waveforms of mobile signals between RUs and DUs. The CPRI specification defines eleven line rates, ranging from 614.4 Mbit/s (option 1) to 24.33 Gbit/s (option 10) [38]. The lower rate options (1 to 7) use the 8B/10B encoding scheme while the higher rate options (7A to 10) use 64B/66B encoding. The 64B/66B encoding is much more efficient (3% overhead) than the 8B/10B encoding (20% overhead) [41].

To transport digitised radio samples over a fronthaul requires the sampled output of the inverse fast Fourier transform (iFFT) to be quantised prior to it being framed for transportation [42]. The I and Q samples are quantised with a 16-bit resolution and then inserted into the payload section of a generic framing structure. The required data rate that will need to be accommodated by the fronthaul per physical antenna port without including the line coding and control overheads of the transport interface is given by equation 2.1 [43]:

$$\text{Data rate} = 2(I/Q)\ x\ S_r\ x\ N\ \ x\ f \tag{2.1}$$

where:

2(I/Q) is a multiplication factor for in-phase (I) and quadrature-phase (Q) data;

S_r is the sampling rate used for digitisation (sample/s/carrier, e.g., 15.36 MHz per 10MHz LTE bandwidth, 30.72 MHz per 20 MHz LTE bandwidth);

N is the sample width (bits/sample);

f is a carrier aggregation factor normalised to a 20 MHz channel.

With the addition of the overhead information in the CPRI frames, the CPRI data rate is given by equation 2.2 [43]:

$$\text{Data rate} = M\ x\ S_r\ x\ N\ x\ 2(I/Q)\ x\ C_w\ x\ C \tag{2.2}$$

where:

M is the number of antennas per sector;

S_r is the sampling rate used for digitisation (sample/s/carrier, e.g., 15.36 MHz per 10MHz LTE bandwidth, 30.72 MHz per 20 MHz LTE bandwidth);

N is the sample width (bits/sample);

2(I/Q) is a multiplication factor for in-phase (I) and quadrature-phase (Q) data;

C_w is the factor of CPRI control word;

C is the coding factor (either 10/8 for 8B/10B line coding or 66/64 for 64B/66B line coding).

For a 20 MHz LTE carrier and 2 x 2 MIMO (*M* = 2, S_r = 30.72 MHz, *N* = 15, C_w = 16/15 and *C* = 10/8), the data rate is 2.4576 Gbit/s, which corresponds to line rate option 3 in [38].

$$\text{Data rate} = 2 \text{ x } 30.72\ MHz \text{ x } 15 \text{ x } 2 \text{ x } \frac{16}{15} \text{ x } \frac{10}{8} = 2.4576\ Gbit/s$$

and for a 20 MHz LTE-Advanced carrier and 4 x 4 MIMO (*M* = 4, S_r = 30.72 MHz, *N* = 15, C_w = 16/15 and *C* = 10/8), the data rate is 4.9152 Gbit/s, which corresponds to line rate option 5 in [38].

$$\text{Data rate} = 4 \text{ x } 30.72\ MHz \text{ x } 15 \text{ x } 2 \text{ x } \frac{16}{15} \text{ x } \frac{10}{8} = 4.9152\ Gbit/s$$

The CPRI line rate is constant regardless of the amount of traffic being transmitted. With the use of more bandwidth and a higher number of antenna ports and MIMO layers, the data rates of 5G networks are expected to be potentially 1000x more than the existing LTE bandwidth. Table *2.2* shows the CPRI data rate requirements using various parameter values [43].

Table 2.2: CPRI data rates for LTE and 5G [43].

Channel Bandwidth (MHz)	**Sample rate (MHz)**	**CPRI data rate (Gbit/s)**					
		Antennas per sector					
		1	*2*	*4*	*8*	*16*	*64*
20	30.72	1.2288	2.4576	4.9152	9.8304	16.2202	64.8806
40	61.44	2.4576	4.9152	9.8304	16.2202	32.4403	129.7613
80	122.88	4.9152	9.8304	16.2202	32.4403	64.8806	259.5226
100	153.6	6.1440	10.1376	20.2752	40.5504	81.1008	324.4032

In Table *2.2*, only the data rates with values equal to or less than 10.1376 Gbit/s (in green) are currently supported by the CPRI specification [38]. To support data rates listed in Table *2.2* greater than 10.1376 Gbit/s (in yellow), solutions such as defining new interface points ('splits') within the radio protocol stack or using data compression techniques have been proffered [44], [45], [46], [47].

Another solution is using Ethernet as a transport protocol in the fronthaul [48]. One of the ways to handle CPRI traffic in a packet fronthaul network is by encapsulating CPRI traffic directly over Ethernet [49]. In 2017, the CPRI cooperation group released an enhanced version of CPRI, known as eCPRI [50], to meet the needs of 5G networks and provide support for packet-based Ethernet and IP fronthaul transport networks.

There are various challenges associated with using Ethernet in the fronthaul that must be overcome. One of these is the issue of timing and synchronisation between different RRHs. These challenges were studied by the EU Horizon 2020 research project – The intelligent Converged network consolidating Radio and optical access aRound User equipment (iCIRRUS) [48], [51] from 2015 to 2017. Ethernet is not a synchronous protocol like CPRI, so in order to meet the demands of 5G networks, synchronisation

must be provided for mobile traffic that will be transmitted through an asynchronous Ethernet network. There are several ways to achieve synchronisation over Ethernet and they are discussed in section 2.2.2.

2.2.1 Latency Requirements

The user plane latency is the contribution of the radio network to the time from when the source sends a packet to when the destination receives it [52]. The key issues to consider for latency in a fronthaul network are the network entry or physical random access channel (PRACH) procedure and the hybrid automatic repeat request (HARQ) protocol [47]. PRACH is used to access the network by the UE. The main latency constraint is imposed by the HARQ protocol, which is used to handle transmission errors between the UE and the BBU. HARQ is a combination of the Automatic Repeat Request (ARQ) and Forward Error Correction (FEC) schemes. It is used to check for errors in which the receiver sends an Acknowledgement (ACK) or a Negative Acknowledgement (NACK) back to the transmitter to indicate if an error or no error is detected in the received subframe.

According to the HARQ timing requirement in an LTE network, the UE should receive an ACK/NACK from the eNodeB (RRH and BBU at the cell site) in three subframes, i.e., in the fourth subframe after sending uplink data. Otherwise, the UE retransmits the data. Given that the transmission of an LTE subframe is 1 ms, the period between the time of the *n* subframe and the time of the *n + 4* subframe is 5 ms, of which 2 ms are for uplink and downlink data transmission, as shown in Figure *2.8*. Therefore, only 3 ms is available for the additional delays and baseband processing time at the eNodeB.

In the C-RAN architecture, the fronthaul network between the RRH and the BBU is mostly implemented using optical fibre. The separation of the RRH and the BBU in the fronthaul network introduces additional delays such as delay due to optical transmission and delay by active equipment in the fronthaul network (e.g., active wavelength division multiplexing, passive optical networks). Figure *2.8* shows the timing diagram for the downlink and uplink HARQ process in a C-RAN.

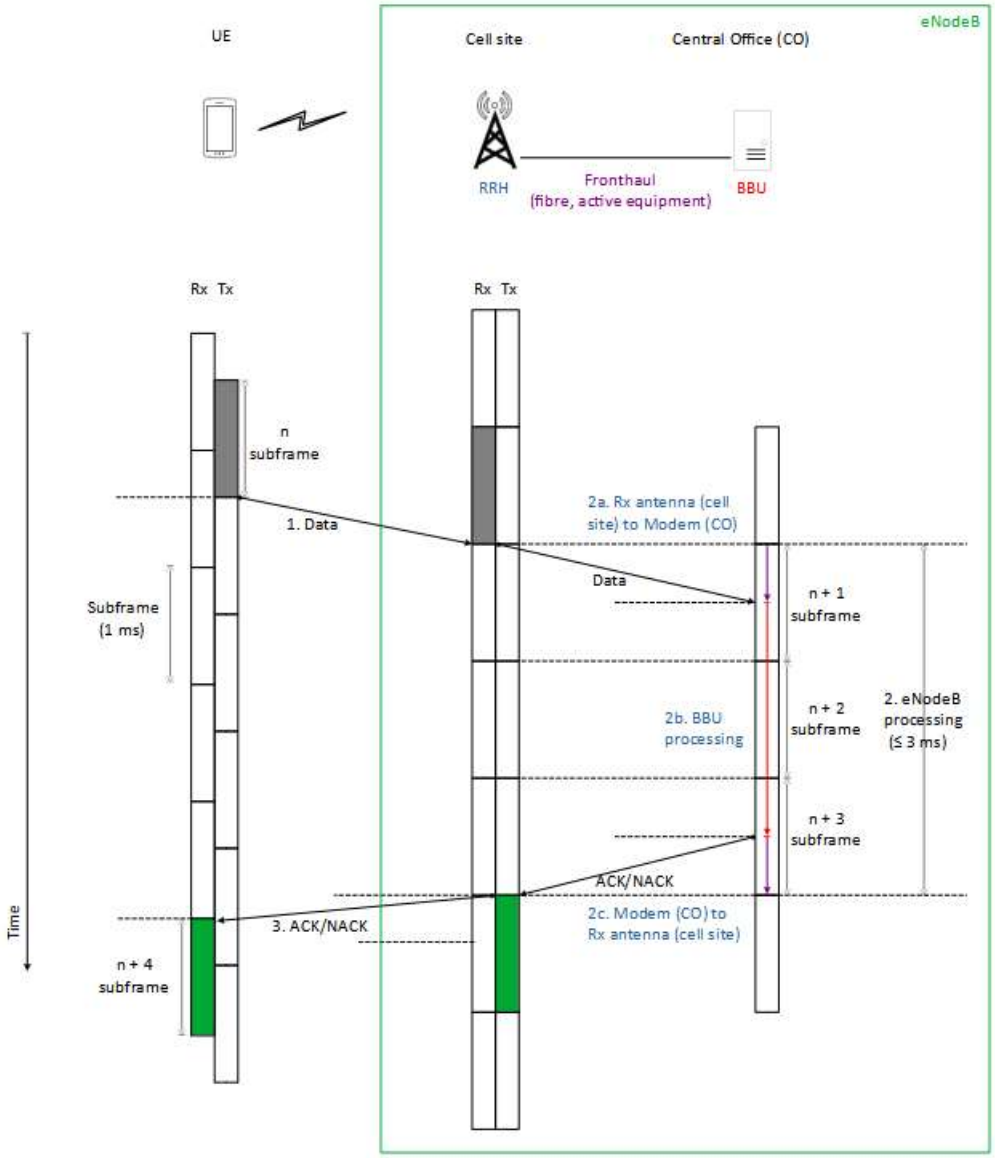


Figure 2.8: LTE HARQ timing diagram in a C-RAN [53].

In order to maintain the timing requirement for eNodeB processing, the telecommunication equipment vendors design the eNodeB to complete the processing and send ACK/NACK within 2.75 ms, instead of 3 ms [53]. Therefore, about 250 µs can be allowed in the fronthaul network. The delay components occurring in the fronthaul network are listed in Table *2.3*.

Table 2.3: Delay components in the fronthaul network (uplink latency) [53].

2a. Rx antenna (cell site) to Modem (CO)	**2b. BBU processing (CO)**	**2c. Modem (CO) to Rx antenna (cell site)**
(i) RRH/RF processing (UL)	(v) BBU/CPRI processing	(x) Optical fibre latency (BBU to RRH)
(ii) RRH/CPRI processing (UL)	(vi) UL frame decoding in PHY layer	(xi) Active equipment processing
(iii) Optical fibre latency (RRH to BBU)	(vii) ACK/NACK creation in MAC layer	(xii) RRH/CPRI processing (DL)
(iv) Active equipment processing	(viii) DL frame creation in PHY layer	(xiii) RRH/RF processing (DL)
	(ix) BBU/CPRI processing	

In a fronthaul network built with active wavelength division multiplexing (WDM), the delay components involved in the data transmission after the RRH receives data from the UE and before it sends an ACK/NACK to the UE are listed in Table *2.4* below [53].

Table 2.4: Optical Fibre Distance between RRH (cell site) and BBU (CO) [53].

Delay Components	**Related Network Equipment**	**Description**	**Typical values**
(a) Round trip RF processing time	RRH	i + xiii	≈ 25 – 40 µs
(b) Round trip CPRI processing time	RRH, BBU	ii + v + ix + xii	≈ 10 µs
(c) BBU round trip baseband processing time	BBU	vi + vii + viii	≈ 2700 µs
(d) Fronthaul equipment round trip processing delay	Fronthaul requirements	iv + xi	≈ 40 µs (OTN encapsulation) ≈ few µs (non-OTN encapsulation)

To determine how the optical fibre delay for non-OTN encapsulation (250 µs) is obtained, the sum of the delay components is subtracted from the eNodeB processing time (3 ms) and is calculated in equation 2.3 [53]:

$$Optical\ fibre\ delay = eNodeB_{processing} - \Sigma[(a) + (b) + (c) + (d)] \quad (2.3)$$

where:

eNodeB$_{processing}$ is the eNodeB processing time = 3000 µs;

(a) is the round trip RF processing time = 40 µs (Table 2.4);

(b) is the round trip CPRI processing time = 10 µs (Table 2.4);

(c) is the BBU round trip baseband processing time = 2700 µs (Table 2.4);

(d) is the fronthaul equipment round trip processing delay = 4 µs (Table 2.4).

$$Optical\ fiber\ delay = 3000\ \mu s - (40\ \mu s + 10\ \mu s + 2700\ \mu s + 4\ \mu s)$$

$$= 3000\ \mu s - 2754\ \mu s$$

$$= 246\ \mu s$$

$$\approx 250\ \mu s$$

From the above calculation, the one-way latency over the fronthaul network is expected to be 250 µs or less. Due to the 250 µs latency requirement in the fronthaul, the maximum separation distance between the RRH and the BBU is limited and can be calculated as shown in equation 2.4 [53]:

$$Optical\ fibre\ distance = \frac{optical\ fibre\ delay}{\left(5\frac{\mu s}{km}\right)\ x\ 2} \quad (2.4)$$

$$= \frac{246\ \mu s}{10\ \frac{\mu s}{km}}$$

$$= 24.6\ km$$

$$\approx 25\ km$$

The value of the one-way transmission delay time per km (≈ 5 µs/km) is obtained using the speed of light in optical fibre, as given in equation 2.5 [54]:

$$Speed\ of\ light\ in\ a\ vacuum\ (c) = 299{,}792{,}458\ \mathrm{m/s} = \ 299.792458\ \mathrm{m/\mu s} \qquad (2.5)$$

A refractive index of 1.47 can typically be used for fused silica or quartz which is found in optical fibre. This value takes into account two types of single mode optical fibres (G.652 and G.655) for three optical transmission windows (wavelength, λ = 1310 nm, 1550 nm and 1625 nm) [55] [56] and is used in equation 2.6:

$$Refractive\ index\ of\ a\ medium\ (n) = \frac{speed\ of\ light\ in\ a\ vacuum\ (c)}{speed\ of\ light\ in\ that\ medium\ (v)} \qquad (2.6)$$

$$Speed\ of\ light\ in\ that\ medium\ (v) = \frac{speed\ of\ light\ in\ a\ vacuum\ (c)}{refractive\ index\ of\ a\ medium\ (n)}$$

$$= \frac{299.792458\ m/\mu s}{1.47}$$

$$= 203.94\ m/\mu s$$

$$\frac{1}{v} = 0.0049033\ \frac{\mu s}{m} = 0.0049033\ x\ 10^3\ \mathrm{\mu s/km}\ = 4.9033\ \mu s/km\ \approx 5\ \mu s/km$$

To meet the latency requirement, the fronthaul distance should not exceed 25 km. For most telecom network operators, a typical 5G fronthaul link is limited to 10 km [57].

Different values for the one-way latency can be found in industry reports, white papers and standards published by the 3GPP [36], CPRI cooperation group [58], the O-RAN Alliance [37], the Institute of Electrical and Electronics Engineers (IEEE) [59], the NGMN Alliance [60] and the Small Cell Forum (SCF) [61]. The value only depends on the specific vendor implementation of the HARQ loop [62]. However, there is agreement on the latency requirement for user plane latency (one way) for eMBB which is 4 ms [62]. The importance of latency in a network is key to deploying current and future time-sensitive and mission-critical applications such as industrial IoT in factories, cloud gaming, intra-rack communications in edge data centres, holographic-type communications and vehicle-to-everything (V2X) communication. In this thesis, the 250 µs latency value is used in the simulation model that will be discussed in Chapters 4 and 5.

2.2.2 Synchronisation Requirements

Precise synchronisation and timing are required for the reliable transmission of data in a mobile network. Globally, mobile networks are equipped with a high precision clock which has an accurate primary reference. The reference is in the form of signals transmitted by global navigation satellite systems (GNSS) such as the United States' Global Positioning System (GPS), Russia's Global Navigation Satellite System (GLONASS), the European Union's Galileo, China's BeiDou Navigation Satellite System (BDS), Japan’s Quasi-Zenith Satellite System (QZSS) and the Indian Regional Navigation Satellite System

(IRNSS) [63] [64]. Synchronisation requirements can be classified into three different types: frequency, phase and time [65].

Frequency synchronisation

Synchronisation in the frequency domain relates to the alignment of clocks in frequency, i.e., the leading edges of the clock pulses occur at the same time [65]. It also refers to the frequency accuracy measured by comparing the actual frequency generated by the base station in the air interface (carrier frequency) to a Primary Reference Clock (PRC) or Primary Reference Source (PRS) based on a GNSS receiver. This frequency accuracy is measured in units of ppb (parts per billion) or ppm (parts per million).

Frequency synchronisation is always required in mobile networks, typically in order to allow handover of the mobile devices or user equipment (UE) between cells [66]. In a 5G FDD network, the carrier frequency must be within 50 ppb of the allocated frequency for wide area base stations (macro cells) and ±100 ppb for medium range (micro cells) or local area (pico-cells) base stations [65] [67]. For example, using a 700 MHz frequency band, the frequency accuracy requirement of 50 ppb means that the carrier frequency could be anywhere between 699,999,965 Hz and 700,00,035 Hz.

For the deployment of 2G and 3G mobile networks, only frequency synchronisation was required. With 4G, requirements for phase and time were added to enable advanced features such as coordinated multi-point (CoMP) [68] and enhanced inter-cell interference coordination (eICIC) to be included [19].

Phase synchronisation

Phase synchronisation refers to the process of aligning clocks in a telecommunication network with respect to phase (i.e., the time difference between two events) [69]. It is typically represented in microseconds (μs) or nanoseconds (ns). Phase synchronisation is needed to support requirements for the air interface of time division duplexing (TDD) mobile networks [63]. For 5G TDD networks, base stations must be synchronised to within 3 μs of each other. This cell phase synchronisation accuracy is defined as applying at the base station antenna connectors [70].

Time synchronisation

Time synchronisation refers to the distribution of a time reference to different nodes in a telecommunication network [69] and the stringent requirements apply to the generation of signals over the air interface of mobile networks [63].

There are various ways to achieve synchronisation over Ethernet in the fronthaul. They include using the IEEE 1588 Precision Time Protocol (PTP) [71] and/or the ITU-T Synchronous Ethernet (SyncE) [72]. IEEE 1588 provides both frequency and phase synchronisation while SyncE can only deliver frequency synchronisation.

2.2.3 Reliability Requirements

Reliability relates to the capability of transmitting a given amount of traffic within a predetermined time duration with a high success probability [52]. The reliability requirement for transmitting a 32-byte packet in URLLC use cases, such as industrial automation, intelligent transportation and remote healthcare, is $1 - 10^{-5}$ (99.999%) with

a user plane latency of 1 ms [52]. For eMBB use cases, such as enhanced multimedia and consumer entertainment, the reliability requirement is $1 - 10^{-3}$ (99.9%) [73] with a user plane latency of 4 ms (one way). The reliability requirements for both use cases are relevant when considering the proportion of frames meeting the latency requirements, as will be discussed in section 5.4.2.

Transport networks are designed to achieve very low bit error rate (BER) [74]. According to the CPRI/eCPRI specification, the maximum allowed BER and one-way frame loss rate for fronthaul is 10^{-12} and 10^{-7}, respectively [38] [58].

2.3 5G RAN Architecture

The 5G RAN (NG-RAN) architecture, as illustrated in Figure *2.9* [75], consists of a set of 5G radio base stations (known as gNBs) connected to the 5G core network (5GC) via the *NG* logical interface. The gNBs are interconnected through the *Xn* logical interface. The gNB is further split into two units: the Central Unit (CU) and the Distributed Unit (DU) which are connected via the *F1* logical interface.

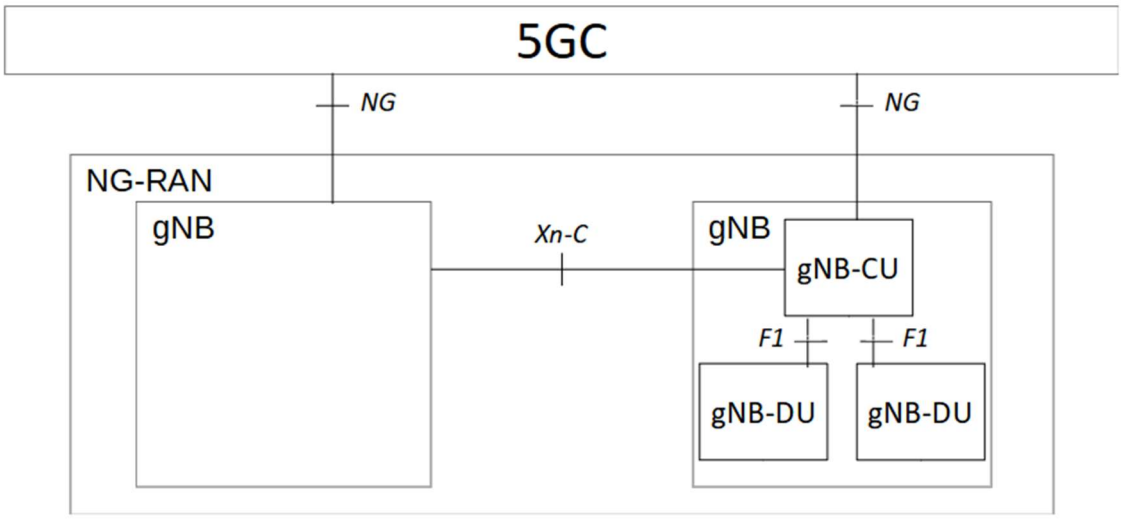


Figure 2.9: 5G RAN architecture [75].

In section 2.2, the terms fronthaul, midhaul and backhaul, were described. These three terms can be mapped to the three logical interfaces (*NG*, *Xn* and *F1*) used in the 5G RAN architecture, as shown in Table *2.5*.

Table 2.5: 3GPP and ITU-T terminology for RAN interfaces [76].

3GPP logical interface or CPRI name	**ITU-T transport network name**	**Interface description**
CPRI/eCPRI	Fronthaul	Interface between RU and CU
F1	Midhaul	Logical interface between the CU and DU
NG	Backhaul	Logical interface between the gNB (CU) and the 5GC
Xn	Midhaul or backhaul	Logical interface between gNB nodes

2.3.1 Radio Protocol Stack

Two different radio protocol stacks exist in the 5G architecture: user plane protocol stack and control protocol stack, as illustrated in Figure *2.10* and Figure *2.11*, respectively [77]. The user plane protocol stack carries user data between different applications in the UE and gNB. The control plane protocol stack carries control information between the UE and 5G core network.

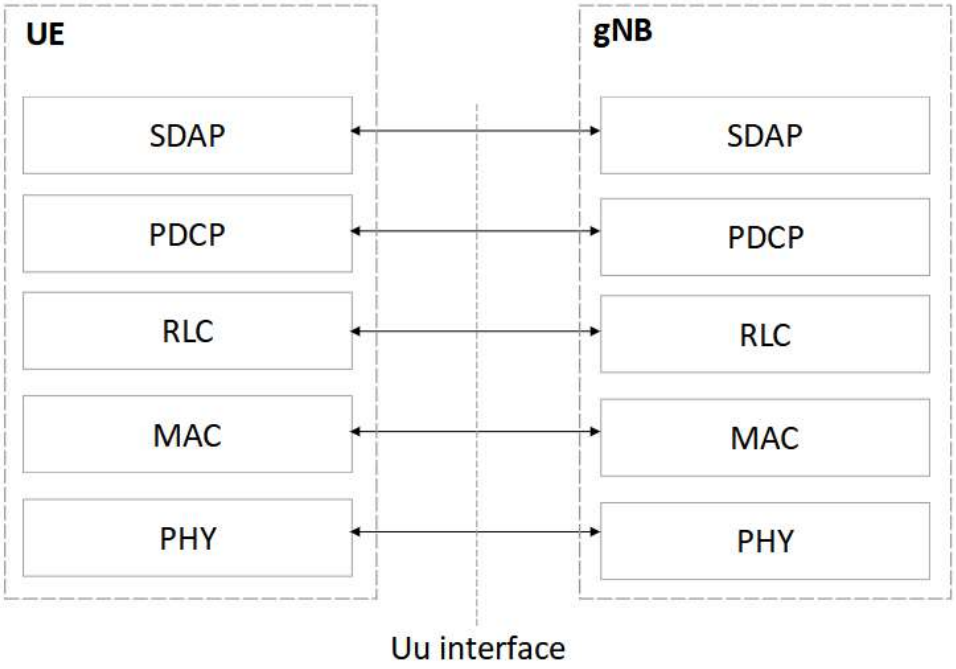


Figure 2.10: 5G NR user plane protocol stack [77].

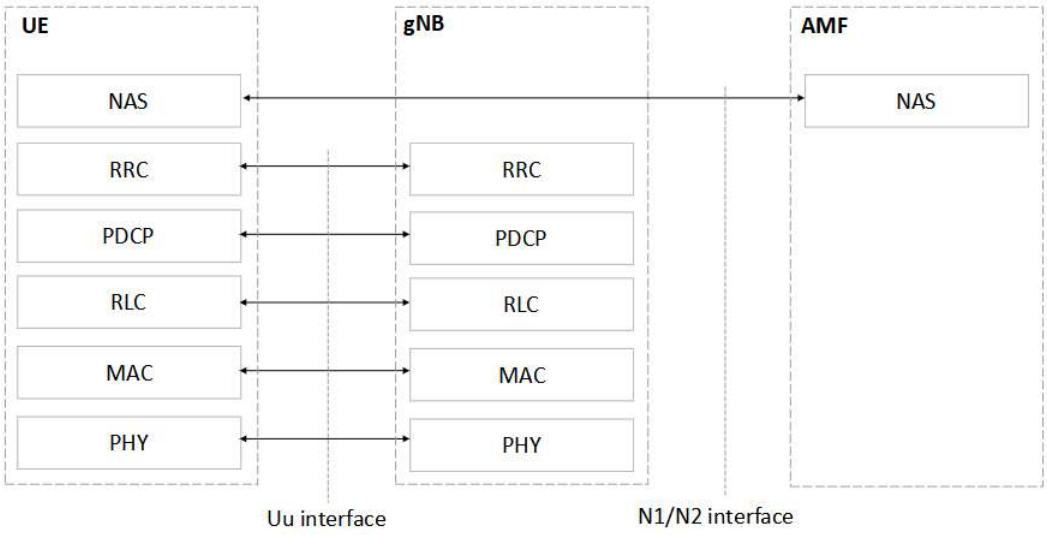


Figure 2.11: 5G NR control plane protocol stack [77].

The NR radio protocol stack is slightly different for the user plane and the control plane. The PHY, MAC, RLC and PDCP layers are common for both user and control plane, while the RRC layer only resides in the control plane. On top of the RRC is the Non-Access Stratum (NAS) layer, which terminate in the UE and the AMF (Access and Mobility Management Function) of the 5G core network but not in the gNB. In 5G, a new user plane layer has been added above the PDCP, named the Service Data Adaptation Protocol (SDAP).

The arrowed horizonal lines in Figure *2.10* and Figure *2.11* depict the logical connections between the UE and the gNB for each protocol using 5G NR interfaces. The UE connects to the gNB at layers 1 – 3 via the Uu interface. N1 is the interface between the UE and the AMF through the gNB. It is used to transfer UE information (related to connection, mobility and sessions) to the AMF. N2 is the interface between the gNB and the AMF. It handles control-plane signalling.

Due to the gNB split architecture, the CU hosts all the upper layer protocols: RRC, SDAP and PDCP while the DU hosts all the lower layer protocols: RLC, MAC and PHY. In the 3GPP standards [77], 5G NR protocol stack is made up of three layers. The 5G layer 1 is the PHY (physical) layer. The 5G layer 2 includes MAC, RLC, PDCP. The 5G layer 3 is the RRC layer. As shown in Figure *2.12*, layer 1 and layer 2 in the 5G NR protocol stack is mapped to the data link and physical layers in the seven-layer Open Systems Interconnection (OSI) network reference model [41] [78].

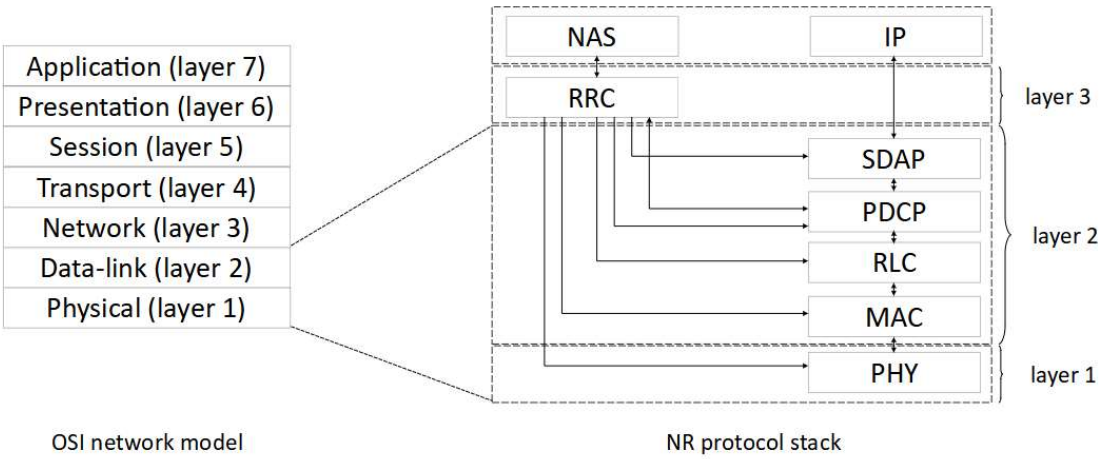


Figure 2.12: Mapping of 5G NR protocol stack (user plane and control plane) to OSI network reference model [34].

The protocols in the control and user planes are described below:

- **Physical layer (PHY)** [79]

  It is responsible for radio related functions such as frequency/time synchronisation, modulation/demodulation, encoding/decoding, and MIMO antenna processing. The physical layer offers services to the MAC layer via the transport channels.

- **Medium Access Control (MAC)** [80]

  This is the layer located below the RLC layer and above the PHY layer. It is responsible for mapping between logical and transport channels, multiplexing and demultiplexing of MAC Service Data Units (SDUs), error correction using the Hybrid ARQ (HARQ) technique and the scheduling of radio resources for both uplink and downlink. The MAC layer provides services to the RLC layer via the logical channels.

- **Radio Link Control (RLC)** [81]

  This is the layer located below the PDCP layer and above the MAC layer. It is responsible for the concatenation, segmentation and reassembly of data/IP packets known as RLC SDUs. It also performs error correction using the Automatic Repeat

reQuest (ARQ) technique. The RLC layer provides services to the PDCP layer via RLC channels.

- **Packet Data Convergence Protocol (PDCP)** [82]

This is the layer located below the RRC layer and above the RLC layer. It is responsible for header compression of IP data flows using the Robust Header Compression (ROHC) protocol. It also handles ciphering and deciphering of user and control plane data, and integrity protection of control plane data. The PDCP layer provides services to the SDAP layer via radio bearers.

- **Radio Resource Control (RRC)** [83]

This is the layer located above the PDCP layer. It is responsible for handling all the control plane signalling between the UE and the eNodeB. These signalling messages are needed to regulate the UE behaviour in order to comply with the network procedures.

- **Non-Access Stratum (NAS)** [84]

This is the highest layer in the control plane and sits above the RRC. The NAS protocol supports the mobility of the UE and the session management procedures to establish and maintain the IP (Internet Protocol) connectivity between the UE and a packet data network gateway.

- **Service Data Application Protocol (SDAP)** [85]

This is a new layer introduced in the NR user plane and sits above the PDCP. Its main functions are the mapping between Quality of Service (QoS) flows and data radio bearers (DRBs), and marking QoS flow identifiers (IDs) in downlink (DL) and uplink (UL) packets. The SDAP layer provides services to the 5G core via QoS flows.

2.3.2 RAN Functional Split

In order to reduce the fronthaul bandwidth required by the CPRI protocol, functional splits were introduced in the RAN. The 3GPP standards organisation defined eight functional split options for the 5G NR protocol stack [36], as illustrated in Figure *2.13*. Each functional split comes with different requirements for the fronthaul transport protocol based on the functions each layer performs [19], as described in section 3.3.1.

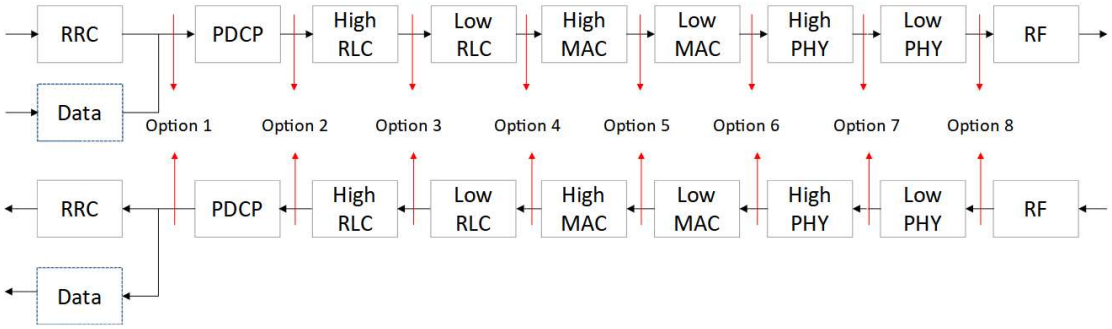


Figure 2.13: 5G RAN functional split architecture [36].

Two functional split groups were identified by 3GPP based on their latency requirements – a high layer split (HLS) and a low layer split (LLS). Options 1 to 3 make up the HLS group, located within the non-real time functions of the RAN, and options 4 to 8 form the LLS group, which are located within the real time functions of the RAN, i.e., has to operate synchronously to the radio [62]. Split options 1, 2 and 3 have the most relaxed latency (1 – 10 ms) requirements, similar to backhaul [86], while split options 4 to 8 have more stringent latency requirements (< 250 μs).

One split option was chosen for each group – Option 2 for HLS and the sub-options of Option 7 (7.x) for LLS, e.g., Option 7-2x by the O-RAN Alliance [87], as shown in Figure *2.14*.

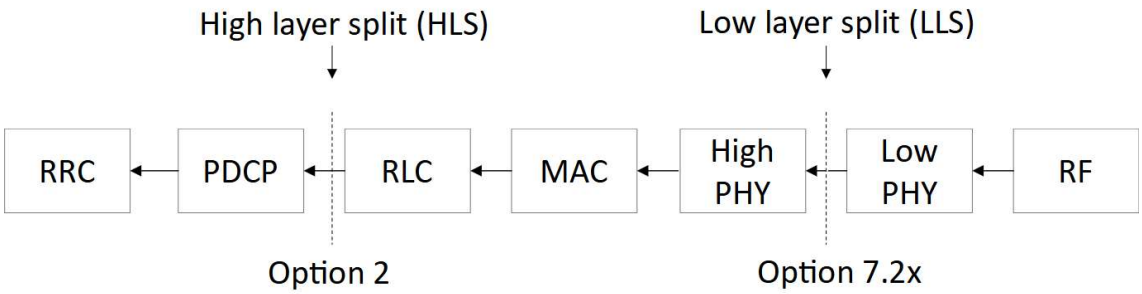


Figure 2.14: High layer and low layer splits.

The eight functional split options between the CU and DU, shown in Figure *2.13*, are described below:

- **Option 1**

  In this option, also referred to as RRC – PDCP split, the RRC is in the CU. The PDCP, RLC, MAC, physical layer and RF are in the DU, thus the entire user plane is in the DU.

- **Option 2**

  In this split option, also referred to as PDCP – RLC split, the RRC, PDCP are in the CU. The RLC, MAC, physical layer and RF are in the DU.

- **Option 3**

  In this option, also referred to as High RLC – Low RLC split, the Low RLC (partial function of RLC), MAC, physical layer and RF are in DU. The PDCP and High RLC (the other partial function of RLC) are in the CU. There are two sub-options – Option 3.1 and Option 3.2.

- **Option 4**

In this split option, also referred to as RLC – MAC split, the MAC, physical layer and RF are in DU. The PDCP and RLC are in the CU.

- **Option 5**

In this option, also referred to as High MAC – Low MAC split or Intra-MAC split, the RF, physical layer and lower part of the MAC layer (Low MAC) are in the DU. The higher part of the MAC layer (High MAC), RLC and PDCP in the CU. Therefore, by splitting the MAC layer into 2 entities (e.g., High MAC and Low MAC), the services and functions provided by the MAC layer will be located in the CU, in the DU, or in both.

- **Option 6**

In this split option, also referred to as MAC – PHY split, the MAC and upper layers are in the CU. The PHY layer and RF are in the DU.

- **Option 7**

In this option, also referred to as High PHY – Low PHY split, the split point is within the physical layer. Part of the physical layer function and RF are in the DU while the upper layers are in the CU. Multiple realisations of this option are possible, including asymmetrical options which allow to obtain benefits of different sub-options for UL and DL independently.

    - **Option 7-3 (Only for DL)**

      This option is only applicable to the DL. The encoder is in the CU and the rest of the PHY functions are in the DU.

- **Option 7-2**

  In this option, the DU performs Fast Fourier Transform (FFT), cyclic prefix (CP) removal, and resource element mapping in UL, and inverse FFT (iFFT), CP insertion, resource element mapping and precoding in DL. All other functions are performed in the CU.

- **Option 7-1**

  In this option, the DU performs FFT and CP removal in UL and iFFT, and CP insertion in DL. All other functions are performed in the CU.

- **Option 8**

In this option, also referred to as PHY – RF split, the RF functionality is in the DU and the upper layers are in the CU. This split permit centralisation of processes at all protocol layer levels, resulting in very tight coordination of the RAN. This allows efficient support of functions such as coordinated multi-point (CoMP), multiple-input, multiple-output (MIMO), load balancing, and mobility.

Table *2.6* shows a summary of the functional split options between the CU and DU. As can be seen, the lower the split option, the less functions are in the CU and the higher the split option, the less functions are in the DU.

Table 2.6: Functional split options between the CU and DU.

Split Option	Central Unit (CU)	Distributed Unit (DU)	Description
Option 1	RRC	PDCP RLC MAC PHY RF	PDCP – RLC split
Option 2	PDCP RRC	RLC MAC PHY RF	PDCP – RLC split
Option 3	High-RLC PDCP RRC	Low-RLC MAC PHY RF	High RLC – Low RLC split
Option 4	RLC PDCP RRC	MAC PHY RF	RLC – MAC split
Option 5	High-MAC RLC PDCP RRC	Low-MAC PHY RF	High MAC – Low MAC split or Intra-MAC split
Option 6	MAC RLC PDCP RRC	PHY RF	MAC – PHY split
Option 7	High-PHY MAC RLC PDCP RRC	Low-PHY RF	High PHY – Low PHY split
Option 8	PHY MAC RLC PDCP RRC	RF	PHY – RF split

Table *2.7* summarises the latency and bandwidth requirements for the eight split options, as defined by 3GPP [36], for the downlink (DL) and uplink (UL). Note that according to ITU-T [62], the Options 7a, 7b and 7c used in Table *2.7* are not equivalent to Options 7-1, 7-2 and 7-3.

Table 2.7: Latency and bandwidth requirements for RAN functional splits [36].

Split option	Required bandwidth (Downlink)	Required bandwidth (Uplink)	Max one-way latency
Option 1	4 Gbit/s	3 Gbit/s	10 ms
Option 2	4.016 Gbit/s	3.024 Gbit/s	1.5 – 10 ms
Option 3	Lower than Option 2 for uplink/downlink		1.5 – 10 ms
Option 4	4 Gbit/s	3 Gbit/s	≈ 100 µs
Option 5	4 Gbit/s	3 Gbit/s	Hundreds of µs
Option 6	4.133 Gbit/s	5.640 Gbit/s	250 µs
Option 7a	10.1 – 22.2 Gbit/s	16.6 – 21.6 Gbit/s	250 µs
Option 7b	37.8 – 86.1 Gbit/s	53.8 – 86.1 Gbit/s	250 µs
Option 7c	10.1 – 22.2 Gbit/s	53.8 – 86.1 Gbit/s	250 µs
Option 8	157.3 Gbit/s	157.3 Gbit/s	250 µs

The parameters used for the required bandwidth in Table *2.7* are listed in Table *2.8*.

Table 2.8: Parameters for bandwidth requirements [88].

Channel bandwidth (MHz)	Modulation scheme	Number of MIMO layers	IQ bitwidth (bit)	Number of antenna ports
100 (DL/UL)	DL: 256 QAM UL: 64 QAM	DL: 8 Ul: 8	DL: 2x (7 – 16) UL: 2x (10 – 16)	32 (DL/UL)

The Small Cell Forum (SCF) adopted Option 6 as the basis for its 5G network functional application platform interface (5G nFAPI) used in small cell deployments [89]. The O-RAN Alliance adopted Option 7-2x for its Open Fronthaul Interface [87]. Option 7-2x has two variants based on where the precoding operation occurs. If precoding and resource element mapping take place in the DU and the RU handles beamforming, then it is a 7-2a split. If precoding happens in the RU, then it is a 7-2b split.

The method for calculating the fronthaul bandwidth (BW) for split option 7-2x is based on the O-RAN Alliance specifications [37]. Fronthaul bandwidth is dimensioned per cell site, where a number of carriers are deployed at different frequency bands (Low- and

mid-bands in Sub-6 and millimetre wave in high band), and at different sectors. The fronthaul bandwidth for a cell site depends on the following key parameters:

- Number of sectors
- Radio channel bandwidth of each carrier
- MIMO order of each carrier

Various sizes of site configuration may be considered for the eCPRI fronthaul traffic:

- Small: single sector, carriers in either millimetre wave or Sub-6 frequency range with low MIMO order
- Medium: multiple sectors, carriers in both Sub-6 and millimetre wave frequency range with medium MIMO order
- Large: multiple sectors, carriers in both millimetre wave and Sub-6 frequency range with Massive MIMO

The millimetre wave frequency band consists of the following frequencies: 24.25 – 27.5 GHz, 37 – 43.5 GHz, 45.5 – 47 GHz, 47.2 – 48.2 and 66-71 GHz while Sub-6 are frequencies under 6 GHz, made up of low-band (< 1 GHz) and mid-band (2.5 GHz, 3.5 GHz, and 3.7 – 4.2 GHz).

Table *2.9* and Table *2.10* provides typical peak fronthaul bandwidths for the various sizes of site configurations. CBW is the aggregated channel bandwidth of all carriers in a sector, ABW is the antenna bandwidth (= number of sectors x number of MIMO layers x CBW) and fronthaul BW is the fronthaul bandwidth for Option 7-2x (transport protocol overhead not included). Table *2.11* provides the total fronthaul bandwidth for Option

7.2x by adding the fronthaul bandwidths of the Sub-6 and millimetre wave frequency bands from Table *2.9* and Table *2.10*.

Table 2.9: Fronthaul bandwidth for Option 7-2x in Sub-6 [37].

	Sub-6				
	Number of sectors	**Total CBW (MHz)**	**MIMO layers**	**ABW (MHz)**	**Fronthaul BW (Gbit/s)**
Small	1				0
Medium	3	100	4	1200	23
Large	3	100	16	4800	90

Table 2.10: Fronthaul bandwidth for Option 7-2x in millimetre wave [37].

	Millimetre wave				
	Number of sectors	**Total CBW (MHz)**	**MIMO layers**	**ABW (MHz)**	**Fronthaul BW (Gbit/s)**
Small	1	400	2	800	15
Medium	3	400	4	4800	87
Large	3	800	4	9600	175

Table 2.11: Total fronthaul bandwidth for Option 7-2x in Sub-6 and millimetre wave [37].

	Sub-6 Fronthaul BW (Gbit/s)	**Millimetre wave Fronthaul BW (Gbit/s)**	**Total Fronthaul BW (Gbit/s)**
Small	0	15	15
Medium	23	87	110
Large	90	175	265

The formula used for the fronthaul BW calculation in Table *2.9* and Table *2.10* is given in equation 2.7 [37]:

$$Fronthaul\ BW = 2 \times 10^{-9}(1+c)\frac{v_{layer} N_{PRB}\left(12N_{mantissa}+N_{exponent}\right)}{T_s^{\mu}} (\text{Gbit/s}) \qquad (2.7)$$

where:

v_{layer} is the maximum number of supported layers;

N_{PRB} is the maximum resource block (RB) allocation for a given channel bandwidth and numerology μ;

$N_{mantissa}$ is the number of mantissa bits. $N_{mantissa}$ = 9 used in Table *2.9* and Table *2.10*;

$N_{exponent}$ is the number of the exponent bits. $N_{exponent}$ = 4 used in Table *2.9* and Table *2.10*;

T_s^{μ} is the average OFDM symbol duration in a subframe for a given numerology, i.e., $T_s^{\mu} = \frac{10^{-3}}{14 \text{ x } 2^{\mu}}$ (second);

c is the overhead from the control-plane. For downlink, *c* ≈ 10% and for uplink *c* ≈ 0, as the control-plane is primarily the downlink traffic in O-RAN specification [87]. The overhead may also take different value depending on vender specific implementation.

The fronthaul bandwidth calculation in equation 2.7 does not include the protocol encapsulation overhead (e.g., eCPRI header, Ethernet header, and IP header).

In this thesis, the focus is on the lower layer split which is used for the fronthaul transport network and supports the coordinated multipoint scheme, a key technology in 5G RAN. This split includes options 6 and 7-2x, which have stringent latency requirements and their bitrates dynamically change in proportion to the user plane traffic. In a disaggregated RAN, the split option 6 is ideal for centralised scheduling with a virtualised CU and small cell deployments, while option 7-2x are typically suited for ultra-reliable, low-latency communication use cases and virtualised DUs running on general purpose processing platforms located at near-edge data centres.

2.4 Conclusion

In this chapter, an overview of the mobile network architecture, mobile technology generations and new RAN architectures were discussed. The latency, synchronisation and reliability requirements for the 5G fronthaul network were described. The 5G radio access network architecture and the radio stack were described while the fronthaul bandwidth for the different functional splits was calculated.

The evolution of the transport network architecture from a centralised architecture to a disaggregated and virtualised architecture has led to meeting different bandwidth and latency requirements for the fronthaul, midhaul and backhaul transport segments. Various transport options have been suggested in the literature to meet these requirements including passive optical networks. Passive optical networks, which are widely deployed in broadband access networks for residential and business use, have become a cost-effective option to meet the transport requirements for current and future communication networks. Passive optical networks will be discussed in detail in Chapter 3.

In 3G and 4G mobile systems, the CPRI interface is used in the fronthaul and provides a constant bitrate transmission regardless of the amount of traffic being transmitted. With the RAN functional splits in 5G mobile systems and using various fronthaul interfaces the required bandwidth in the fronthaul varies and is in proportion to the mobile traffic. The focus of this thesis is the lower layer split options 6 or 7-2x, which is

used by the 5G fronthaul network and is chosen based on the specific use case and deployment type.

3. Passive Optical Networks

3.1 Introduction

This chapter provides an overview of passive optical networks – their architecture and standards. The different types of dynamic bandwidth allocation (DBA) algorithms used are discussed with a focus on ITU-T PON DBAs followed by the various mechanisms by which DBA algorithms work.

3.1.1 Architecture

A passive optical network (PON) is a point-to-multipoint optical access network in which the transmission path between the central office (CO) and the customer premises is passive, i.e., it does not contain active electronic devices and does not require any electrical power supply. In contrast, an active optical network uses active electronic devices, such as an Ethernet switch or router in the transmission path, which require electrical power. In a PON, power is only required at the CO or at the customer premises. PONs are mainly used for very short distances in the 'last mile' (the last segment of the network seen by end users) or 'first mile' (the first segment of the network seen by end users), connecting the service providers' central office (CO) or local exchange to individual subscribers or end users. PONs are widely used to provide high capacity broadband access to residential and business users in fibre-to-the-x (FTTx) systems, where 'x' can be 'node,' 'curb,' 'building,' 'home'.

The major FTTx systems are described below and illustrated in Figure *3.1* [90]:

- FTTN (Fibre to the node, -neighborhood): In FTTN, optical fibre is terminated in a street cabinet that is close to the end user's premises. FTTN relies on copper for the final connections.
- FTTC (Fibre to the curb, -cabinet): FTTC is similar to FTTN in that the optical fibre is terminated in a street cabinet but with a shorter distance to the end user's premises.
- FTTB (Fibre to the building, -basement): In FTTB, the optical fibre terminates in the end user's premises, such as an apartment building or multi-dwelling unit (MDU). The final connection to the individual home is provided by wireline or wireless technologies.
- FTTH (Fibre to the home): In FTTH, the optical fibre is terminated at the boundary of an individual's home, usually a box on the wall.

In Figure *3.1*, the following technologies are shown: asymmetrical digital subscriber line (ADSL), very-high-bit-rate digital subscriber line (VDSL), hybrid fibre-coaxial (HFC) and cable TV (CATV) with various network devices such the digital subscriber line access multiplexer (DSLAM), and modulator-demodulator (modem). The configuration in FTTx systems can be point to point (PtP) or point to multipoint (PtMP) like a PON.

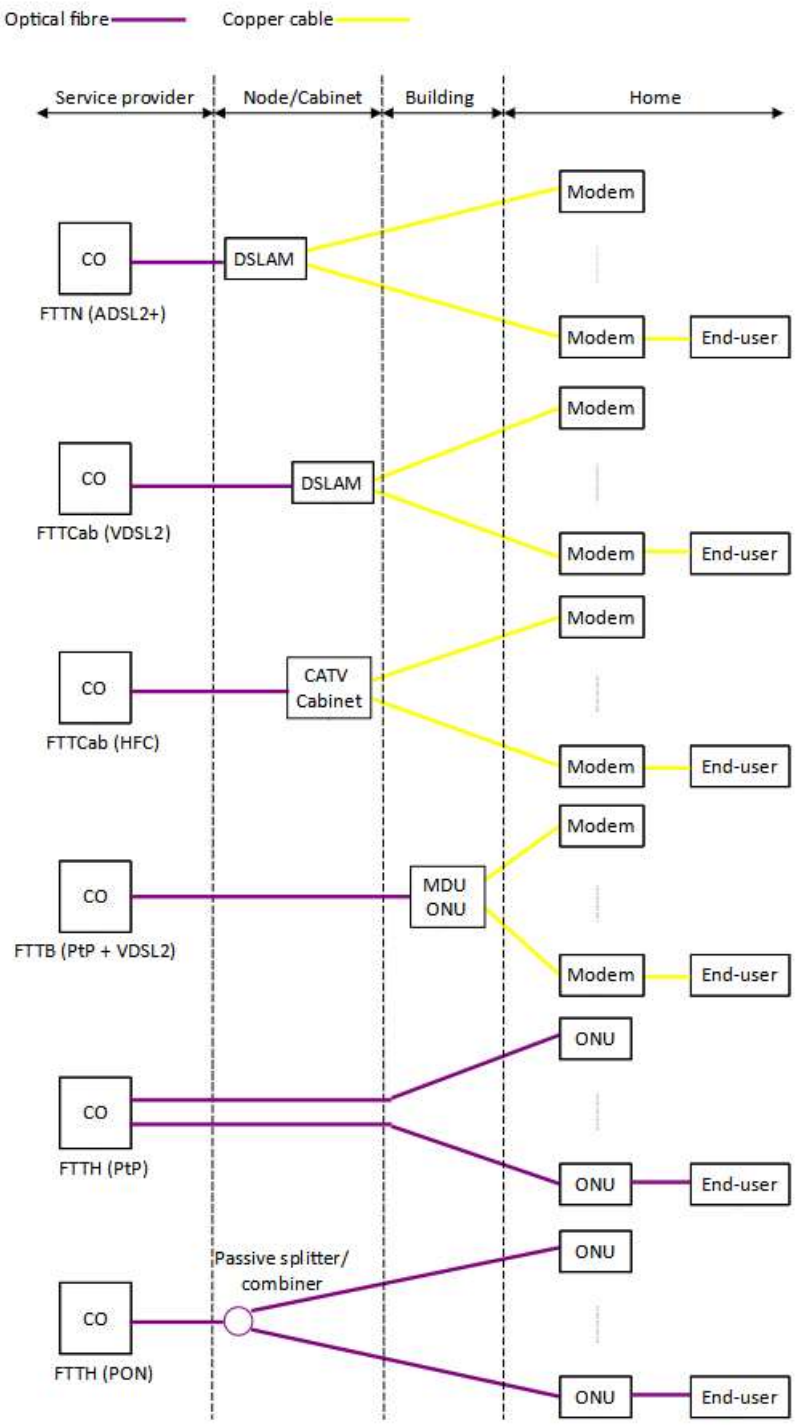


Figure 3.1: FTTx network architectures [90].

There are three main network elements in a PON – an optical line terminal (OLT), an optical network unit (ONU) and an optical distribution network (ODN), as shown in

Figure *3.2*. The OLT is located at the service providers' central office (CO) or local exchange. It provides the interface between the ODN and the backbone network. The ONU is located near subscribers' or end users' premises. The ODN is the interconnection between the OLT and the ONU and it consists of optical fibres passing through a passive optical splitter/combiner, splices, connectors and filters. The feeder fibre from the OLT is directly interconnected to distribution (or drop) fibres going to the ONUs.

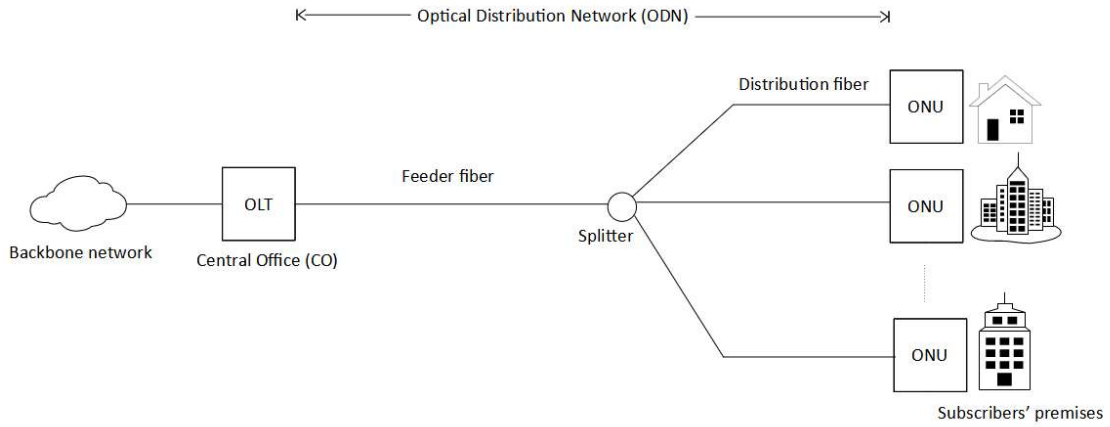


Figure 3.2: PON architecture.

The OLT broadcasts data to all ONUs in the downstream direction (from OLT to ONU). In the upstream direction (from ONU to OLT), each ONU is able to send data to the OLT but not directly with one another. In summary, the downstream direction of a PON can be said to have a point-to-multipoint architecture while in the upstream direction, a multipoint-to-point architecture. The optical splitter/combiner is used to divide (split) and distribute the downstream optical signal into multiple, equal but low power signals to the ONUs and combine (couple) the upstream optical signals from the ONUs. The optical splitter in the ODN has a split ratio of 1: N, where N can typically range from 4 to 256. The distance between the OLT and each ONU ranges from 10 to 20 km.

3.1.2 Multiple Access and Multiplexing Methods

There are various methods of handling traffic in the upstream and downstream directions in a PON. Multiplexing allows the combination of multiple data streams, each dedicated to one or several terminals, over the transmission medium in the downstream direction. The multiple access method allows several terminals to share the same transmission medium in the upstream direction [91].

- **Time Division Multiple Access (TDMA)**: This method allows the OLT to provide different time slots to each of the connected ONUs. Each ONU has a dedicated bandwidth to transmit their data for the duration of its allocated time slot. To avoid any collision of the data in the upstream, each ONU needs to negotiate with the OLT when it is permitted to send its data. This negotiation process is accomplished by using bandwidth allocation algorithms. To realise a TDMA system, a burst-mode optical receiver and transmitter are required at the OLT, as well as a passive optical power splitter at the optical distribution node (ODN) or remote node (RN).
- **Wavelength Division Multiple Access (WDMA)**: In this method, each ONU is assigned a pair of dedicated wavelengths. This means that each ONU can send data to the OLT at any time. To implement WDMA, a WDM multiplexer is used at the RN and a WDM demultiplexer is located at the CO to separate the multiple wavelength signals at the OLT.
- **Code Division Multiple Access (CDMA)**: CDMA uses the CDM multiplexing method to allow multiple signals from multiple users to share the same communication channel. This is achieved by assigning each channel a unique

and orthogonal code to separate it from one another. Each user can transmit at any time regardless of when the others are transmitting.

- **Orthogonal Frequency Division Multiple Access (OFDMA)**: In OFDMA, the OFDM modulation scheme is used, whereby a high-bandwidth signal is divided into a number of orthogonal channels (i.e., subcarriers). A typical OFDM-PON architecture is discussed in section 3.2.3.

3.2 PON Architectures

The three major PON architectures, differentiated by their multiplexing scheme, are time division multiplexing (TDM) PON, wavelength division multiplexing (WDM) PON and orthogonal frequency division multiplexing (OFDM) PON. There is a fourth type of PON technology, time and wavelength division multiplexed (TWDM) PON, which is a combination of the TDM and WDM schemes. The focus of this thesis is on TDM PONs.

3.2.1 Time Division Multiplexed PON (TDM-PON)

TDM-PON is the most widely used PON architecture today. It was invented in the late 1980s at the British Telecom Research Laboratories (BRTL), now BT Labs, to provide last mile broadband access [92][93][94][95]. In a TDM-PON, downstream traffic is broadcast by the OLT to all ONUs that share the same optical fibre. Since ONUs normally have different distances to the OLT, the data bursts from these ONUs must be scheduled carefully for a collision-free and efficient upstream data communication. In the upstream, each ONU is granted time slots in which to transmit their data to the OLT, as shown in Figure *3.3*. An arbitration mechanism is required so that only a single ONU is

allowed to transmit data at a given point in time because of the shared upstream channel. The start time and length of a transmission timeslot for each ONU are scheduled using a bandwidth allocation scheme.

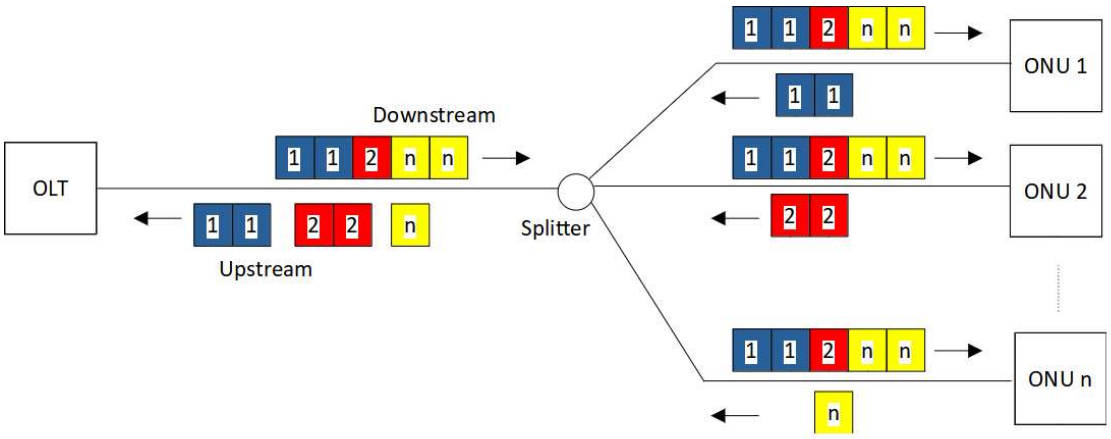


Figure 3.3: TDM-PON architecture.

3.2.2 Wavelength Division Multiplexed PON (WDM-PON)

Wavelength division multiplexed (WDM) PONs were proposed in the late 1980s by a group at Bell Communication Research (Bellcore), USA [96][97][98]. In a WDM-PON, one optical fibre is used to carry different wavelengths to each end user. Each ONU has its own upstream and downstream wavelength for communication with the OLT e.g., the separate wavelengths allow each ONU to have a point-to-point connection to the OLT. In a WDM-PON, a passive wavelength splitter such as a thermal arrayed waveguide grating (AWG) is used at the remote node (RN). The function of the AWG is to (de-) multiplex optical wavelengths. The point-to-point connection enables each wavelength channel (λ) to run at a different data rate (e.g., 1.25 Gbit/s, 2.5 Gbit/s, 10 Gbit/s) and with a different protocol, as shown in Figure *3.4*.

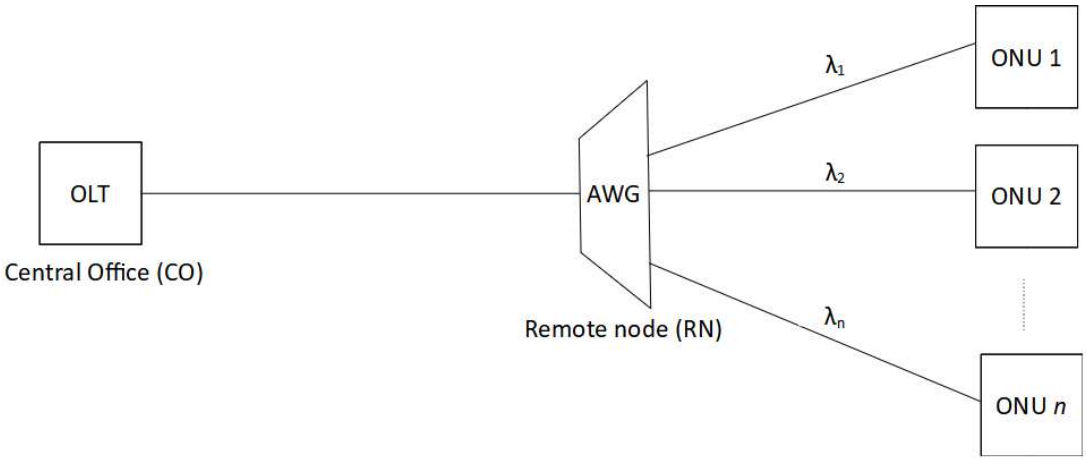


Figure 3.4: WDM-PON architecture.

3.2.3 Orthogonal Frequency Division Multiplexed PON (OFDM-PON)

In an OFDM-PON, the upstream/downstream traffic in the PON is transmitted over one wavelength channel, where each OFDM subcarrier is dynamically assigned to different ONUs in different time slots, as illustrated in Figure *3.5*. Due to their orthogonal nature, there will be no interference among the subcarriers from different ONUs.

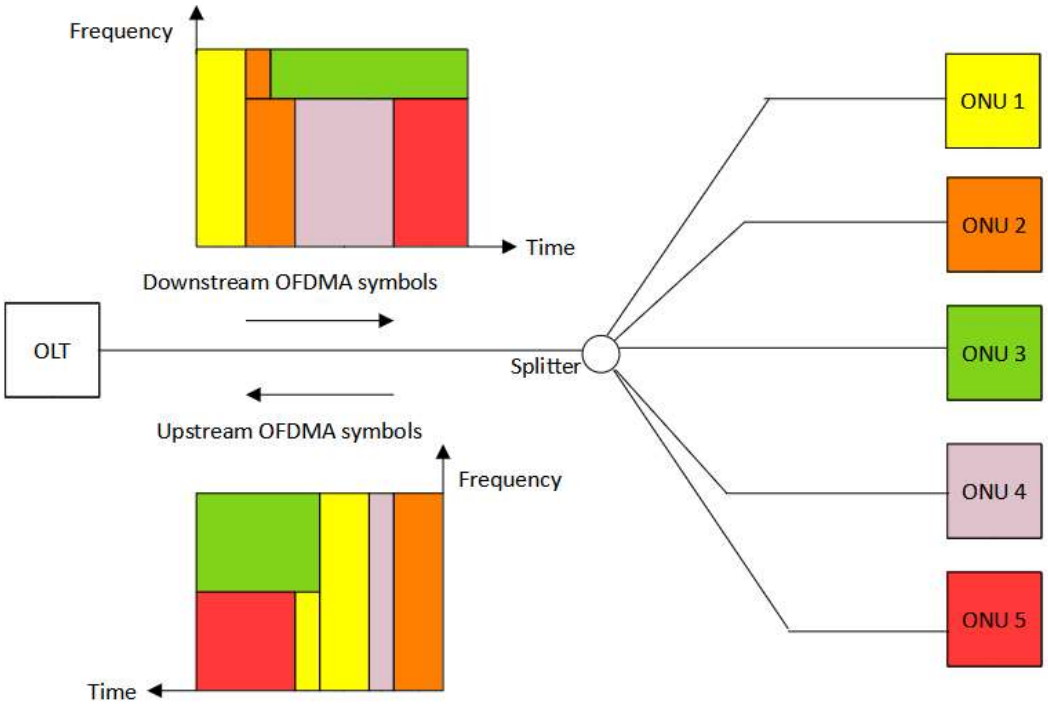


Figure 3.5: OFDM-PON architecture [99].

3.2.4 Time and Wavelength Division Multiplexed PON (TWDM-PON)

The time- and wavelength-division multiplexed PON (TWDM-PON) is based on a combination of the TDM and WDM schemes. In this case, the optical spectrum is divided into frequency channels, and each channel is divided into time slots.

3.3 PON Standards

The main organisations involved in PON standardisation are the International Telecommunication Union (ITU) through its Telecommunication Standardization Sector (ITU-T) and the Institute of Electrical and Electronics Engineers (IEEE). The Full Service Access Network (FSAN) Group initiated the development of passive optical network

systems in 1995. The FSAN is not a standards development organisation but they offer contributions to the ITU-T SG15/Q2 group (the Question 2 subgroup under Study Group 15: Networks, Technologies and Infrastructures for Transport, Access and Home is responsible for writing recommendations for passive optical networks (PONs)) [100].

3.3.1 ITU-T PON Standards

The first TDM-PON standard (G.983.x series) was based on the asynchronous-transfer mode (ATM) communication protocol and became known as ATM PON (APON) [101]. It evolved into the broadband PON (B-PON) [101] and led to the development of the Gigabit-capable PON (GPON), based on the G.984 series of recommendations [102]. Other ITU-T PON standards include the G.987 series for 10-Gigabit-capable PON (XG-PON) [103]–[106], G.9807 series for 10-Gigabit-capable symmetric PON (XGS-PON) [107], G.989 series for next-generation PON stage 2 (NG-PON2) [108] and G.9804 series for Higher speed PON [109] [110] [111]. A summary of the major ITU-T PON standards is given in Table *3.1*.

Table 3.1: Summary of ITU-T PON standards.

	GPON	**XG-PON**	**XGS-PON**	**NG-PON2/ TWDM-PON**	**Higher Speed PON**
ITU-T Standard	G.984	G.987	G.9807	G.989	G.9804
Downstream Line Rate	2.5 Gbit/s	10 Gbit/s	10 Gbit/s	2.5, 10 Gbit/s (4 channels)	12.5, 25 Gbit/s
Upstream Line Rate	1.25 Gbit/s	2.5 Gbit/s	10 Gbit/s	2.5, 10 Gbit/s (4 channels)	50 Gbit/s
Transmission Wavelengths (Downstream)	1480–1500 nm	1575–1580 nm	1575–1580 nm	1596–1603 nm	O-band
Transmission Wavelengths (Upstream)	1260–1360 nm (regular) 1290–1330 nm (reduced) 1300–1320 nm (narrow)	1260–1280 nm	1260–1280 nm	1524–1544 nm (wide), 1528–1540 nm (reduced), 1532–1540 nm (narrow)	O-band

3.3.2 IEEE PON Standards

The PON standards specified by IEEE include IEEE 802.3ah (Ethernet Passive Optical Network (EPON)) [112], 802.3av (10 Gbit/s EPON) [113], 802.3ca (100G-EPON – 25, 50, and 100 Gbit/s over EPONs) [114] and 802.3cs (“Super-PON”) which defines an increased-reach, 10 Gbit/s optical access with at least 50 km reach and 1:64 split ratio per wavelength pair, 16 wavelength pairs and is scheduled for completion in 2022 [115]. A summary of the IEEE PON standards is listed in Table *3.2*.

Table 3.2: Summary of IEEE PON standards.

	EPON	**10G-EPON**	**NG-EPON**
IEEE Standard	802.3ah	802.3av	802.3ca
Downstream Line Rate	1 Gbit/s	10 Gbit/s	25, 50 Gbit/s
Upstream Line Rate	1 Gbit/s	1, 10 Gbit/s	10, 25, 50 Gbit/s
Transmission Wavelengths (Downstream)	1480–1500 nm	1575–1580 nm	1590–1610 nm, 1524–1544 nm
Transmission Wavelengths (Upstream)	1260–1360 nm	1260–1280 nm (for 10 Gbit/s upstream)/ 1260–1360 nm (for 1 Gbit/s upstream)	1280–1360 nm

As shown in Table *3.1* and Table *3.2*, the upstream and downstream wavelength bands are essentially the same for the ITU-T and IEEE PON standards. This has led to the use of common optical components by both standard organisations [116].

3.4 Dynamic Bandwidth Allocation (DBA)

The dynamic bandwidth allocation (DBA) represents a method to allow the OLT to reallocate bandwidth quickly across the entire PON according to current traffic conditions. The DBA would allocate upstream time slots to different ONUs such that no collision is observed at the OLT. This requires accurate ranging and timing to measure the round-trip time (RTT) between the OLT and ONU such that the OLT can appropriately offset the upstream time slots granted to ONUs to avoid upstream packet collision.

3.4.1 Bandwidth Management in PON

To effectively utilise the bandwidth in the upstream direction in a TDM-PON, three processes are usually used: bandwidth negotiation, scheduling and bandwidth allocation [117]. These processes take place at the data link layer of the TDM-PON.

Bandwidth Negotiation: Bandwidth negotiation is the process by which information is exchanged between the OLT and ONUs. This is required to enable each ONU to request a time slot to transmit its data and for the OLT to send its bandwidth arbitration decision to each ONU. In TDM-PONs, the MAC layer defines a protocol that supports bandwidth negotiation between the OLT and ONUs.

Scheduling: The scheduling process is used to decide the transmission order of data from different ONUs to the OLT. This is usually determined by a scheduling algorithm. The most widely-used scheduling algorithm is round-robin (RR). RR is simple to implement but is not adaptive to instantaneous traffic conditions at each ONU. This issue was addressed by [118] who proposed two adaptive scheduling algorithm: the longest-queue-first (LQF) algorithm and the earliest-packet-first (EPF) algorithm.

Bandwidth Allocation: In order to allocate bandwidth to each ONU, the OLT uses a bandwidth allocation protocol. There are two broad categories: static bandwidth allocation (SBA) and dynamic bandwidth allocation (DBA) [117], [119]. In an SBA, each ONU is granted a predefined bandwidth allocation. As shown in Figure *3.6*, the traffic from the users arrives at the ONU at a fixed rate. Once an ONU is idle, its allocated bandwidth is unavailable to other ONUs in the network. For example, in Figure *3.6*, ONU

2 cannot transmit data using the idle time slots made available by ONU 1 and ONU 3. This leads to a higher latency for ONU 2 and a reduced performance in overall upstream utilisation [120]. In a DBA, the OLT adjusts the bandwidth allocation to the varying (bursty) amount of traffic in the network. As shown in Figure *3.7*, the bandwidth not used by ONU 1 and ONU 3 is allocated to ONU 2. Also, ONU 1 uses the timeslot not used by ONU 3. This leads to a more efficient bandwidth utilisation in the network. Using a DBA also offers a statistical multiplexing gain and a better quality of service (QoS) to end users [120].

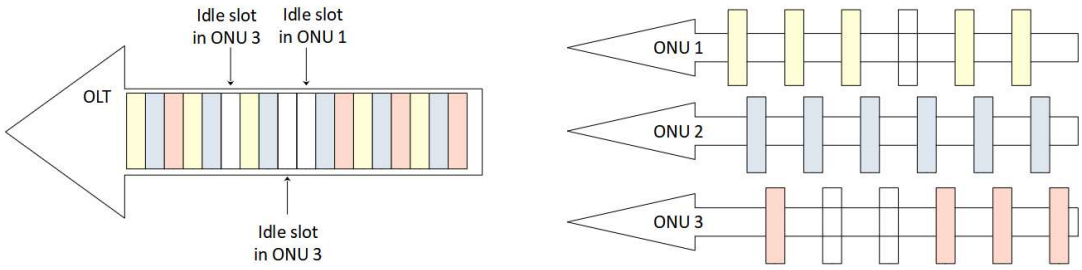


Figure 3.6: Static Bandwidth Allocation Example [120].

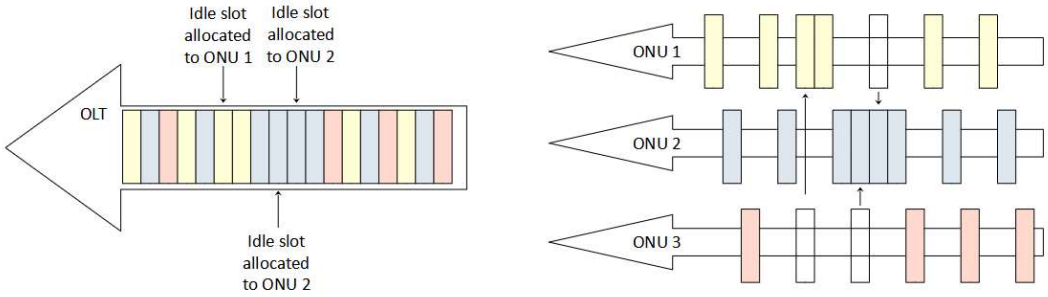


Figure 3.7: Dynamic Bandwidth Allocation Example [120].

3.4.2 Ranging and Discovery

In order to prevent collisions as each ONU transmits its data to the OLT, a waiting period is adhered to before transmission. However, the various ONUs have varying distances from the OLT and this affects the time at which each received upstream data reaches the OLT. This variation in propagation time may eventually result in collision. Ranging is a method used to determine the round-trip time (RTT) from the OLT to the ONU. The RTT is calculated as the difference between the OLT local time and the incoming timestamp [121]. The ranging process is performed by the OLT during ONU initialisation before the ONU is permitted to transmit any upstream data.

3.4.3 ITU-T DBA Methods

The ITU-T PON standards supports three DBA methods to infer the buffer occupancy status of each ONU [122] – status reporting DBA (SR-DBA), traffic monitoring DBA (TM-DBA) and cooperative DBA (CO-DBA). An OLT can implement one or all three of these methods.

Status Reporting DBA (SR-DBA): In a SR-DBA, all ONUs report their upstream data queue occupancy to the OLT periodically using the dynamic bandwidth report upstream (DBRu) field of the upstream frames. Each ONU may have several transmission containers (T-CONTs), each with its own traffic class. The OLT computes the upstream bandwidth allocations for each T-CONT in an ONU based on the received reports using a DBA algorithm. It then sends the bandwidth allocation to the ONUs as a grant via the bandwidth map (BWmap) field of the downstream frames, as shown in Figure *3.8*. When

an ONU has no information to send, it sends an idle frame (a GEM frame with zero payload length, consisting of an all-zero GEM header) upstream to indicate that its buffer is empty. This informs the OLT that the grants for that T-CONT can be assigned to other T-CONTs.

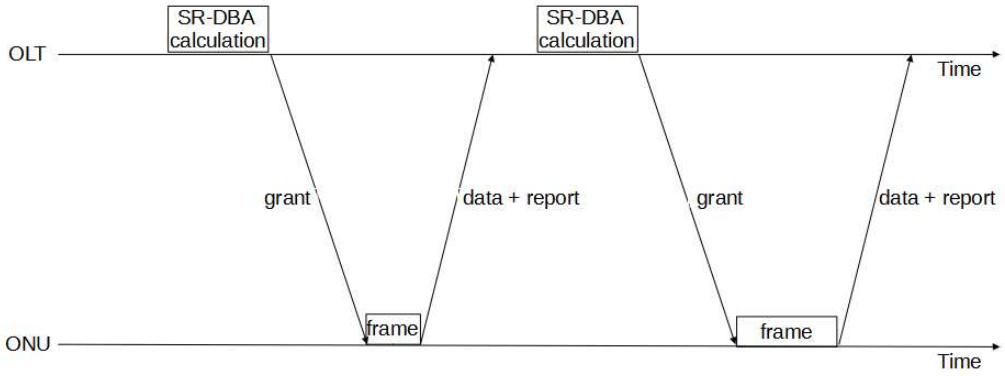


Figure 3.8: Timing diagram of SR-DBA process.

Traffic Monitoring DBA (TM-DBA): In a TM-DBA, ONUs do not provide explicit queue occupancy information. Instead, the OLT estimates the ONU queue status, typically based on the actual transmission in the previous cycle. For example, if an ONU has no traffic to send, it transmits idle frames during its allocation period. The OLT would then observe the idle frames and decrease the bandwidth allocation to that ONU in the following cycle. In the opposite case, the OLT constantly increases the allocation size until idles are detected, slowly adjusting to growing traffic.

Cooperative DBA (CO-DBA): In a CO-DBA, which is mainly used for use cases involving PON-based fronthaul, the OLT receives mobile scheduling information from the CU/DU

and allocates bandwidth to the ONUs in advance of the request from the RU, as shown in Figure *3.9*. Chapter 5 describes in detail the CO-DBA method.

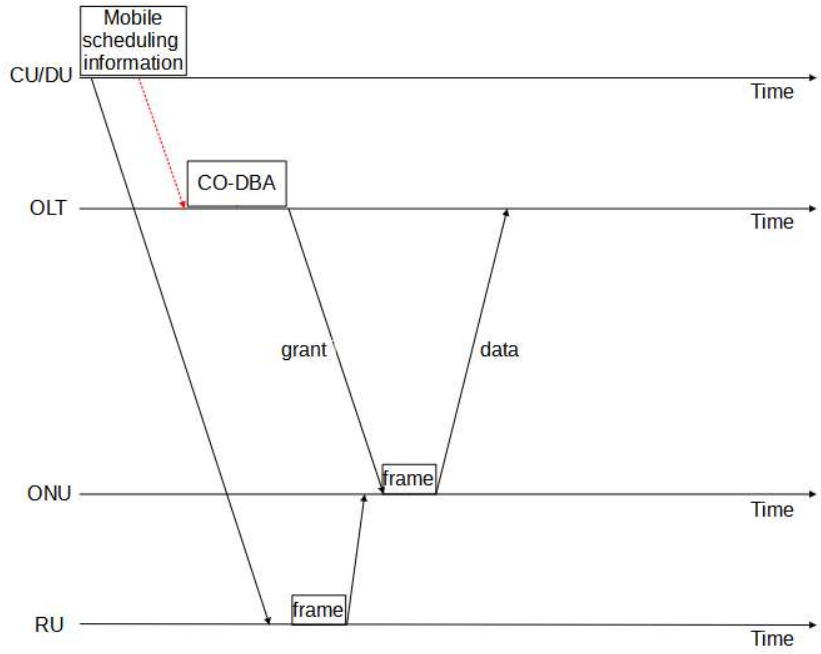


Figure 3.9: Timing diagram of CO-DBA process.

3.4.4 GPON Encapsulation Method (GEM)

For transport over the PON, GPON must encapsulate service data units (SDUs). SDUs include user data (Ethernet) frames and OMCI messages. The result of the encapsulation is called a **GEM frame**: GPON encapsulation method, as shown in Figure *3.10*. This occurs at the GTC layer. Each GEM frame is tagged with a GEM port ID, which uniquely identifies a particular flow on the PON. GEM frame encapsulation serves two functions: (1) multiplexing of GEM ports and (2) payload data fragmentation. GEM frames allow the GPON protocol to support variable-size frames such as Ethernet. The GEM frame consists of a header and a payload (which could be a full frame or a frame fragment). In

IEEE PONs, the user data is transmitted as native (normal) Ethernet frames in the PON [123].

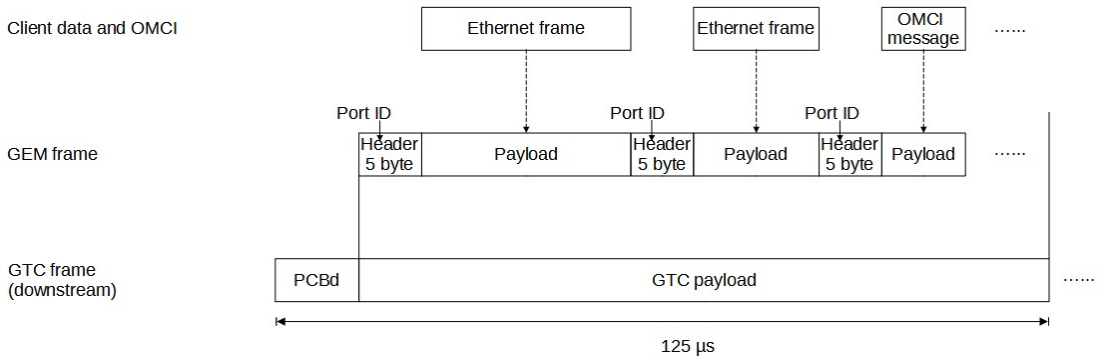


Figure 3.10: GPON Encapsulation Method (GEM).

The downstream and upstream GTC frames have a fixed duration of 125 µs. In the downstream direction, GEM frames are transmitted from the OLT to the ONUs using the GTC frame payload section, as shown in Figure *3.11*. The payload section carries variable-size GEM frames. The header section is known as the downstream physical control block (PCBd). The PCBd section consists of a fixed part and a variable part. The fixed part contains the following fields:

- A 4-byte physical synchronisation (PSync) field,
- A 4-byte Identification (Ident) field and
- A 13-byte downstream physical layer operations administration and maintenance (PLOAMd) field.

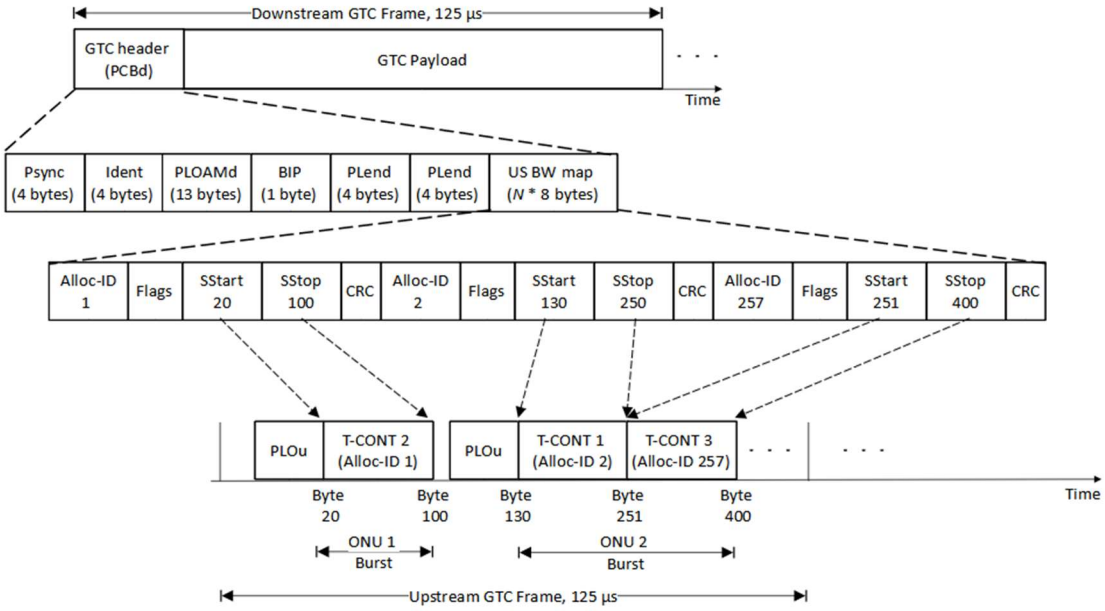


Figure 3.11: Frame structure in GPON [11].

The fields in the fixed part of the PCBd are protected by a 1-byte bit-interleaved parity check (BIP). The variable part of the PCBd contains a duplicated 4-byte downstream payload length indicator (PLend), which gives the length of the upstream bandwidth allocation map (US BW map) in the T-CONT. The US BW map is *N* x 8 bytes long and contains *N* entries associated with *N* time-slot allocation identifications for the ONUs. Each entry in the US BW map consists of:

- A 12-bit Alloc-ID of a T-CONT
- A 2-byte (or 16-bit) start pointer field (SStart) that indicates when the upstream transmission window starts
- A 2-byte (or 16-bit) stop pointer field (SStop) that indicates when the upstream transmission window stops

- A 1-byte (or 8-bit) cyclic redundancy check (CRC). The access structure is protected by a CRC-8 that provides 2-bit error detection and 1-bit error correction.
- A 12-bit flags field that indicates how the allocation should be used.

In the upstream direction, frames are transmitted from ONU to OLT using the configured GEM allocation time. The ONU buffers GEM frames as they arrive, and then sends them in bursts when allocated time to do so by the OLT. The OLT receives the frames and multiplexes them with the frames from the other ONUs. The upstream GTC frame consists of four types of overhead fields and a variable-length user data payload that contains a burst of transmission. The four overhead fields are:

- The upstream physical layer overhead (PLOu)
- The upstream physical layer OAM (PLOAMu)
- The upstream power levelling sequence (PLSu)
- The upstream dynamic bandwidth report (DBRu)

Transmission of the PLOAMu, PLSu and DBRu fields are optional depending on the downstream flags in the US BW map. A DBRu field is tied to each T-CONT for reporting the upstream traffic status of the associated T-CONT and is protected against bit errors by a CRC. The DBRu field contains the dynamic bandwidth allocation (DBA) report which is used to indicate the upstream queue length for dynamic bandwidth allocation. The DBA report contains the number of GEM blocks (which are 48-byte in length) waiting in the upstream traffic.

Each upstream transmission burst starts with a PLOu and one or more bandwidth allocation intervals associated with individual Alloc-IDs. The US BW map dictates the arrangement of the bursts within the frame and the allocation intervals within each burst. Each allocation interval is controlled by a specific allocation structure of the US BW map. In order to accommodate traffic from multiple ONUs, it is often desirable to fragment an Ethernet frame. GEM allows for the fragmentation of Ethernet frames in order to increase bandwidth utilisation. A large Ethernet frame may be split into fragments spanning multiple GEM payload fields.

The GEM process in GPON is similar to those used in XG-PON [105] and XGS-PON [107] standards, where it is called the XG-PON and XGS-PON encapsulation method (XGEM), respectively.

3.5 PON DBA Algorithms

In the various PON standards (IEEE and ITU-T), the DBA algorithm itself is not standardised. This has allowed for a lot of research in finding efficient DBA algorithms for PON.

3.5.1 ITU-T PON DBA Algorithms

There are several ITU-T TDM-PON (GPON, XG-PON) DBAs such as the GigaPON Access NeTwork (GIANT) [124], [125], [126], Immediate Allocation with Colorless Grant (IACG) [127], Efficient Bandwidth Utilization (EBU) [128], GPON Redundancy Eraser Algorithm for Long-Reach (GREAL) [129], [130], Balance Transfer (BLT) [131] [132] and the Offset-

based Scheduling with Flexible Intervals (OFSI)/OSFI using Predictive Colorless Grants (PCG-OSFI) [133]. A detailed survey of ITU-T PON DBAs can be found in [134]. In this section, the focus will be on the GIANT, IACG and EBU DBAs, which are the only algorithms to have been physically implemented [135].

GigaPON Access NeTwork (GIANT) DBA: The GIANT DBA was the first DBA algorithm compliant with the ITU-T GPON standard. It was proposed and implemented [136] to counter the bursty nature of the upstream traffic in a network, i.e., there are periods when the traffic source generates packets at a very high rate compared to the average rate. The DBA supports four traffic service classes – transmission containers (T-CONT) types 1, 2, 3 and 4 (T1, T2, T3 and T4) as defined in [137]. Each T-CONT type has two service parameters, a service interval (SI) and an allocation byte (AB). The SI determines how often the T-CONT gets served, while the AB determines how many bytes on the upstream frame can be assigned to the T-CONT. The SI is expressed in multiples of the frame duration (i.e., 125 μs) and the AB is expressed in bytes.

The algorithm allocates the available bandwidth to the T-CONTs based on two service parameters (SI_{max} or SI_{min}, and AB_{min} or AB_{sur}) in order to estimate the respective share of each T-CONT demand. SI_{max} and SI_{min} denote the maximum and minimum service interval, respectively while AB_{min} and AB_{sur} denote the minimum and surplus allocation bytes, respectively.

Each T-CONT has a down counter (SI_{max_timer} or SI_{min_timer}) for bandwidth allocation. The down counter represents the SI of the corresponding T-CONT and is decreased by one

every frame duration. The OLT can only allocate bandwidth to a T-CONT when the down counter of the queue reaches 1 (i.e., the timer has expired). The expiration of each down counter invokes the allocation of allocation bytes for the corresponding T-CONT and the down counter is reset to its service interval value. For bandwidth allocation to the T-CONT types, GIANT uses a fixed allocation for T1 and reservation-based allocation for T2 – T4. Table *3.3* shows the service parameters and counters of the GIANT DBA algorithm for each T-CONT.

Table 3.3: GIANT DBA service parameters and counters.

T-CONT type	**Bandwidth**	**Service Parameters**	**Down Counters**	**Applications**
T1	Fixed	SI_{max}, AB_{min}	SI_{max_timer}	Constant bit rate (CBR) traffic
T2	Assured	SI_{max}, AB_{min}	SI_{max_timer}	Variable bit rate (VBR) traffic
T3	Assured	SI_{max}, AB_{min}	SI_{max_timer}	Better than best effort
	Non-assured (Surplus)	SI_{min}, AB_{sur}	SI_{min_timer}	
T4	Best-effort	SI_{max}, AB_{min} (polling) SI_{min}, AB_{sur} (surplus)	SI_{max_timer} SI_{min_timer}	Best effort

The allocated bandwidth is calculated as the fraction of the allocation byte over the service interval (i.e., service rate, $R = AB/SI$). The assured service rate is expressed as: AB_{min}/SI_{max}, while the surplus (non-assured) service rate is expressed as: AB_{sur}/SI_{min}.

The GIANT bandwidth scheduling mechanism includes two phases – guaranteed phase allocation (GPA) and surplus phase allocation (SPA). It executes GPA and SPA recursively. This assures fairness of allocation by picking T-CONTs in a round-robin manner during each GPA and SPA. Both GPA and SPA are executed during a service interval. The

scheduling priority is in order of the assured bandwidth of T2, the assured bandwidth of T3, the polling bandwidth of T4, the surplus bandwidth of the T3, and the surplus bandwidth of T4.

The GIANT MAC algorithm has the drawback that each T-CONT can obtain the transmission time slot only when its down counter reaches 1. This drawback results in performance degradation. It also does not consider the unallocated remainder of the transmission time slot. If the total queue length at the OLT is less than the upstream frame size, then the upstream transmission time slot is not fully allocated.

Immediate Allocation with Colorless Grant (IACG) DBA: The IACG DBA is based on the GIANT DBA and was proposed [127] to overcome the drawbacks in GIANT. This was achieved in IACG with a colorless grant phase that utilises the unallocated remainder of the upstream frame and assigns this unallocated remainder to each ONU with equal share. Also, IACG makes use of an additional parameter (available byte counter) to increase the grant allocation frequency, which purely relied on the service timer expiry of each T-CONT. The IACG DBA bandwidth scheduling mechanism is discussed in section 3.6.

Efficient Bandwidth Utilization (EBU) DBA: The EBU DBA was proposed [128] for XG-PON as an improved version of IACG. EBU seeks to solve the utilisation problem of unused bandwidth which was not addressed by the GIANT, IACG and PCG-OSFI algorithms. It achieves this by allowing any unused bandwidth to be utilised by other

ONUs. However, the implementation complexity of EBU is higher than that of IACG because an extra stage is required for the update of the available byte counters.

3.5.2 IEEE PON DBA Algorithms

A variety of DBA algorithms have been proposed for IEEE PONs and a detailed study of some of these algorithms can be found in [138], [117], [119]. The DBA algorithms for IEEE PONs are not compatible with ITU-T PONs and they are compared in Table *3.4* [139].

Table 3.4: Comparison of ITU-T and IEEE PON DBA algorithms [139].

ITU-T PON DBA	**IEEE PON DBA**
Designed for transfer of time-critical data	Designed for non-time-critical data transfers
Upstream frame is synchronised with downstream frame	Frames are not synchronised
Uses a single frame format for both upstream and downstream communications	Uses different frame formats for communication, like GATE and REPORT
Strict bandwidth requirements for a service class	No strict requirements
Different frame structure – uses a GEM (GPON Encapsulation method) frame	Frame structure has close resemblance to the Ethernet structure – easier to study and test
Allows fragmentation by using a segmentation and reassembly function	Does not allow fragmentation
Both upstream and downstream frames have a fixed duration of 125 µs	Frames do not have a fixed size
Uses line coding e.g., non-return-to-zero (NRZ)	Uses block coding e.g., 8B/10B

3.6 IACG DBA Bandwidth Scheduling

The IACG DBA uses the same two service parameters (SI_{max} or SI_{min}, and AB_{min} or AB_{sur}) and down counters (SI_{max_timer} or SI_{min_timer}) as the GIANT DBA and an additional parameter, known as the available byte counter (V_a or V_s). The available byte counter,

which represents the remaining available allocation bytes, is used to immediately allocate the transmission bandwidth for a T-CONT regardless of the down counter value. To limit the maximum allocation amount for a T-CONT, the algorithm cannot allocate more than the value of the available byte counter, and the available byte counter is decreased by the allocation amount.

The bandwidth scheduling mechanism of IACG consists of three phases – the Guaranteed Phase Allocation (GPA), Surplus Phase Allocation (SPA) and Colorless Grant (CG) phase. In the CG phase, the algorithm assigns the unused bandwidth equally among all the ONU T-CONTs at the end of a DBA cycle. During the bandwidth scheduling mechanism, the GPA is executed first, followed by the SPA, then the CG phase. In the GPA phase, bandwidth is allocated to the assured bandwidth of T-CONT 2 and the assured bandwidth of T-CONT 3 while in the SPA phase, bandwidth is allocated to the surplus bandwidth of T-CONT 3 and the best-effort bandwidth of T-CONT 4. After the SPA phase, the CG phase takes place whereby the remaining unassigned bandwidth is equally allocated to the ONUs using T-CONT 5. The ONUs uses the remaining unassigned bandwidth in a strict class priority order: the assured bandwidth of T-CONT 2, the assured bandwidth of T-CONT 3, the surplus bandwidth of T-CONT 3 and the best-effort bandwidth of T-CONT 4.

IACG uses the two service parameters (SI_{max} or SI_{min} and AB_{min} or AB_{sur}) for each T-CONT in an upstream transmission in order to allocate the available bandwidth. IACG uses the two counters to monitor:

- the available bytes that can be allocated to ONUs within the SI (available byte counter for the assured bandwidth, V_a or surplus bandwidth, V_s), and
- the remaining duration of the SI itself (down counter for assured bandwidth, SI_{max_timer} or surplus bandwidth, SI_{min_timer}).

Table *3.5* shows the service parameters and counters for each T-CONT type used in IACG.

Table 3.5: IACG DBA service parameters and counters.

T-CONT type	Bandwidth	Service Parameters	Down Counters	Available Byte Counters
T2	Assured	SI_{max}, AB_{min}	SI_{max_timer}	V_a
T3	Assured	SI_{max}, AB_{min}	SI_{max_timer}	V_a
	Non-assured (Surplus)	SI_{min}, AB_{sur}	SI_{min_timer}	V_s
T4	Best-effort	SI_{min}, AB_{sur}	SI_{min_timer}	V_s

Note that the maximum value of the available FB in every cycle (125 µs) is 38,880 bytes for a 2.48832 Gbit/s (approximately 2.5 Gbit/s) upstream line rate and 155,520 bytes for a 9.95328 Gbit/s (approximately 10 Gbit/s) upstream line rate. The DBA process is described as follows: if the available byte counter (V_a or V_s) and available frame bytes (FB) are greater than zero, then each ONU is granted a bandwidth allocation specified by the minimum of its request, the maximum allocation bytes of the T-CONT and the available FB ($grant = min(V_a, request, FB)$) in the following order: the assured bandwidth of T2 (shown in Figure *3.12*), assured bandwidth of T3 (shown in Figure *3.13*), surplus bandwidth of T3 (shown in Figure *3.13*) and best-effort bandwidth of T4 (shown in Figure *3.14*). The available byte counter is decreased by the grant amount and recharged to the allocation bytes ($V_a = AB_{min}$ or $Vs = AB_{sur}$) when its down counter

(SI_{max_timer} or SI_{min_timer}) reaches 1. At the end of the bandwidth allocation, any unallocated remainder of the available FB is distributed equally to all ONUs using T5 (shown in Figure *3.14*).

The execution of the three phases of the bandwidth assignment process and sending bandwidth grants in every downstream cycle aid to reduce the upstream delay by decreasing the grant time, i.e., the time after which the DBA algorithm is executed and allocations are sent to ONUs. This ultimately leads to the reduction of the OLT time, which depends on the grant time and DBA processing time.

Each ONU transmits a bandwidth report for each T-CONT, if requested by the OLT, in the dynamic bandwidth report upstream (DBRu) field of an upstream frame. The bandwidth report reflects the actual queue length of each T-CONT, which is used by the DBA algorithm in the OLT to create bandwidth grants which are sent in a downstream frame as the bandwidth map (BWmap) to each ONU every 125 µs. The BWmap defines the transmission timeslot allocated to each frame from a particular T-CONT in the upstream. Based on the bandwidth grants received, each ONU can begin to transmit their T-CONT frames in the upstream to the OLT.

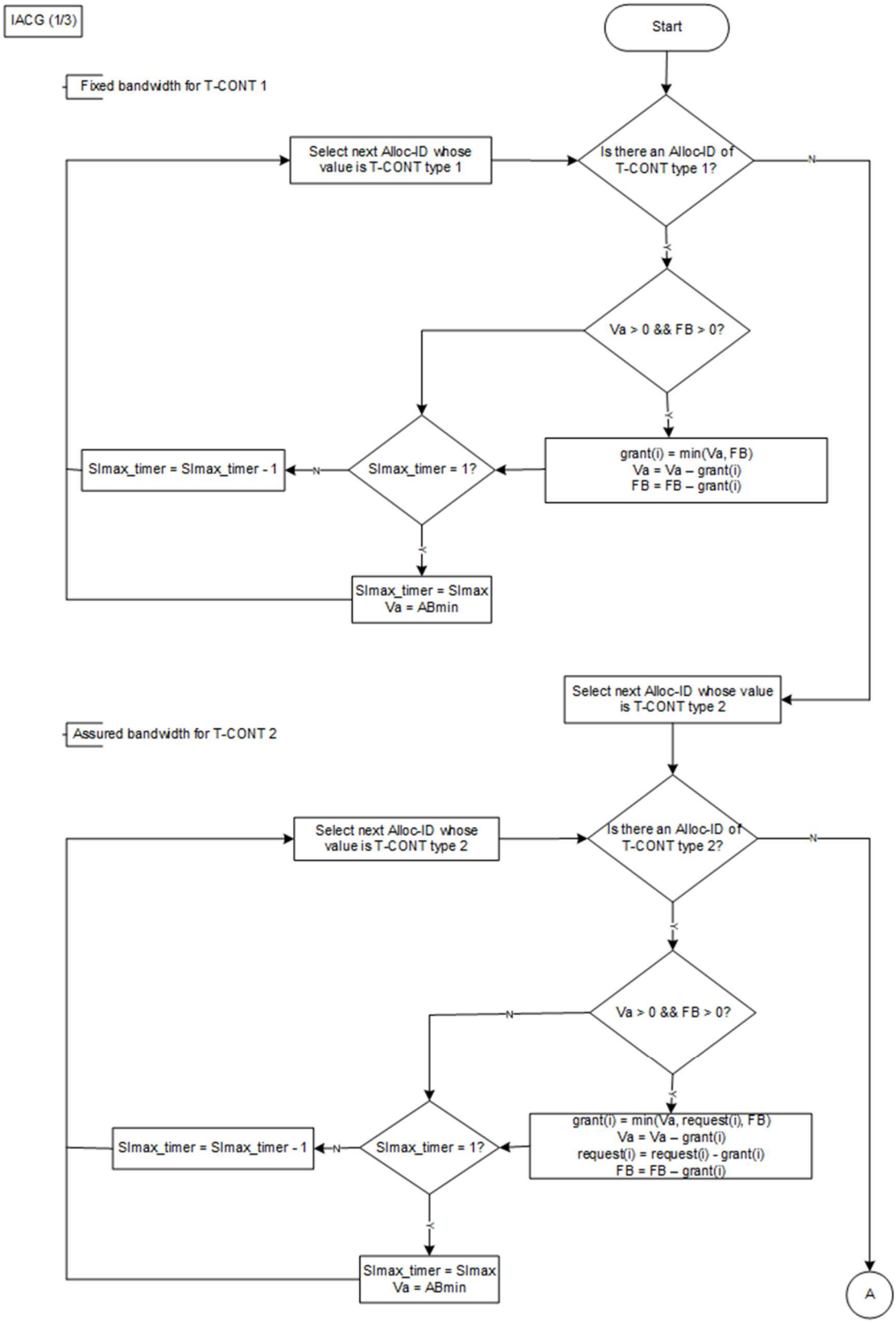


Figure 3.12: IACG DBA algorithm grant process flowchart for T1 (fixed bandwidth) and T2 (assured bandwidth).

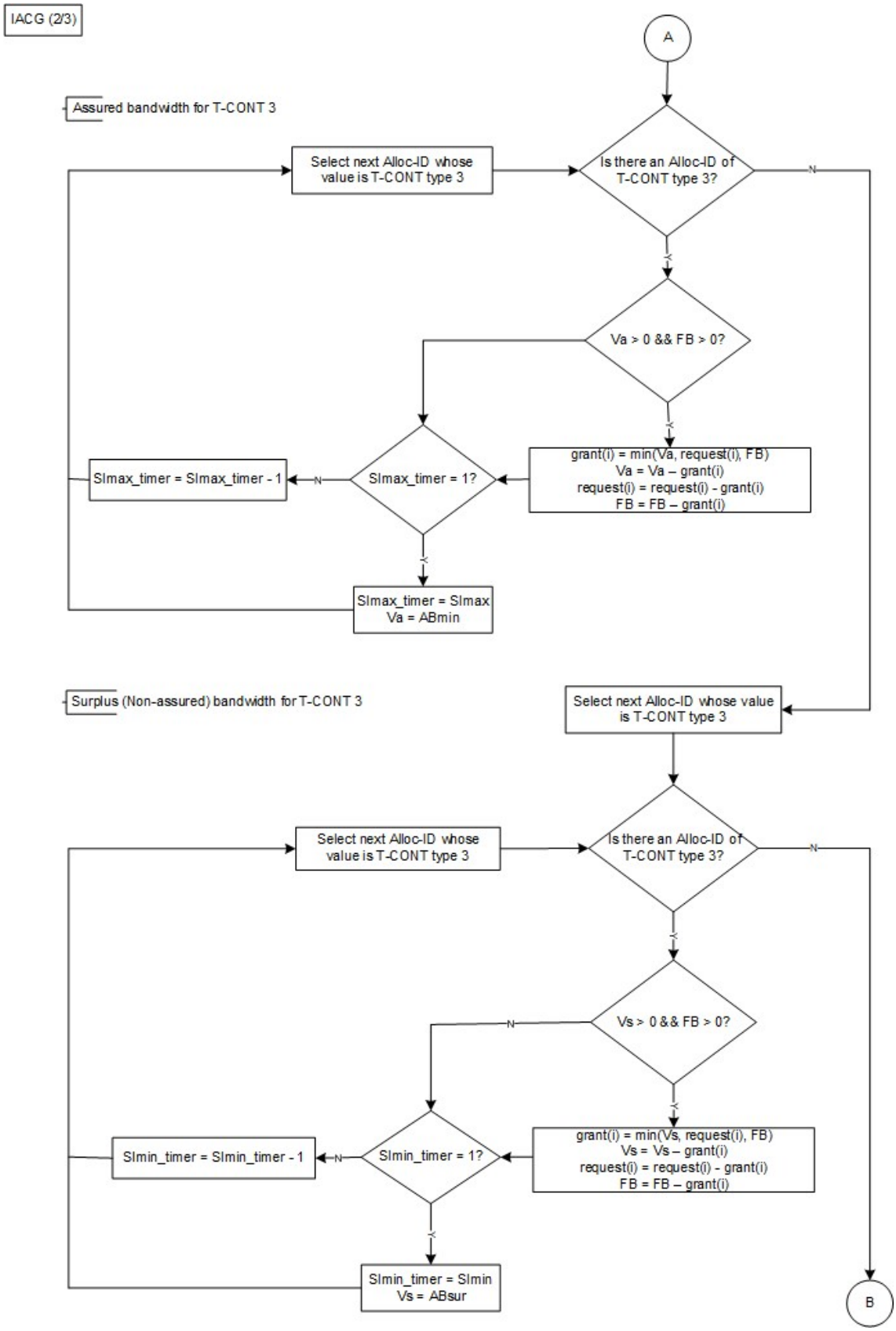


Figure 3.13: IACG DBA algorithm grant process flowchart for T3 (assured bandwidth and non-assured or surplus bandwidth).

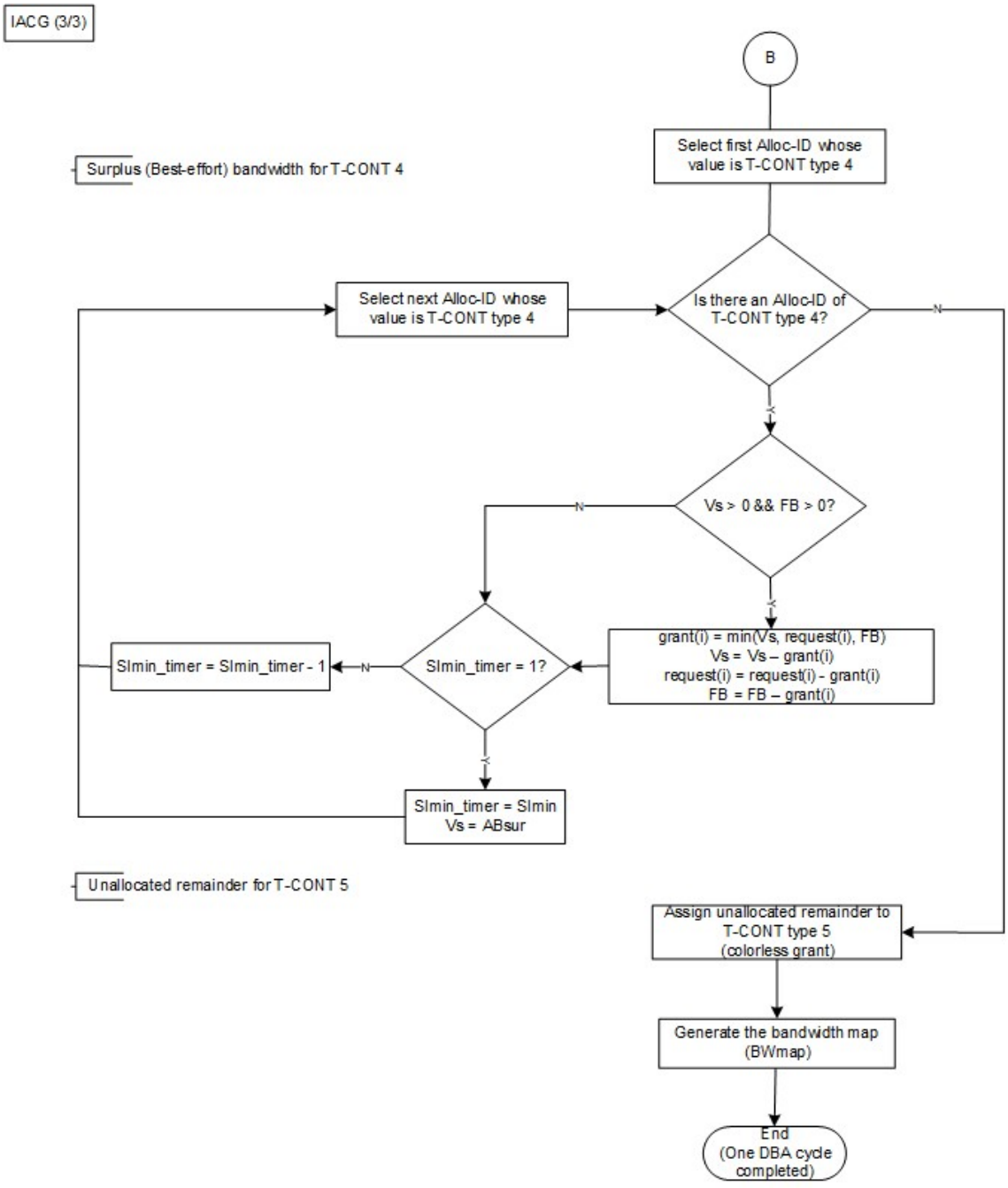


Figure 3.14: IACG DBA algorithm grant process flowchart for T4 (best-effort bandwidth) and T5 (colorless grant).

As discussed in section 3.5.1, the colorless grant phase in IACG allows for a more efficient allocation of bandwidth. This is because the algorithm sends bandwidth grants to the ONUs in every downstream frame during a SI and it assigns the unallocated remainder of the upstream frame equally to each ONU.

3.7 Conclusion

In this chapter, the architecture and standards of passive optical networks were reviewed. In addition, the various DBA algorithms based on the ITU-T and IEEE PON standards were discussed with a focus on ITU-T DBA methods and algorithms. In particular, the different DBA methods were explained, in addition to the bandwidth scheduling mechanism of two ITU-T standard compliant status reporting DBA algorithms – GIANT DBA and IACG DBA.

The DBA process is a key component in meeting the latency requirements and supporting different types of traffic in a TDM-PON for 5G and future networks. A combination of existing DBA methods can be used to handle both fixed and mobile traffic in a TDM-PON. Some advantages of the IACG DBA over the GIANT DBA include the ability to send bandwidth grants in every downstream frame and its use of a colorless grant phase which enables it to assign unallocated bandwidth to each ONU equally, leading to a higher bandwidth efficiency. The performance of the IACG DBA algorithm in meeting the latency requirements of a PON-based converged fronthaul will be evaluated in Chapter 4.

4. Dynamic Bandwidth Allocation Algorithm for Fronthaul

4.1 Introduction

This chapter presents the performance evaluation of a PON dynamic bandwidth allocation algorithm with regard to meeting the xHaul transport latency requirements. The focus is on the performance of the DBA in terms of the queueing delay, the average ONU upstream delay and the proportion of frames meeting the latency requirement. The chapter discusses the bandwidth grant cycle process and how it affects the delay between the ONU bandwidth requests and the arrival of the bandwidth grants from the OLT. Also, three different frame sizes are evaluated to determine their effect on the queueing delay.

4.2 Related Work

Passive optical networks (PONs) are fundamental to the implementation of 5G mobile networks and next-generation fixed networks [140], [141], which has been discussed in Chapters 2 and 3. A large proportion of the research on DBA algorithms for the fronthaul have focused on IEEE TDM-PON (EPON and 10G-EPON) [142], [143], [144], [145]. Several ITU TDM-PON (GPON, XG-PON, XGS-PON) DBA algorithms have been studied for use in the fronthaul. These include the Round Robin DBA (RR-DBA) [146], Group Assured GIANT (gGIANT) DBA [147] – both based on the well-known GigaPON Access NeTwork (GIANT) DBA [125], [126] – and Optimized-Round Robin (Optimized-RR) DBA [148]. In [148], the authors improve the gGIANT and RR DBAs by redistributing unused bandwidth of lightly-loaded transmission containers (T-CONTs) equally to heavily-loaded T-CONTs, resulting in sub-300 µs upstream delays. The authors in [149] demonstrate sub-250 µs

delay values for eCPRI functional split fronthaul traffic using the fixed-elastic DBA (FE-DBA) algorithm, which uses the ITU-T fixed bandwidth type for fronthaul traffic and the other bandwidth types (assured, non-assured and best-effort) for backhaul and FTTH traffic in the same TDM-PON. A self-adjusting DBA algorithm for the support of 5G fronthaul over PON networks was proposed in [150] by dynamically adjusting the allocation intervals to the current required fronthaul throughput based on the requests reported from the ONUs to guarantee the maximum latency delay of 250 µs. The authors in [151] proposed a PON-based mobile fronthaul transport architecture to maintain low fronthaul transport latency (≈ 100 µs) under varying mobile traffic conditions.

4.2.1 IACG DBA Algorithm Implementation

The OMNeT++ implementation of the IACG DBA algorithm is based on work done in [13]. In the work of this thesis, the original IACG DBA model has been extended to fit the requirements for use in a converged network. The major modifications to the model include the following:

- **A frame counter at the ONU.** The original IACG DBA model does not use a frame counter. The original model used the tri-modal distribution traffic generation model which generated frames of three different sizes – 64, 500, 1500 bytes. In order to keep track of the number of frames being sent from the ONU to the OLT, a fixed frame size of 1500 bytes was chosen. ITU PON standards (GPON, XG-PON and XGS-PON) transmit GEM or XGEM frames (described in section 3.4.4), which allows frames to be fragmented into several smaller frame fragments of various sizes. Each

frame fragment is transmitted individually over one or more upstream transmissions. All the frame fragments of a 1500-byte size frame must be transmitted before subsequent arriving frames in the queue are sent. The frame fragments are identified using the GEM Port ID and arrival time of the frame at the T-CONT buffer ingress in the ONU. A 1500-byte size frame that is fragmented and transmitted from the ONU cannot be said to have arrived at the OLT until the last fragment of the frame is received.

- **Traffic generators for each T-CONT type in each of the 16 ONUs in the User module.** The original model had one traffic generator for all the T-CONT types in one ONU. This was modified to have distinct traffic generators for each T-CONT type in an ONU and assigning them different bandwidth service classes with priorities.

- **Use of a 10 Gbit/s upstream line rate.** The original model used an upstream line rate of 2.5 Gbit/s, which is insufficient for the traffic needs of a converged network. In order to support the traffic in the fronthaul and obtain lower delay values, the minimum upstream line rate that is required is 10 Gbit/s [152]. In addition, a 10 Gbit/s upstream line rate increases the number of frames available for the upstream bandwidth transmission which will lead to increasing the number of frames that meet the latency and reliability requirements. This required recalculation of the downstream and upstream arrival rates in the initialisation file and changes to several modules (ONU and OLT modules) in the original model.

The aim of the ONU frame counter algorithm is to keep track of the number of frames (of size 1500 bytes) extracted from the ONU buffer to be sent to the OLT in the upstream direction and record the time of exit from the buffer to determine the queueing delay. In the original model, the queueing delay of all the frame fragments of a frame sent to the OLT was taken into account when computing the queueing delay. This problem is addressed using the frame counter such that only the queueing delay of the last frame fragment of a frame sent to the OLT is used in the computation of the queueing delay.

The flowchart for the (e.g., T-CONT 2 (T2)) frame counter in the ONU module is shown in Figure *4.1*. The algorithm starts by checking if there is a frame to be extracted from the ONU buffer (`Frame size != 0`). If there is no frame to be extracted, the algorithm ends. If there is a frame to be extracted, its frame size is added to the total frame size (`T2 total = T2 total + Frame size`), which is initially assigned a zero value (`T2 total = 0`). As more frames arrive in the queue, the algorithm proceeds to check if the total frame size (`T2 total`) is equal or greater than 1500 bytes (`T2 total >= 1500`). If the total frame size is less than 1500 bytes, the algorithm waits until the remaining frame fragment(s) for that specific frame to arrive in the ONU buffer. If the total frame size is equal or greater than 1500 bytes, the value of 1500 is subtracted from the total frame size (`T2 total = T2 total - 1500`) and the frame counter is incremented by one (`T2 counter = T2 counter + 1`) and the exit time of the frame is noted. The value of 1500 continues to be subtracted until the total frame size is no longer equal or greater than 1500 bytes.

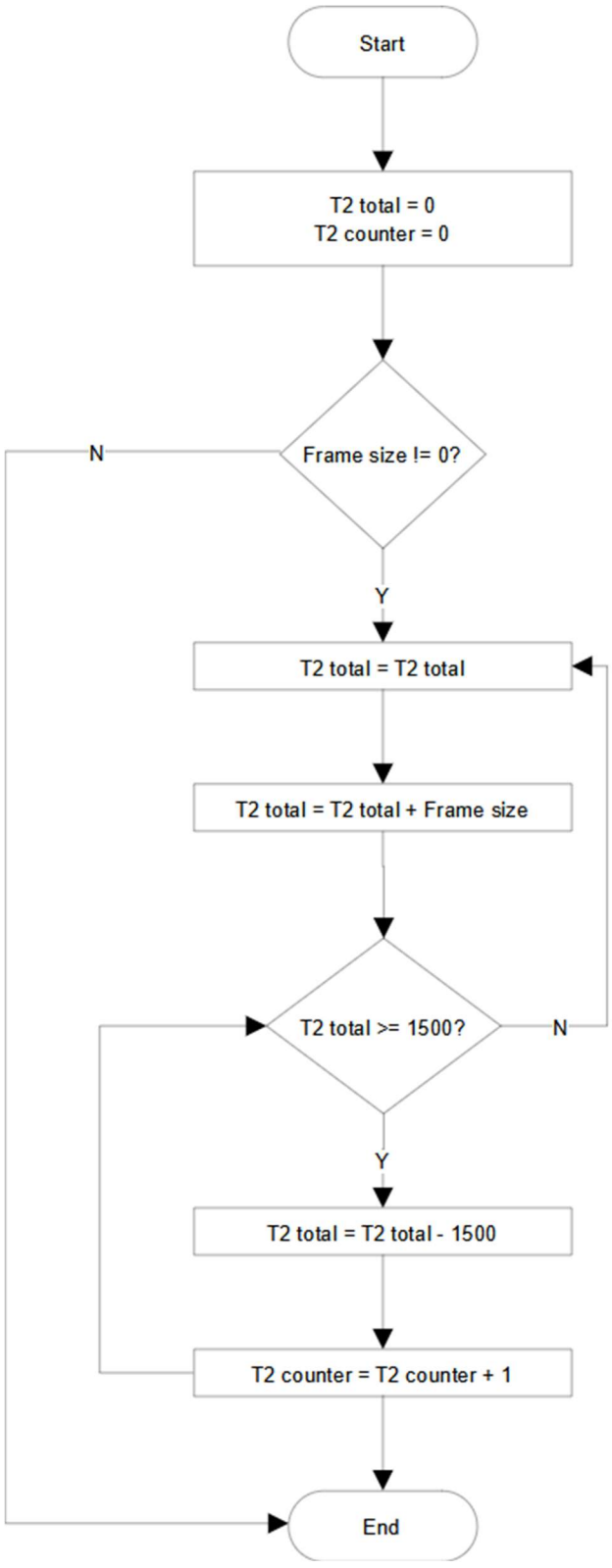


Figure 4.1: Flowchart for T2 frame counter in the ONU module.

4.3 Bandwidth Grant Cycle Process

This section investigates the bandwidth grant cycle process in relation to the ONU-OLT distance and propagation delay. This is important in order to determine how long it takes for the OLT to fulfill the bandwidth request of the ONUs. An ONU-OLT distance of 10 km (equivalent to a propagation delay of 50 μs) is used which represents the typical value for fronthaul links in a PON [140]. The remaining system parameters used are listed in Table *4.1*.

During the PON transmission, after an ONU sends its queue report to the OLT, it must wait for a time period before receiving the bandwidth grant which informs it to send the frames in its queue. While waiting for the bandwidth grant from the OLT (which occurs every 125 μs), the ONU may use the available extra bandwidth (T5) to send T-CONT frames in the upstream. Also, frames from the users continue to arrive at the ONU while waiting for the bandwidth grant from the OLT.

Figure *4.2* to Figure *4.5* show the packet trace extracts from a simulation with the timing for the bandwidth requests of the ONU and the bandwidth grants from the OLT. Column A (Grant Cycle) shows the grant cycle number with the corresponding start time of each cycle in seconds in Column B (OLT time). The ONU number for each of the 16 ONUs is shown in Column C (ONU ID) with their corresponding bandwidth grants (in bytes) for T-CONT 2 (T2), T-CONT 3 (T3), T-CONT 4 (T4) and T-CONT 5 (T5 or colorless grant, CG) in Columns D, E, F, and G, respectively. Columns A to G take place at the OLT and grants are sent in batches of 16. Column H (ONU time) show the start of the

upstream transmission cycle at the ONU in seconds with the corresponding ONU number in Column I (ONU ID). The received bandwidth allocations (in bytes) from the OLT for each T-CONT are shown in Columns J to M. The queue lengths for T-CONT 2, 3 and 4 at the start of the ONU upstream transmission cycle are shown in Columns O, P and Q, respectively. Columns H to Q take place at the ONU.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Grant	**OLT**		Bandwidth Grants				**ONU**		Received Allocations					Queue Lengths		
2	Cycle	**time**	ONU ID	T2	T3	T4	CG (T5)	**time**	ONU ID	T2	T3	T4	T5		T2	T3	T4
3	1	0.000125															
4			0	0	0	0	2430	0.00012501	0	0	0	0	0		0	0	0
5			1	0	0	0	2430	0.00013201	1	0	0	0	0		0	0	0
6			2	0	0	0	2430	0.00013901	2	0	0	0	0		0	0	0
7			3	0	0	0	2430	0.00014601	3	0	0	0	0		0	0	0
8			4	0	0	0	2430	0.00015301	4	0	0	0	0		0	0	0
9			5	0	0	0	2430	0.00016001	5	0	0	0	0		0	0	0
10			6	0	0	0	2430	0.00016701	6	0	0	0	0		0	0	0
11			7	0	0	0	2430	0.00017401	7	0	0	0	0		0	0	0
12			8	0	0	0	2430	0.00018101	8	0	0	0	2430		0	0	0
13			9	0	0	0	2430	0.00018801	9	0	0	0	2430		0	0	0
14			10	0	0	0	2430	0.00019501	10	0	0	0	2430		0	0	0
15			11	0	0	0	2430	0.00020201	11	0	0	0	2430		0	0	0
16			12	0	0	0	2430	0.00020901	12	0	0	0	2430		0	0	0
17			13	0	0	0	2430	0.00021601	13	0	0	0	2430		0	0	0
18			14	0	0	0	2430	0.00022301	14	0	0	0	2430		0	0	0
19			15	0	0	0	2430	0.00023001	15	0	0	0	2430		0	0	0

Figure 4.2: Packet trace extract showing first grant cycle at 0.000125 s (125 µs).

The first frames from the users arrive at the ONUs at 0.000274 s (274 µs). In the first example, the queue length information of ONU 5 (shown in red in Figure *4.3*) is sent to the OLT at time = 0.00028501 s (285.01 µs) and arrives at the OLT at 0.00033501 s (335.01 µs), i.e., 50 µs later. The propagation delay of 50 µs ensures that the bandwidth request of ONU 5 can only be taken into account by the OLT in the third grant cycle at time = 0.000375 s (375 µs), as shown in blue in Figure *4.4*. The OLT grants a bandwidth allocation of 1500 bytes (T2) to ONU 5, which arrives at the ONU at 425 µs.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Grant	**OLT**		Bandwidth Grants				**ONU**		Received Allocations					Queue Lengths		
2	Cycle	**time**	ONU ID	T2	T3	T4	CG (T5)	**time**	ONU ID	T2	T3	T4	T5		T2	T3	T4
20	2	0.000250															
21			0	0	0	0	2430	0.00025001	0	0	0	0	2430		0	0	0
22			1	0	0	0	2430	0.00025701	1	0	0	0	2430		0	0	0
23			2	0	0	0	2430	0.00026401	2	0	0	0	2430		0	0	0
24			3	0	0	0	2430	0.00027101	3	0	0	0	2430		0	0	0
25			4	0	0	0	2430	0.00027801	4	0	0	0	2430		0	0	0
26			5	0	0	0	2430	0.00028501	5	0	0	0	2430		1500	0	0
27			6	0	0	0	2430	0.00029201	6	0	0	0	2430		1500	0	0
28			7	0	0	0	2430	0.00029901	7	0	0	0	2430		1500	0	0
29			8	0	0	0	2430	0.00030601	8	0	0	0	2430		1500	0	0
30			9	0	0	0	2430	0.00031301	9	0	0	0	2430		0	1500	0
31			10	0	0	0	2430	0.00032001	10	0	0	0	2430		0	1500	0
32			11	0	0	0	2430	0.00032701	11	0	0	0	2430		0	1500	0
33			12	0	0	0	2430	0.00033401	12	0	0	0	2430		0	1500	0
34			13	0	0	0	2430	0.00034101	13	0	0	0	2430		0	0	1500
35			14	0	0	0	2430	0.00034801	14	0	0	0	2430		0	0	1500
36			15	0	0	0	2430	0.00035501	15	0	0	0	2430		0	0	1500

Figure 4.3: Packet trace extract showing second grant cycle at 0.000250 s (250 µs).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Grant	**OLT**		Bandwidth Grants				**ONU**		Received Allocations					Queue Lengths		
2	Cycle	**time**	ONU ID	T2	T3	T4	CG (T5)	**time**	ONU ID	T2	T3	T4	T5		T2	T3	T4
37	3	0.000375															
38			0	0	0	0	1957	0.00037501	0	0	0	0	2430		1500	0	0
39			1	0	0	0	1957	0.00038201	1	0	0	0	2430		1500	0	0
40			2	0	0	0	1957	0.00038901	2	0	0	0	2430		1500	0	0
41			3	0	0	0	1957	0.00039601	3	0	0	0	2430		1500	0	0
42			4	0	0	0	1957	0.00040301	4	0	0	0	2430		3000	0	0
43			5	1500	0	0	1957	0.00041001	5	0	0	0	2430		0	0	0
44			6	1500	0	0	1957	0.00041701	6	0	0	0	2430		1500	0	0
45			7	1500	0	0	1957	0.00042401	7	0	0	0	2430		0	0	0
46			8	1500	0	0	1957	0.00043101	8	1500	0	0	1957		0	0	0
47			9	0	780	0	1957	0.00043801	9	0	780	0	1957		0	1500	0
48			10	0	780	0	1957	0.00044501	10	0	780	0	1957		0	3000	0
49			11	0	0	0	1957	0.00045201	11	0	0	0	1957		0	0	0
50			12	0	0	0	1957	0.00045901	12	0	0	0	1957		0	1500	0
51			13	0	0	0	1957	0.00046601	13	0	0	0	1957		0	0	0
52			14	0	0	0	1957	0.00047301	14	0	0	0	1957		0	0	1500
53			15	0	0	0	1957	0.00048001	15	0	0	0	1957		0	0	0

Figure 4.4: Packet trace extract showing third grant cycle at 0.000375 s (375 µs).

The received allocation (1500 bytes) for T2 in ONU 5 is used to transmit the 1500-byte T2 frame in the queue at 0.00053501 s (535.01 µs) to the OLT, as shown in green in Figure *4.5*. It can be observed that it took 250 µs or two grant cycles before ONU 5 was able to receive a bandwidth allocation from the OLT for its bandwidth request.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Grant	**OLT**		Bandwidth Grants				**ONU**		Received Allocations					Queue Lengths		
2	Cycle	**time**	ONU ID	T2	T3	T4	CG (T5)	**time**	ONU ID	T2	T3	T4	T5		T2	T3	T4
54	4	0.000500															
55			0	1500	0	0	1696	0.00050001	0	0	0	0	1957		1500	0	0
56			1	1500	0	0	1696	0.00050701	1	0	0	0	1957		1500	0	0
57			2	1500	0	0	1696	0.00051401	2	0	0	0	1957		0	0	0
58			3	1500	0	0	1696	0.00052101	3	0	0	0	1957		0	0	0
59			4	3000	0	0	1696	0.00052801	4	0	0	0	1957		2070	0	0
60			5	0	0	0	1696	0.00053501	5	1500	0	0	1957		1500	0	0
61			6	0	0	0	1696	0.00054201	6	1500	0	0	1957		1500	0	0
62			7	0	0	0	1696	0.00054901	7	1500	0	0	1957		1500	0	0
63			8	0	0	0	1696	0.00055601	8	0	0	0	1696		1500	0	0
64			9	0		0	1696	0.00056301	9	0	0	0	1696		0	1500	0
65			10	0		0	1696	0.00057001	10	0	0	0	1696		0	1763	0
66			11	0	780	0	1696	0.00057701	11	0	780	0	1696		0	1500	0
67			12	0	780	0	1696	0.00058401	12	0	780	0	1696		0	0	0
68			13	0	0	390	1696	0.00059101	13	0	0	390	1696		0	0	1500
69			14	0	0	390	1696	0.00059801	14	0	0	390	1696		0	0	1500
70			15	0	0	390	1696	0.00060501	15	0	0	390	1696		0	0	3000

Figure 4.5: Packet trace extract showing fourth grant cycle at 0.000500 s (500 µs).

It should be noted that the colorless grant or T5 bandwidth allocation to the ONUs is made available to T2, T3, and T4 in that order. If T2 has frames in its queue, and it doesn't have any or insufficient bandwidth allocation from the OLT, then it makes use of the colorless grant (T5). The unused bandwidth allocation is then made available to T3 and subsequently T4, if needed. After cycling through the three T-CONTs, the unused bandwidth allocation is set to zero.

4.3.1 Results and Discussion

The ONU-OLT distance and corresponding propagation delay is a major factor in determining how quickly the bandwidth grant process in the OLT responds to bandwidth requests made by the ONUs. A shorter waiting time (250 µs) is achieved when using an ONU-OLT distance of 10 km rather than a distance of 20 km, which has a longer waiting time of 375 µs. This leads to the likelihood that T2 will use its own bandwidth grant allocation to transmit frames rather than using the colorless grant (T5).

4.4 IACG DBA with a 10 Gbit/s Upstream Line Rate (XGS-PON)

The 2.5 Gbit/s upstream line rate, as used in the IACG DBA algorithm in the simulation model in [15], is insufficient for carrying 5G traffic. In order to support the traffic in the fronthaul and obtain lower delay values, the minimum upstream line rate that is required is 10 Gbit/s [152]. By examining the queueing delay against normalised load, the results obtained can be made more generally applicable, especially for higher upstream data rates beyond 10 Gbit/s, such as 40 Gbit/s (NG-PON2) [108] and higher speed PONs (25, 50, 100 Gbit/s) [109] [153] [154] [115].

4.4.1 System Model

The system model consists of three sets of ONUs at the customer premises to accommodate the following types of traffic: (1) fronthaul (LLS), (2) midhaul (HLS), (3) backhaul, and (4) fixed access (FTTx), as illustrated in Figure 4.6. It also includes an OLT at a central office (CO) and an optical distribution network (ODN) that includes a passive optical splitter/combiner. The simulations involve 16 ONUs at a distance of 10 km from the OLT. Two simulation scenarios are used to evaluate the IACG DBA algorithm.

The network architecture being modelled represents a PON-based 5G xHaul transport network with both higher layer split (HLS) and lower layer split (LLS) configurations. A mixed functional split scenario such as this may represent a network in which there is a need to provide for legacy and new deployments and differentiated services, e.g., dense small cells, business and residential broadband services. It could also provide for functions such as coordinated multi-point (CoMP), interference cancellation, and

differing degrees of RAN centralisation [141], [62]. The PON elements (OLT, ONU) are deployed to connect the RAN elements (CU, DU, RU). The PON must then also meet the throughput and latency requirements of the CU-DU (midhaul) and DU-RU (fronthaul) transport. In addition, to provide for comprehensive convergence possibilities, it is assumed that backhaul and fixed access traffic may be carried over the same PON.

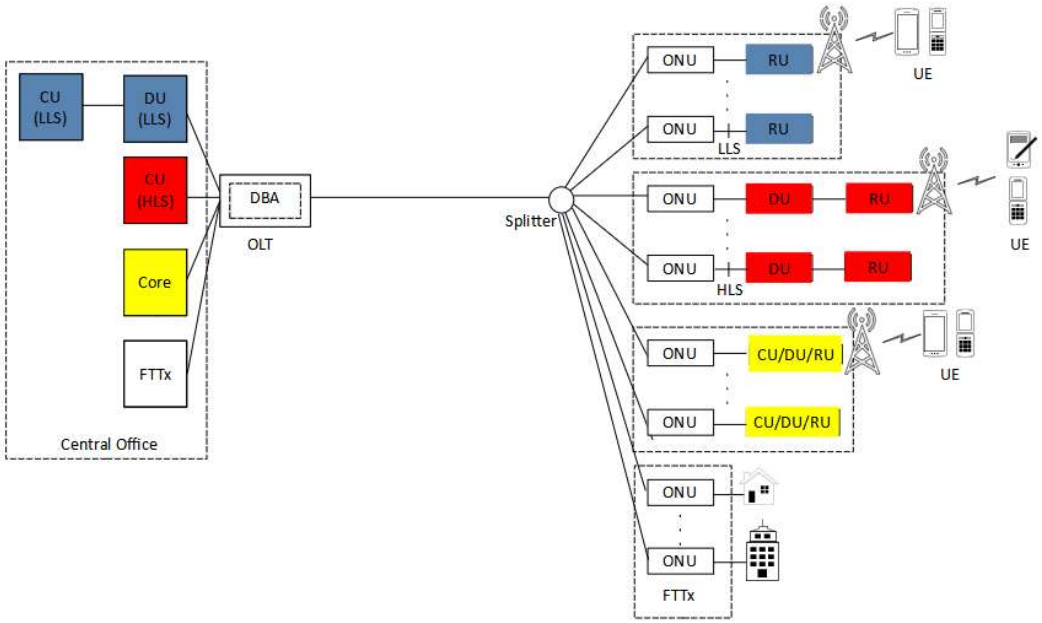


Figure 4.6: PON-based xHaul network architecture.

4.4.2 Evaluation Metrics

In the upstream direction, the PON transports Ethernet frames for the various types of traffic, such as fronthaul, midhaul, backhaul and fixed access. The upstream delay of the frames will consist of delay components such as propagation delay (fixed), serialisation delay (fixed, for a specific frame length) and queueing delay (variable). According to the 3GPP standards organisation, the one-way delay requirement for 5G fronthaul traffic is set at 250 µs [36]. Accordingly, a new queueing delay threshold of

140 µs is obtained after deducting the following delays from 250 µs: propagation delay of a 10 km optical fibre link (50 µs), DBA processing time (≈ 40 µs), optical-electrical-optical (OEO) conversion delay (≈ 15 µs) and FEC coding/decoding (≈ 5 µs) [155], [156]. The queueing delay is the only variable delay component of the upstream delay and is thus the focus in this chapter. Ethernet frames can vary in length but a 1500-byte fixed length is assumed in order to have a fixed serialisation delay value and simplify the analysis. The following performance metrics were used in the simulation:

Queueing Delay

The queueing delay is the waiting time of a frame in the ONU buffer from its arrival at the buffer's ingress to its departure from the buffer's egress. For each T-CONT, the granting of bandwidth in IACG can result in 1500-byte frames being fragmented and split over one or more upstream transmissions. The frame delay is then considered to be the queueing delays of the last frame fragments leaving the ONU buffer of each 1500-byte frame arriving at the ONU buffer. In the simulation model, this is achieved by using a counter at the ONU to keep track of all the frame fragments of each 1500-byte frame sent from the ONU to the OLT. The operation of the frame counter is discussed in section 4.2.

Average ONU Upstream Delay

To compute the ONU upstream delay of each T-CONT at the OLT, the serialisation delay is added to the queueing delay. To transmit a 1500-byte packet on a 10 Gbit/s interface, it would take 1.2 µs to serialise. In the simulations, using the ONU upstream delays for many frames, the average ONU upstream delay is computed, and its cumulative

distribution function (CDF) is obtained. This is done for each T-CONT aggregating the upstream delays obtained at all ONUs.

Percentage of Frames Meeting the Queueing Delay Requirement

The number of frames meeting the 140 μs queueing delay requirement is calculated as a percentage from the ONU upstream delay or as a percentage of frames lost (1 – 100%). This value is used to examine the distribution of delay for each T-CONT frame.

4.4.3 Simulation Setup

An XGS-PON system of 16 ONUs with downstream and upstream line rates of 10 Gbit/s [107] is implemented in the OMNeT++ open-source discrete event network simulator [157]. The simulation model is depicted in Figure *4.7*, and shows the different modules in the simulation. The User module acts as the upstream traffic generator and a sink for downstream traffic. There are traffic generators for each T-CONT type (T-CONTs 2, 3 and 4) of an ONU and this ensures the random frame generation of traffic in every ONU. The traffic generator uses an exponential distribution for the inter-arrival times, as the load varies from 0.1 to 0.9. The OLT is connected to the Server module which acts as a downstream traffic generator and upstream sink. The ONU-OLT distance is 10 km and each ONU has a buffer size of 1 MB [127]. To meet the latency requirement, the fronthaul distance should not exceed 25 km. For most telecom network operators, a typical 5G fronthaul link is limited to 10 km [57].

T-CONT 1 uses a fixed bandwidth type and is not considered in the simulation since it is mainly suited for constant bit-rate traffic, and would result in inefficient bandwidth utilisation in a network with varying user traffic. T-CONT 5 supports a combination of one or more of the other four bandwidth types (T2 – T4) [11], [105].

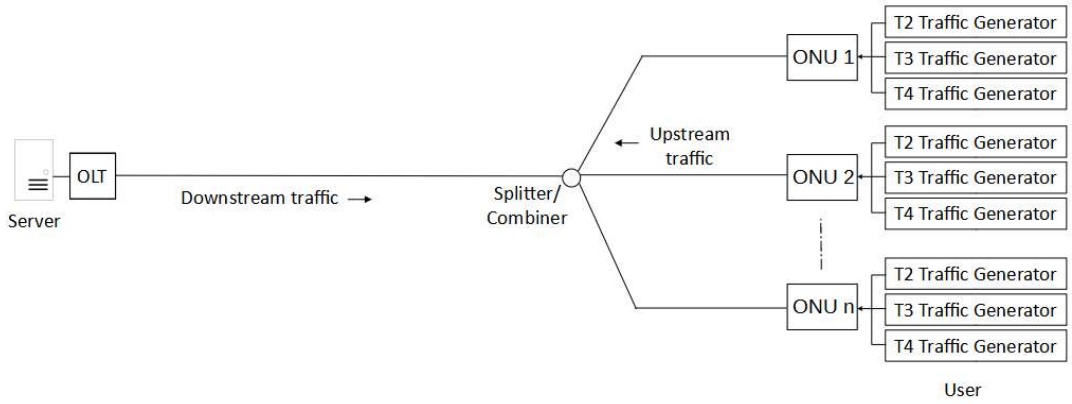


Figure 4.7: Simulation setup as implemented in OMNeT++.

The simulation parameters for the simulation setup are listed in Table *4.1* and focus on the data link layer where the DBA operates. Physical layer parameters such as the fibre type, transmission wavelength, optical modulation index of optical transmitters, and responsivity of optical detectors are not considered.

Table 4.1: Simulation parameters for IACG DBA.

Parameter	Value
OLT	1
Total number of ONUs	16
Upstream line rate	10 Gbit/s
Downstream line rate	10 Gbit/s
Link rate	200 Mbit/s
Buffer size	1 MB
Ethernet frame size	1500 bytes
T-CONT types	T2, T3 and T4
Distance between ONU and OLT	10 km
Propagation delay	5 µs/km

The two simulation scenarios used to evaluate the IACG DBA are:

Scenario 1 (T2 as fronthaul): T-CONT 2 (T2) carries fronthaul traffic, T-CONT 3 (T3) carries midhaul traffic and T-CONT 4 (T4) carries backhaul & fixed access traffic.

Scenario 2 (T2 and T3 as fronthaul): T2 carries control and signalling (fronthaul) traffic, T3 carries user data (fronthaul) traffic and T4 carries midhaul traffic.

The use of multiple T-CONTs (T2, T3 and T4) in each ONU for various types of traffic in the network is common to both scenarios.

Scenario 1

In this scenario, the fronthaul traffic with the highest priority is placed in T2, midhaul traffic with medium priority placed in T3 and backhaul & fixed access traffic with the lowest priority placed in T4. The amount of fronthaul traffic generated by the users is made larger due to higher overheads and the possible need to transmit multiple antenna streams. In a packet based 5G transport network, the fronthaul bandwidth (split options 6 or 7 where the CU/DU is located away from the RU) is significantly higher [36][62] due to overheads from scheduling control, synchronisation and the Ethernet frame [158] [159] and the scaling of the bandwidth with the number of antennas. The bandwidth allocations for the T-CONTs in this scenario are made according to the data rate requirements of the split options. The low layer split (options 6 or 7 – fronthaul) requires a much larger bandwidth allocation than the high layer split (option 2 – midhaul), approximately 2.5 times more for the uplink bandwidth [36]. This is reflected in the distribution of traffic to the various T-CONTs by giving T2 most of the traffic generated. The traffic is distributed as follows: 55% (342 Mbit/s per ONU, at 100% load) to T2, 25%

(156 Mbit/s) to T3 and 20% (124 Mbit/s) to T4. Table *4.2* lists the T-CONT bandwidth and traffic types for T2, T3 and T4 in each ONU for scenario 1. The latency requirement in the downstream direction is assumed to be met so the focus is on the traffic sent in the upstream direction.

A baseline or worst-case scenario experiment will show that when fronthaul traffic (T2) is transmitted using 16 ONUs, the average ONU upstream delay and percentage of frames lost are much higher than when fronthaul traffic is transmitted by a specific number of ONUs out of a total of 16 ONUs. In this experiment, the bandwidth allocation to T2 cannot be overallocated as there is no other T-CONT types being generated in the network so the total bandwidth (622 Mbit/s) in each ONU is allocated to T2. The results obtained showed that 97.45% of T2 frames at 80% traffic load met the 140 µs queueing delay requirement with an average ONU upstream delay of 69 µs. The use of an 80% traffic load for the experiment is to show the effect on the number of frames that meet the latency requirement at a higher traffic load because at lower traffic loads (such as 50% and 60%), 100% of frames meet the latency requirement.

Table 4.2: T-CONT bandwidth and traffic type in scenario 1 (T2 as fronthaul traffic) for IACG DBA.

T-CONT type	**Bandwidth Type**	**ONU**	**Traffic Type**
T2	Assured	0 – 15	Fronthaul
T3	Assured and Non-assured (Surplus)	0 – 15	Midhaul
T4	Best-effort	0 – 15	Backhaul

Table 4.3: Initial bandwidth allocation using 16 ONUs in scenario 1 (T2 as fronthaul traffic) for IACG DBA.

T-CONT type	ONU	Parameters	Bandwidth
T2	0 – 15	AB_{min2} = 31248 bytes, SI_{max2} = 5	400 Mbit/s
T3	0 – 15	AB_{min3} = 25000 bytes, SI_{max3} = 10	160 Mbit/s
	0 – 15	AB_{sur3} = 3124 bytes, SI_{min3} = 10	20 Mbit/s
T4	0 – 15	AB_{sur4} = 6248 bytes, SI_{min4} = 10	40 Mbit/s

The bandwidth was initially allocated to each ONU as follows: 400 Mbit/s to T2 (at 100% load), 180 Mbit/s to T3 and 40 Mbit/s to T4, as shown in Table *4.3*. The average ONU upstream delay for T2 was 65 µs at 80% load. However, from the results obtained, it was found that only 99.81% of T2 frames were meeting the 140 µs queueing delay requirement at 80% traffic load while 100% of T3 and T4 frames were meeting the less stringent 1 ms latency requirement. However, some 5G services may require a packet reliability of 99.9% or higher [26]. Therefore, the bandwidth allocation of T2 was increased to 560 Mbit/s to ensure that T2 frames access the network with minimal delay. The bandwidth allocations to T3 and T4 were lowered to 40 Mbit/s and 20 Mbit/s, respectively, as shown in Table *4.4*.

Table 4.4: Bandwidth allocation using 16 ONUs in scenario 1 (T2 as fronthaul traffic) for IACG DBA.

T-CONT type	ONU	Parameters	Bandwidth
T2	0 – 15	AB_{min2} = 43748 bytes, SI_{max2} = 5	560 Mbit/s
T3	0 – 15	AB_{min3} = 1560 bytes, SI_{max3} = 5	20 Mbit/s
	0 – 15	AB_{sur3} = 1560 bytes, SI_{min3} = 5	20 Mbit/s
T4	0 – 15	AB_{sur4} = 1560 bytes, SI_{min4} = 5	20 Mbit/s

Even though, the average ONU upstream delay of T2 indicates that the latency requirement is being met, there may be a significant number of frames that do not meet it. This requires further investigation by examining the distribution of delay for T2 in order to get a realistic analysis of the scenario. Using the new bandwidth allocation values in Table *4.4*, it was observed that 99.96% of T2 frames at 80% traffic load meet the 140 µs queueing delay requirement and that the average ONU upstream delay of T2 at all traffic loads does not exceed 70 µs. At 80% load, the average ONU upstream delay of T3 and T4 is 69.46 µs and 71.37 µs, respectively, thus satisfying the 1 ms requirement for midhaul and backhaul traffic. When compared with the baseline or worst-case experiment, the results for the bandwidth overallocation experiment showed a 2.56% increase in the number of T2 frames meeting the 140 µs queueing delay requirement and a 7.25% decrease in the average ONU upstream delay. Therefore, the overallocation of bandwidth to fronthaul traffic (T2) shows an increased improvement in performance for meeting the 140 µs queueing delay requirement at 80% load.

In order to replicate a mixed functional split scenario, instead of sixteen ONUs (see Table *4.2*) transmitting all three types of T-CONTs as described above, specific ONUs are chosen to transmit T2, T3 and T4 separately. Therefore, nine of the sixteen ONUs (56.25%) transmit only T2, four ONUs (25%) transmit only T3 and three ONUs (18.75%) transmit only T4, as shown in Table *4.5*. In the case where there is no over-allocation of bandwidth, 98.42% of T2 frames at 80% traffic load meet the 140 µs queueing delay requirement. However, when there is an over-allocation of bandwidth for T2, 99.18% of T2 frames meet the 140 µs queueing delay requirement with an average ONU upstream

delay of 66 µs. Figure *4.8* shows the CDF of the queueing delay at 80% traffic load, with the vertical purple line showing the 140 µs queueing delay requirement.

Table 4.5: Bandwidth allocation using specific ONUs in scenario 1 (T2 as fronthaul traffic) for IACG DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 43748 bytes, SI_{max2} = 5	560 Mbit/s
T3	9 – 12	AB_{min3} = 1560 bytes, SI_{max3} = 5	20 Mbit/s
	9 – 12	AB_{sur3} = 1560 bytes, SI_{min3} = 5	20 Mbit/s
T4	13 – 15	AB_{sur4} = 1560 bytes, SI_{min4} = 5	20 Mbit/s

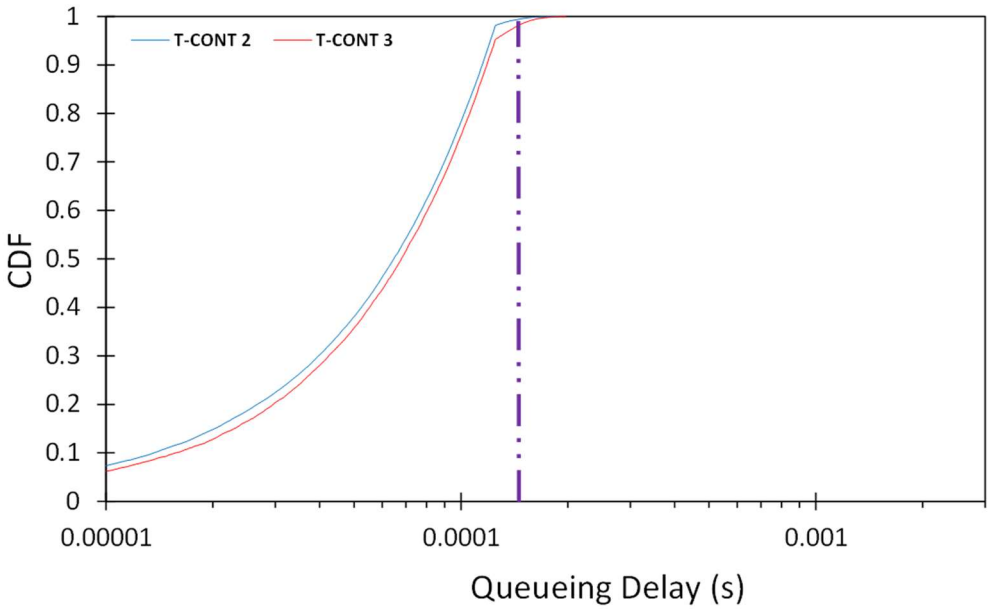


Figure 4.8: CDF of queueing delay at 80% traffic load in scenario 1 with frame size of 1500 bytes for IACG DBA.

Scenario 2

Scenario 2 is much more challenging for the PON since both T2 and T3 carry fronthaul traffic that must meet stringent latency requirements (< 250 µs). In this scenario, nine ONUs (ONU 0 to 8) transmit both control and signalling (fronthaul) traffic and user data (fronthaul) traffic in T2 and T3, respectively (shown in Table *4.6*). The total traffic

generated by the users for ONU 0 to 8 is split between T2 and T3 and allocated as 22% (136.84 Mbit/s per ONU, at 100% offered load) to T2 and 78% (485.18 Mbit/s) to T3. The remaining seven ONUs (ONU 9 to 15) transmit only T4 with 100% (622.08 Mbit/s) of the traffic generated by the users allocated to T4. The 22%:78% ratio is based on having the same number of ONUs for fronthaul traffic (nine) in scenario 1 such that control and signalling (fronthaul) traffic is always a smaller proportion of the user data (fronthaul) traffic.

As shown in Table *4.7*, T3 is allocated the highest amount of bandwidth since it carries user data (fronthaul) traffic. However, T2 has higher priority but requires less bandwidth to transmit control and signalling (fronthaul) traffic. The midhaul traffic carried in T4 has a less stringent latency requirement (1 ms) to meet so it is allocated the least bandwidth. Although the T4 bandwidth allocation is low, it is expected that T4 will use the colorless grant (T5) to transmit its frames. The CDF in Figure *4.9* shows that at 80% traffic load, 100% of T2 frames and 98.97% of T3 frames meet the 140 µs queueing delay requirement, as indicated with the vertical purple line.

Table 4.6: T-CONT bandwidth and traffic type in scenario 2 (T2 and T3 as fronthaul traffic) for IACG DBA.

T-CONT type	**Bandwidth Type**	**ONU**	**Traffic Type**
T2	Assured	0 – 8	Fronthaul (control and signalling)
T3	Assured and Non-assured (Surplus)	0 – 8	Fronthaul (user data)
T4	Best-effort	9 – 15	Midhaul

Table 4.7: Bandwidth allocation using specific ONUs in scenario 2 (T2 and T3 as fronthaul traffic) for IACG DBA.

T-CONT type	ONU	Parameters	Bandwidth
T2	0 – 8	AB_{min2} = 9372 bytes, SI_{max2} = 5	120 Mbit/s
T3	0 – 8	AB_{min3} = 35936 bytes, SI_{max3} = 5	460 Mbit/s
	0 – 8	AB_{sur3} = 1560 bytes, SI_{min3} = 5	20 Mbit/s
T4	9 – 15	AB_{sur4} = 1560 bytes, SI_{min4} = 5	20 Mbit/s

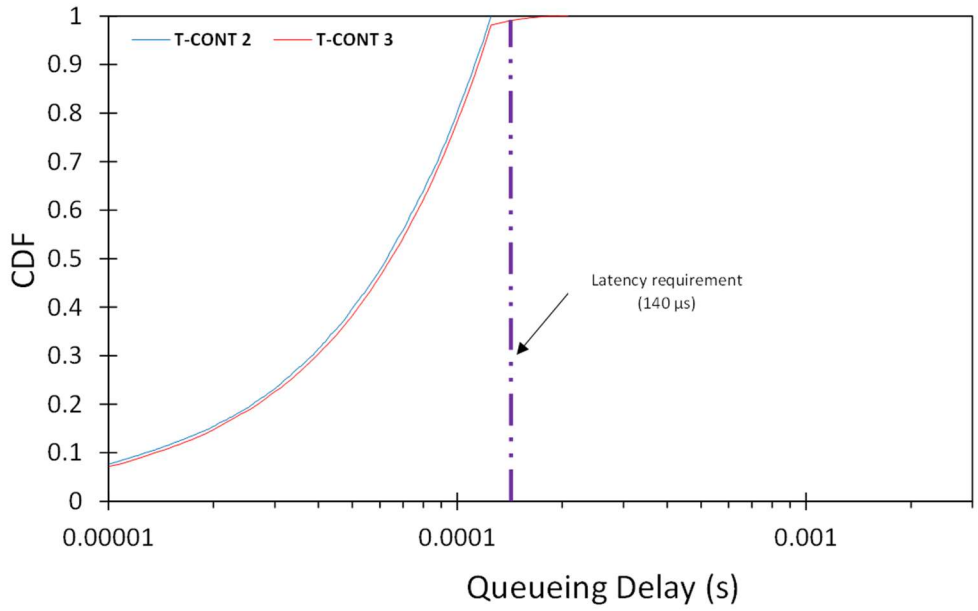


Figure 4.9: CDF of queueing delay at 80% traffic load in scenario 2 with frame size of 1500 bytes for IACG DBA.

4.4.4 Results and Discussion

Meeting the delay requirements largely depends on the amount of bandwidth allocated to each T-CONT and the traffic load. When compared to other DBAs, the IACG DBA [127] provides lower delay values for T2, T3 and T4 at all loads by sending bandwidth grants every downstream frame and assigning the unallocated bandwidth of the upstream frame at the end of the DBA cycle to each ONU equally (colorless grant phase). The GIANT DBA [125], [126] does not have a colorless grant phase and only grants bandwidth

once during a service interval (SI), so it was found to give higher delays for the traffic classes at all loads [146][148]. Also, the RR DBA [146], gGIANT DBA [147] and optimized RR DBA [148], which all use a 2.5 Gbit/s upstream line rate, show average upstream delay values at slightly less or higher than 300 μs at the same per-ONU traffic load used in the optimized RR DBA [148]. Note that the authors in [148] did not simulate traffic for the optimized RR DBA in the LLS point but used a delay requirement for the HLS point. When using a 2.5 Gbit/s upstream rate for the IACG DBA, the results obtained show an average upstream delay of 90.89 μs (< 100 μs). However, the emphasis is on the percentile delay result which shows that 86.06% of T2 frames at 80% traffic load meet the 140 μs queueing delay requirement in scenario 1. As has been discussed earlier in scenarios 1 and 2, the percentage of frames meeting the 140 μs queueing delay requirement is significantly higher when the upstream line rate is 10 Gbit/s. This is because more frames are available for the upstream bandwidth transmission. An upstream line rate of 10 Gbit/s can accommodate 155,520 bytes compared to 38,880 bytes when using a 2.5 Gbit/s upstream line rate.

4.4.5 Frame Size Comparison

In Scenarios 1 and 2, a fixed frame size of 1500 bytes was assumed for all traffic generated by the users. In this section, in order to study the effect of the frame size on the queueing delay, two other frame sizes (1000 and 500 bytes) are investigated using a 10 km ONU-OLT distance for both scenarios.

Figure *4.10* shows the CDF of the queueing delay at 80% traffic load for a frame size of 500 bytes in scenarios 1 and 2. It can be observed that 100% of T2 frames in scenario 1

meet the 140 µs queueing delay requirement, as indicated with the vertical purple line, and in scenario 2, 100% of both T2 frames and T3 frames also meet the 140 µs queueing delay requirement.

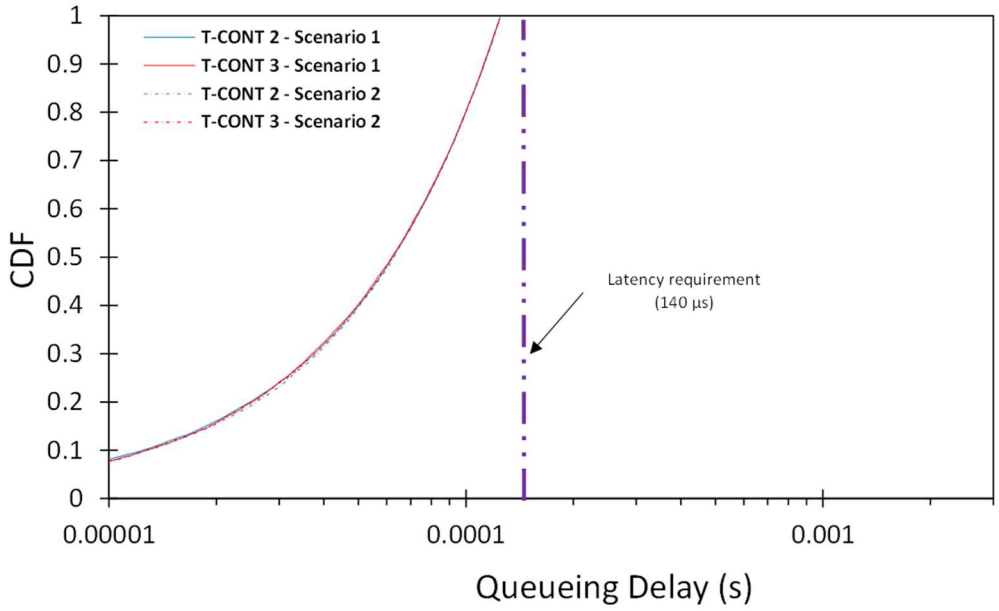


Figure 4.10: CDF of queueing delay at 80% traffic load in scenarios 1 and 2 with frame size of 500 bytes for IACG DBA.

4.4.6 Results and Discussion

A summary of the results for the different frame sizes in scenario 1 and 2 at 80% traffic load are shown in Table *4.8* and Table *4.9*. The results show that as the frame size decreases from 1500 bytes to 500 bytes, the percentage of frames meeting the 140 µs queueing delay requirement increases while the average ONU upstream delay decreases. The reason for the steady increase of T-CONT frames meeting the delay requirement is that during the transmission of frames from the ONU to the OLT, a lot more frames avoid the need for fragmentation and are being transmitted in one upstream cycle leaving fewer frame fragments waiting for subsequent upstream cycles

to be transmitted. In the case where different frame sizes are generated, the maximum Ethernet frame size of 1500 bytes will be the worst-case scenario. For 500-byte frames sizes, 100% of T2 frames will meet the 140 µs queueing delay requirement, as shown in both cases above.

Table 4.8: Frame size comparison in scenario 1 (T2 as fronthaul traffic) at 80% load with a 10 Gbit/s upstream line rate for IACG DBA.

T2		
Frame Size (bytes)	CDF (%)	Average Queueing Delay (µs)
1500	99.18	66.00
1000	100	63.66
500	100	63.52

Table 4.9: Frame size comparison in scenario 2 (T2 and T3 as fronthaul traffic) at 80% load with a 10 Gbit/s upstream line rate for IACG DBA.

	T2		**T3**	
Frame Size (bytes)	CDF (%)	Average Queueing Delay (µs)	CDF (%)	Average Queueing Delay (µs)
1500	100	63.94	98.97	66.05
1000	100	63.94	100	63.71
500	100	63.94	100	63.71

4.5 IACG DBA with a 2.5 Gbit/s Upstream Line Rate (XGS-PON)

In this section, the results obtained from using a 2.5 Gbit/s upstream line rate are discussed and compared with the results in section 4.4. The results, as seen in Table *4.10* and

Table 4.11, show that using a 2.5 Gbit/s upstream line rate worsens the proportion of T2 frames meeting the 140 µs queueing delay requirement by 21.9% and 25.3% in scenarios 1 and 2, respectively when compared to using a 10 Gbit/s upstream line rate (see Table *4.8* and Table *4.9*) and the average ONU upstream delay of T2 in scenarios 1 and 2 decrease by 9.6% and 12.7%, respectively. The proportion of frames meeting the queueing delay requirement for T2 is 89.54% in scenario 1 while its average ONU

upstream delay is 84.54 µs. The proportion of frames meeting the queueing delay requirement scenario 2 is 87.28% and 84.47% for T2 and T3, respectively, while the average ONU upstream delay for T2 and T3 are 85.65 µs and 90.74 µs, respectively. In both scenarios, T4 meets its 1 ms latency requirement. The results therefore justify the need to use a 10 Gbit/s upstream line rate, which is consistent with the recommendation of the O-RAN Alliance specifications [152] for fronthaul traffic in a converged fronthaul network. The frame size comparison at 80% traffic load is shown in *Table 4.10* and Table 4.11 for scenarios 1 and 2, respectively.

Table 4.10: Frame size comparison in scenario 1 (T2 as fronthaul traffic) at 80% load with a 2.5 Gbit/s upstream line rate for IACG DBA.

T2		
Frame Size (bytes)	CDF (%)	Average ONU Upstream Delay (µs)
1500	86.06	90.89
1000	88.85	86.47
500	98.14	71.80

Table 4.11: Frame size comparison in scenario 2 (T2 and T3 as fronthaul traffic) at 80% load with a 2.5 Gbit/s upstream line rate for IACG DBA.

	T2		**T3**	
Frame Size (bytes)	CDF (%)	Average ONU Upstream Delay (µs)	CDF (%)	Average ONU Upstream Delay (µs)
1500	94.51	75.06	78.03	102.47
1000	100	67.17	83.19	94.41
500	100	67.23	97.55	72.53

4.6 Experiments for Scenarios 1 and 2 at all traffic loads

Three more experiments are conducted for scenarios 1 and 2 at all traffic loads in order to further analyse the results. In the first experiment (Bandwidth allocation proportional to maximum load), the total available bandwidth in the upstream is shared in proportion to the number of ONUs carrying T2, T3 and T4 traffic in scenario 1, as follows: 350 Mbit/s to T2, 156 Mbit/s to T3 and 116 Mbit/s to T4, as shown in Table *4.12*. The bandwidth is

totally allocated independent of the actual load. Table *4.13* lists the bandwidth allocation for the first experiment in scenario 2. In the second experiment (Bandwidth allocation proportional to actual load), the bandwidth in the upstream is allocated in proportion to the specific traffic load. In the third experiment, bandwidth overallocation for the fronthaul traffic (T2) is used, in order to reduce the delay and improve the performance of meeting the 140 µs queueing delay requirement at all traffic loads.

Table 4.12: Bandwidth allocation proportional to maximum load in scenario 1 (T2 as fronthaul traffic) for IACG DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 27337 bytes, SI_{max2} = 5	350 Mbit/s
T3	9 – 12	AB_{min3} = 11760 bytes, SI_{max3} = 5	151 Mbit/s
	9 – 12	AB_{sur3} = 390 bytes, SI_{min3} = 5	5 Mbit/s
T4	13 – 15	AB_{sur4} = 9113 bytes, SI_{min4} = 5	116 Mbit/s

Table 4.13: Bandwidth allocation proportional to maximum load in scenario 2 (T2 and T3 as fronthaul traffic) for IACG DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 6015 bytes, SI_{max2} = 5	77 Mbit/s
T3	0 – 8	AB_{min3} = 21150 bytes, SI_{max3} = 5	271 Mbit/s
	0 – 8	AB_{sur3} = 170 bytes, SI_{min3} = 5	2 Mbit/s
T4	9 – 15	AB_{sur4} = 21260 bytes, SI_{min4} = 5	272 Mbit/s

4.6.1 Results and Discussion

Figure *4.11* shows the T2 frame loss ratio for the three experiments in scenario 1 for the IACG DBA at various traffic loads. It can be observed that where there is bandwidth overallocation to T2, the frame loss is the smallest. This is also the case for the T3 frame loss in scenario 2, which carries user data (fronthaul) traffic, as shown in Figure *4.12*. In

scenario 2, the frame loss for T2, which carries control and signalling (fronthaul) traffic, is zero. This is due to its smaller amount of T2 traffic being transmitted and T2 having a higher priority over T3 in bandwidth allocation.

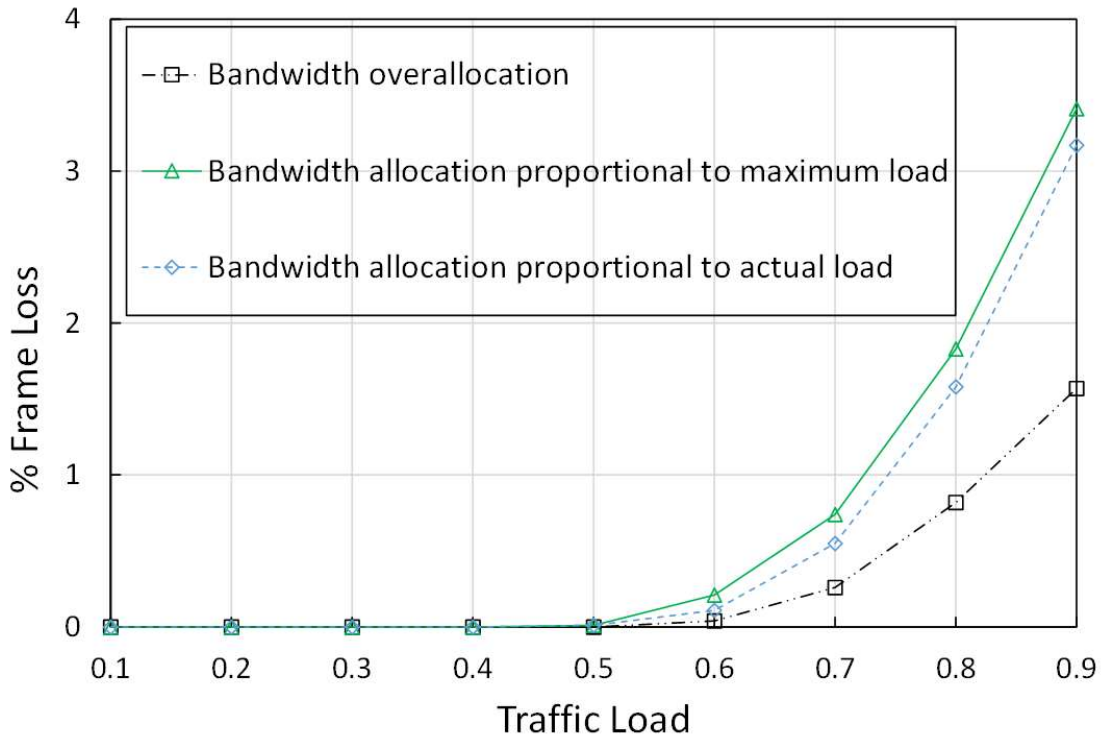


Figure 4.11: Percentage of T2 frame loss vs. traffic load in scenario 1 (T2 as fronthaul) for IACG DBA.

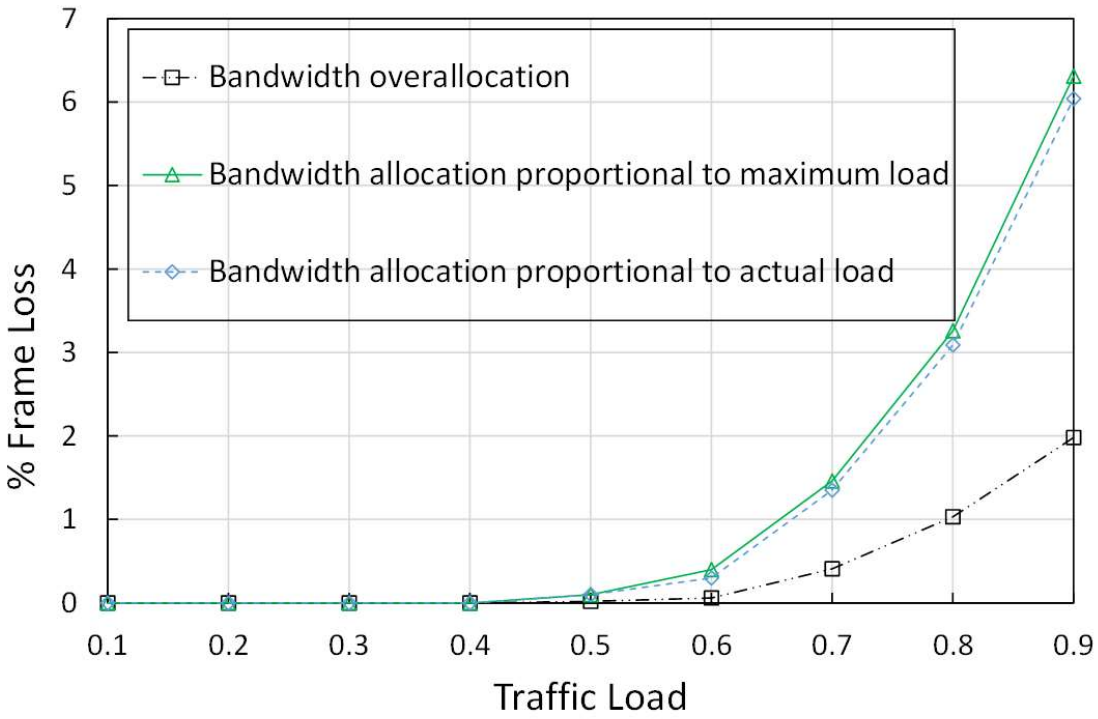


Figure 4.12: Percentage of T3 frame loss vs. traffic load in scenario 2 (T2 and T3 as fronthaul) for IACG DBA.

4.7 Scenarios 1 and 2 with increased number of ONUs

In order to further test the performance of the IACG DBA, a fourth experiment is conducted in which the number of ONUs carrying fronthaul traffic for scenarios 1 and 2 is increased from 9 to 15 ONUs with only 1 ONU carrying midhaul traffic in scenario 2 or both midhaul and backhaul/fixed access traffic in scenario 1. Scenario 1 uses one ONU, with T3 carrying midhaul traffic and T4 carrying backhaul/fixed access traffic in the ratio 57:43 (T3:T4) while scenario 2 uses one ONU with T4 carrying midhaul traffic.

In the IACG DBA, with the increased number of ONUs for fronthaul traffic, it is observed that there is a higher frame loss. In scenario 1, the frame loss for fronthaul traffic at 90% traffic load increased from 1.57% to 4.51%. There is a decrease in the average ONU

upstream delay for midhaul (81 µs → 67 µs) and a significant increase for backhaul traffic (82 µs → 164 µs). This is due to only one ONU being used for both midhaul (T3) and backhaul (T4) traffic with a larger proportion of traffic generated going to midhaul traffic and the backhaul traffic having a lower priority in bandwidth allocation. In scenario 2, the frame loss for control and signalling – fronthaul traffic (T2) remains the same at 0% while user data – fronthaul traffic (T3) increased from 1.98% to 5.87%. The midhaul traffic (T4) has a large increase in average ONU upstream delay from 82 µs to 207 µs. This is due to only one ONU (previously 7 ONUs) carrying T4 traffic and the T4 traffic class having the lowest priority.

4.8 Conclusion

In this chapter, the performance of the IACG DBA was presented as being suitable for meeting the latency requirement in an PON-based converged transport network, provided an adequate allocation of bandwidth is made available for fronthaul traffic. The delay distribution of T-CONT frames in a PON-based converged transport network with different traffic priorities was analysed to show the proportion of frames that meet the different latency requirements. The results obtained using an upstream line rate of 10 Gbit/s (XGS-PON) was also compared with those when using an upstream line rate of 2.5 Gbit/s (XG-PON), which is insufficient for the traffic needs in a converged network. A 10 Gbit/s upstream line rate can accommodate four times more bytes for the upstream bandwidth transmission.

The IACG DBA was used because of its colorless grant process which allocates unallocated bandwidth equally to T-CONTs in the ONU at the end of a DBA cycle, thereby ensuring higher bandwidth efficiency. The results obtained from the simulations show that the average ONU upstream delays for fronthaul traffic is below the 250 µs latency requirement while midhaul and backhaul traffic can meet its 1 ms latency requirement. Further analysis of the delay distribution shows that 99% of fronthaul traffic at 80% load in both scenarios are below the 250 µs latency requirement. The frame loss for fronthaul traffic in scenario 1 for the bandwidth overallocation experiment is the smallest when compared with the experiments for bandwidth allocation proportional to maximum load and actual load at various traffic loads (10% to 90%). In scenario 2, the frame loss for user data (fronthaul) traffic is zero while the frame loss for control and signalling (fronthaul) traffic is the smallest. However, when the number of ONUs for fronthaul traffic is increased, there is a higher frame loss for fronthaul traffic in scenario 1 and user data (fronthaul) traffic in scenario 2 while the frame loss for control and signalling (fronthaul) remains the same at 0%. Overall, the results indicate that the latency requirements of fronthaul traffic can be met by over-allocating bandwidth to it relative to other types of traffic (midhaul, backhaul and fixed access) in the network. The non-fronthaul traffic have less stringent latency requirements and can be assumed that they would make use of the colorless grant to access unused bandwidth. The results obtained can be adapted for higher upstream line rates of 40 Gbit/s, where there will be a higher proportion of frames meeting the queueing delay requirement and the delays are likely to be lower.

With the change of the upstream line rate from 2.5 Gbit/s to 10 Gbit/s, the proportion of T-CONT frames meeting the latency requirement for fronthaul traffic increased by 21.9% and 25.3% in scenarios 1 and 2, respectively. Also, the average ONU upstream delay decreased by 9.6% and 12.7% in scenarios 1 and 2, respectively.

In order to further reduce the upstream latency, it is necessary for the mobile base station to share the mobile scheduling information with the DBA at the OLT in advance of the arrival of uplink data. This concept is implemented in Chapter 5 by proposing a hybrid DBA that operates with the IACG DBA and cooperative DBA.

5. Proposed Hybrid DBA for Converged Fronthaul operating with IACG DBA and Cooperative DBA

5.1 Introduction

In this chapter, a proposed hybrid DBA for low latency in a fronthaul network is introduced. The proposed hybrid DBA operates with a status reporting DBA (SR-DBA) – the IACG DBA [127], which was discussed in Chapter 4.

In order to satisfy the latency requirement in a converged fronthaul, the concept of coordinating scheduling between the mobile scheduler (in the CU/DU) and the PON scheduler (in the OLT) was introduced in [142] and has been recognised by the ITU-T as a DBA method known as the cooperative DBA (CO-DBA) [160]. A corresponding bidirectional interface is needed by the CO-DBA to facilitate the scheduling so an open interface known as the cooperative transport interface (CTI) was specified by the O-RAN Alliance [152], [160]. The CTI is used to receive user equipment (UE) scheduling information from the mobile scheduler and pass it to the PON scheduler.

Tashiro et al. [142] proposed the first CO-DBA scheme for TDM-PON, which was based on the interleaved polling with adaptive cycle time (IPACT) algorithm [161]. In this scheme (M-DBA), the OLT receives the scheduling information from the centralised baseband unit (BBU) and allocates time slots according to the received information. Uzawa et al. [162] proposed a practical M-DBA scheme that properly allocates time slots by estimating the arrival period of the data from the scheduling information. Also,

Uzawa et al. [163], [164] proposed a DBA scheme that combines fixed bandwidth allocation (FBA) with CO-DBA to converge mobile fronthaul and Internet of Things (IoT) networks on a single TDM-PON. The proposed DBA scheme allocates bandwidth differently for each sub-network. However, using an FBA scheme cannot satisfy the demands of varying traffic in a 5G mobile fronthaul and has the disadvantage of low bandwidth allocation efficiency. Zhou et al. [165], [166] proposed a mobile fronthaul architecture based on a PHY functional split with a unified mobile and PON scheduler known as Mobile-PON. The use of the unified Mobile-PON scheduler eliminates the need for additional scheduling delay at the PON by combining the PHY functional split and Mobile-PON mapping scheme. Hisano and Nakayama [145] proposed the introduction of a forwarding order control in CO-DBA to maximise the number of ONUs that can transmit fronthaul streams within the requirement and the bandwidth usage efficiency in a fronthaul link. Hatta et al. [143] proposed a DBA scheme that automatically adjusts the DBA cycle length according to the traffic load in order to achieve low latency. Nomura et al. [167] proposed a DBA scheme that combines both status-reporting DBA and CO-DBA based on EPON, an IEEE PON standard. This scheme is implemented using an NG-PON2 system that only uses three ONUs with two ONUs for fronthaul and one ONU for midhaul traffic.

The DBA schemes mentioned above are mainly focused on the IEEE PON standards (e.g., EPON and 10G-EPON) with a few performing studies on ITU-T PON standards (e.g., GPON, XG-PON, XGS-PON), such as the papers in [145] and [150]. In this thesis, a hybrid DBA based on XGS-PON, an ITU-T PON standard, is proposed. The hybrid DBA is designed to operate with the cooperative DBA which transmits latency-sensitive

fronthaul traffic and IACG DBA, which has a colorless grant phase, and transmits non-fronthaul traffic with less stringent latency requirements. The proposed hybrid DBA enables a mix of different traffic types to be transmitted on the same PON while satisfying the strict latency requirement for fronthaul traffic. The latency-sensitive fronthaul traffic is handled by the cooperative DBA while the latency-tolerant non-fronthaul traffic (midhaul, backhaul and fixed access traffic) is handled by the IACG DBA, which is a ITU-T standard compliant status reporting DBA. By combining the cooperative DBA with the IACG DBA, the hybrid DBA reduces the idle period in a converged network. The idle period is the time during which the ONUs wait for their bandwidth allocation from the OLT. Reducing the idle period leads to a reduction in latency and increases the bandwidth utilisation of frames in a converged network.

In a converged fronthaul network that uses a SR-DBA (e.g., IACG DBA), the CU/DU sends the mobile scheduling information to the UE(s), as shown in Figure 5.1. The UE(s) mobile scheduling request for upstream bandwidth is assumed to be already at the CU/DU. The UE(s) uses this information to transmit its data in the upstream to the RU, which arrives at the ONU. The ONU makes a request to the OLT for bandwidth using the information of the data received from the RU. Meanwhile, the upstream data from the RU waits in the ONU until the OLT allocates bandwidth and sends a grant to the ONU. Using the grant from the OLT, the ONU can finally begin transmission of the data to the CU/DU.

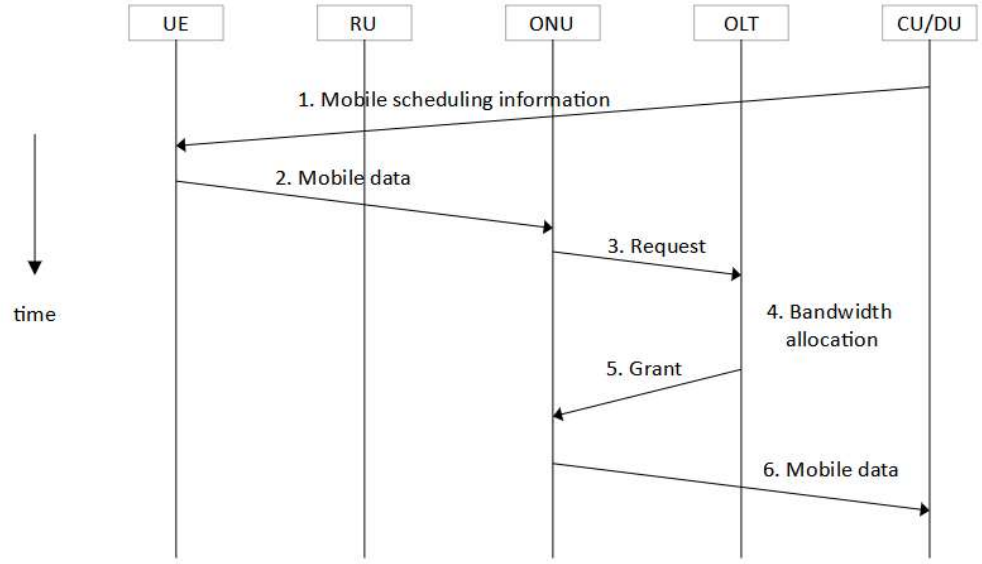


Figure 5.1: Timing diagram for a status reporting DBA in a converged fronthaul network.

The proposed hybrid DBA scheme combines a SR-DBA, in this case the IACG DBA with a CO-DBA, in order to reduce the upstream latency in a converged fronthaul network. Figure 5.2 shows the timing diagram of the CO-DBA. Again, the UE(s) mobile scheduling request is assumed to be already at the CU/DU. Since the OLT and CU/DU share a common interface (CTI), the OLT is able to access and read the scheduling information of the UE(s) at the same time it is being sent by the CU/DU. This enables the OLT to allocate upstream bandwidth in advance before the arrival of the uplink mobile data from the RU at the ONU. Using the mobile scheduling information, the OLT makes available the bandwidth allocation as close as possible to the estimated mobile data arrival time. This process allows the CO-DBA to avoid waiting for the mobile data to arrive in the ONU buffer before allocating bandwidth.

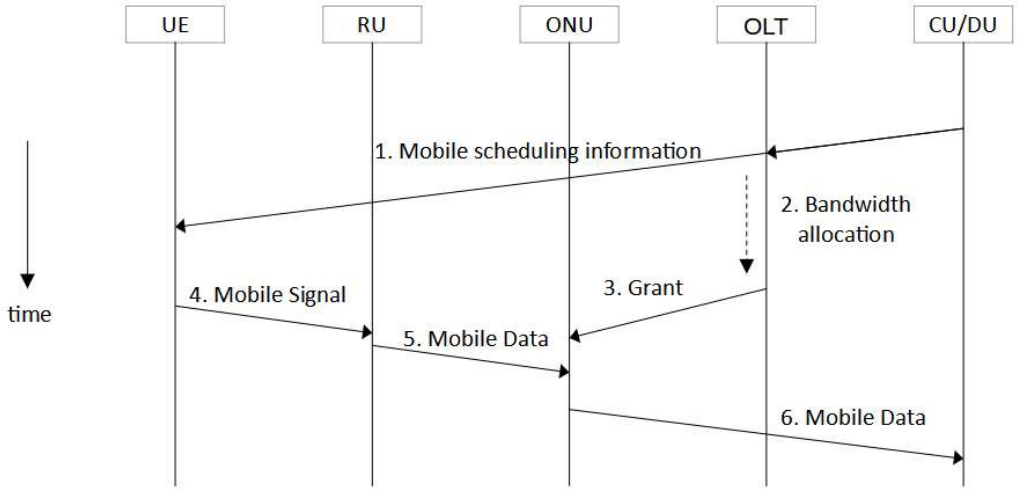


Figure 5.2: Timing diagram for the CO-DBA in a converged fronthaul network.

5.2 System Model

The system model was developed using the OMNeT++ discrete event network simulator [157], as discussed in section 4.4, to emulate how the CO-DBA would work in a converged fronthaul network using a 10 Gbit/s symmetrical PON (XGS-PON) [107]. The system model implementation of the CO-DBA, as shown in Figure 5.3, is designed to emulate its use in a real-world deployment scenario, with random frame generation for traffic in the upstream direction modelling the UE data arrival from the RU. This is achieved by creating two paths – a data path and a message path – in the model, with a relative delay between both paths. The message path represents the mobile scheduling information from the CU/DU and the data path represents the mobile data from the RU, as discussed in section 5.1. A screenshot of the simulation model as implemented in OMNeT++ is shown in Figure *5.4*.

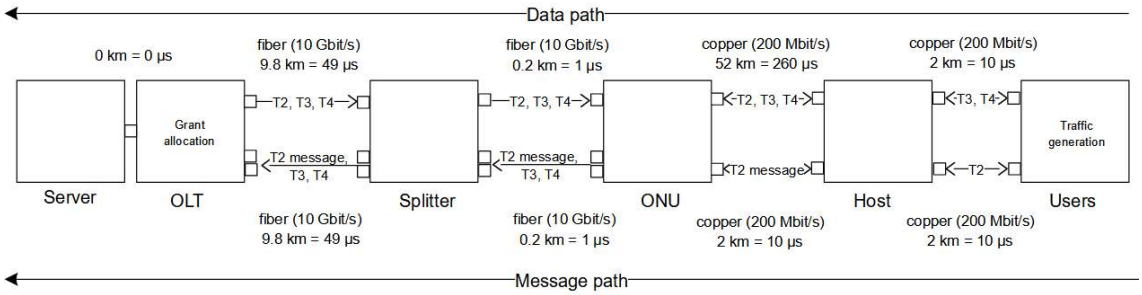


Figure 5.3: Network design diagram of simulation model in scenario 1 (T2 as fronthaul) for proposed hybrid DBA.

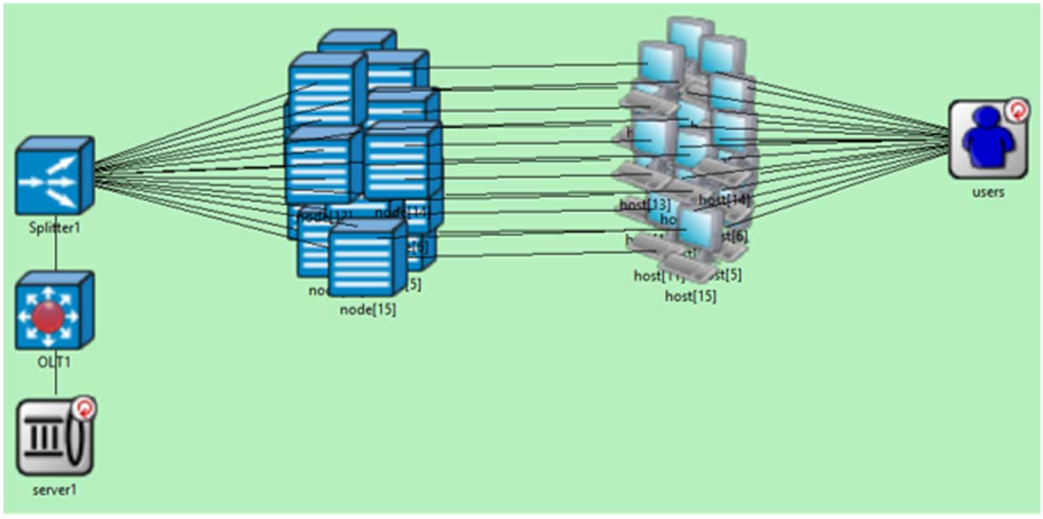


Figure 5.4: Screenshot of simulation model implemented in OMNeT++.

The simulation parameters used for the model are listed in Table *5.1*.

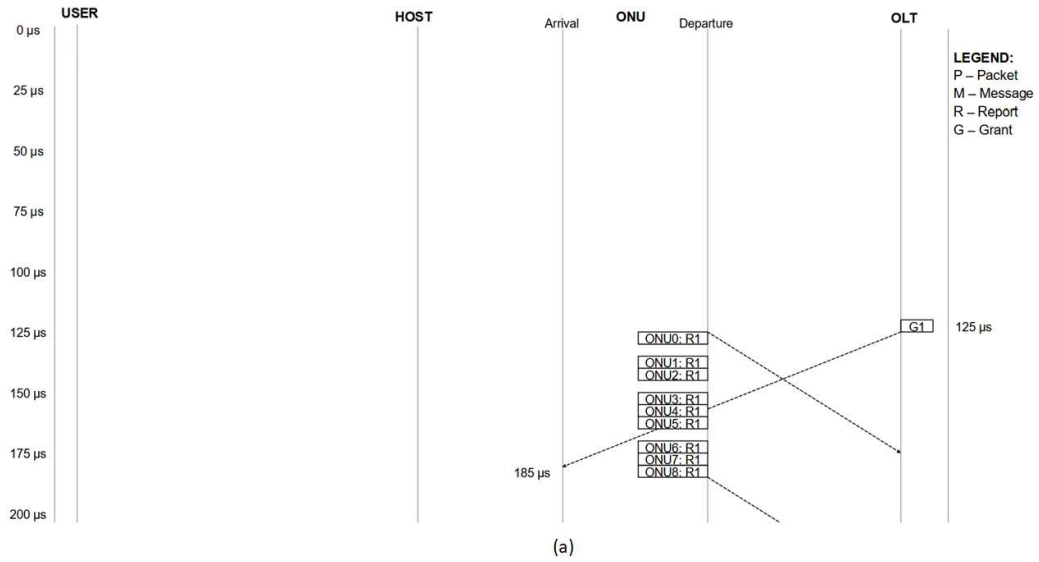
Table 5.1: Simulation parameters for hybrid DBA.

Parameter	**Value**
OLT	1
Total number of ONUs	16
Upstream line rate	10 Gbit/s
Downstream line rate	10 Gbit/s
Link rate	200 Mbit/s
Buffer size	1 MB
Ethernet frame size	1500 bytes
T-CONT types	T2, T3 and T4
Distance between ONU and OLT	10 km
Propagation delay	5 µs/km

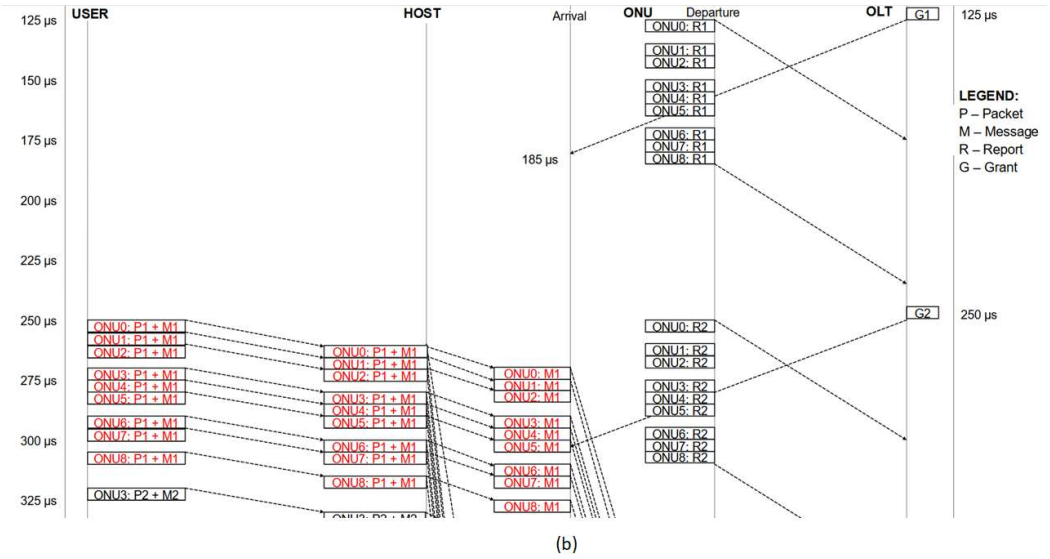
The results in section 4.3 showed that when using an ONU-OLT distance of 10 km, it takes 250 μs before an ONU's bandwidth request is fulfilled by the IACG DBA in the OLT. The 250 μs time period consists of two cycles – a 125 μs packet generation cycle and a 125 μs upstream transmission cycle. The packets generated in one cycle will all fit in the next upstream transmission cycle because of the 250 μs relative delay between the data path and the message path. The OLT sends grants to the ONU every 125 μs, in the form of a bandwidth map. After an ONU receives a grant, it has to wait for the next upstream transmission cycle. The upstream transmission cycle is the time period when packets are allowed to be transmitted from the ONU in the upstream direction. If the grant misses this cycle, it has to wait another 250 μs for the next upstream transmission cycle. A 60 μs delay between the ONU and the OLT is needed to ensure that the grants arrive at the ONU within the upstream transmission cycle. The 60 μs delay consists of a 50 μs propagation delay and 10 μs for the forwarding latency in the switching stages of the ONU and OLT [152]. Figure *5.5* shows the timing diagram for the packet generation cycle and the upstream transmission cycle for nine ONUs (ONU 0 – ONU 8) in both the message path and data path.

As shown in Figure *5.5* (a), the OLT sends the first grant (G1) at 125 μs which arrives at the ONUs 60 μs later at 185 μs. The first set of reports from the 9 ONUs are sent to the OLT during the first upstream transmission cycle from 125.01 μs (ONU0: R1) to 181.01 μs (ONU8: R1) with an interval gap of 7 μs between each ONU. The bandwidth reports arrive at the OLT before the second grant (G2) is sent at 250 μs.

The first packet generation cycle from the users start at 250 µs and the packets are sent to the OLT along with the messages (ONU0: P1 + M1 to ONU8: P1 + M1), shown in red in Figure *5.5* (b). The packets and messages travel on the data and message path, respectively and both arrive at the host 10 µs later. The messages (ONU0: M1 to ONU8: M1), continue on to the ONU while the packets wait at the host to transmit their data. On arrival at the ONU, the mobile scheduling information contained in the messages is used by the ONU in its third bandwidth report to the OLT (ONU0: R3 to ONU8: R3), as shown in Figure *5.5* (c). These reports (ONU0: R3 to ONU8: R3) arrive at the OLT before the fourth grant (G4) at 500 µs. The OLT uses the information received by the reports to make the appropriate bandwidth allocations which arrive in advance of the packets at the ONU. As shown in red in Figure *5.5* (d), the first set of packets (ONU0: P1 to ONU8: P1) which contain the mobile data all arrive at the ONU before the next upstream transmission cycle between 625.01 µs and 681.01 µs. They make use of the bandwidth allocated to send its mobile data to the OLT.

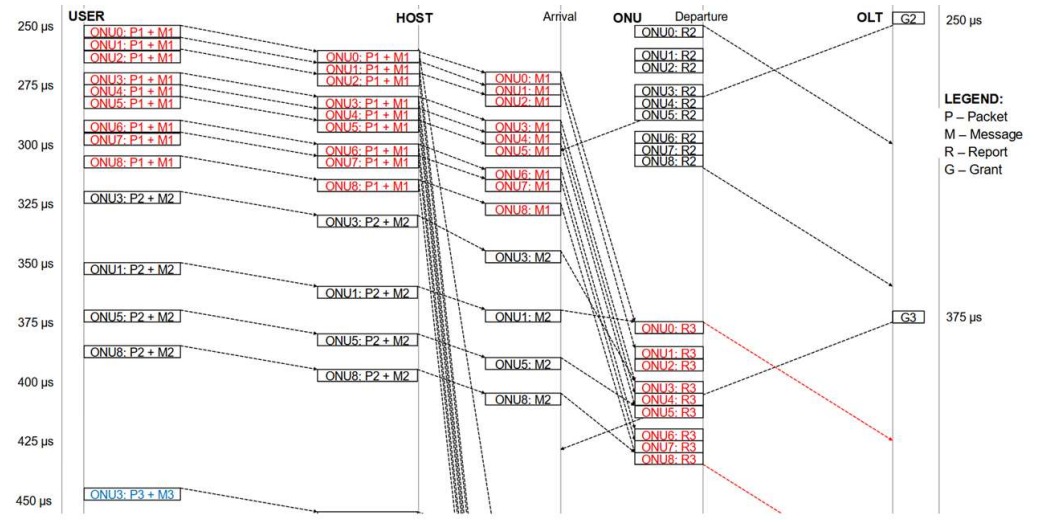


(a) First grant (G1) from OLT at 125 µs for ONUs 0 – 8.



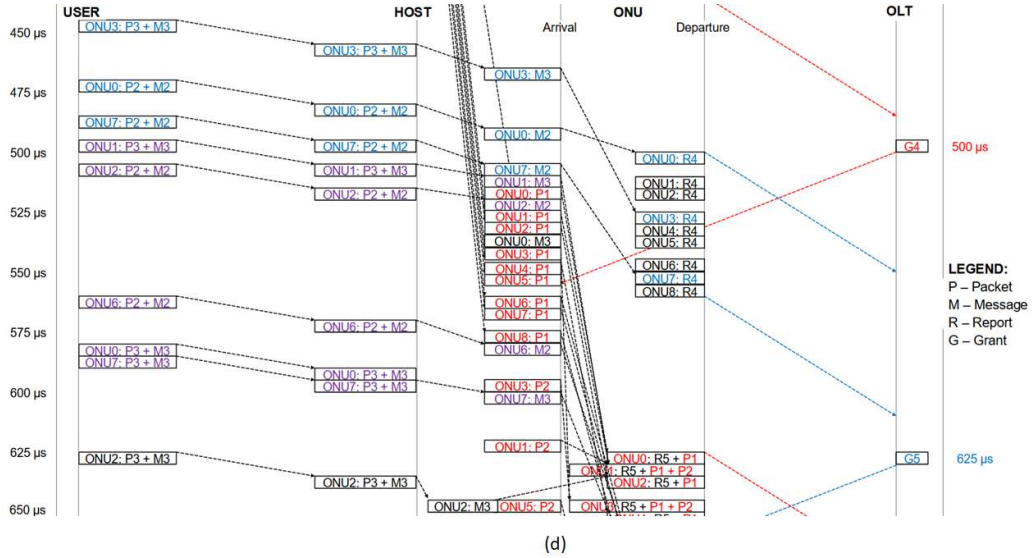
(b)

(b) First set of packets and messages (in red) from users at 250 µs for ONUs 0 – 8.



(c)

(c) First set of messages from users (in red) used in third bandwidth report (in red) to OLT.



(d)

(d) First set of packets (in red) for ONU 0 – 8 sent to OLT.

Figure 5.5: Timing diagram of proposed hybrid DBA implemented in OMNeT++.

5.3 Proposed Hybrid DBA with T2 as fronthaul traffic

In this section, a proposed hybrid DBA with T2 as fronthaul traffic is evaluated using a scenario whereby the bandwidth for T2 is not overallocated as was the case for the IACG DBA in section 4.4. In the proposed hybrid DBA, T2 does not use the DBA process so its bandwidth allocation is zero. The performance of the hybrid DBA is evaluated in terms of the average ONU upstream delay and the proportion of frames meeting the queueing delay requirement and compared with the IACG DBA. A further evaluation of the same scenario is undertaken where the number of ONUs needed to transport T2 traffic is increased. The simulation parameters used are the same listed in Table *5.1* in section 5.2.

5.3.1 Scenario 1 (T2 as fronthaul)

This simulation scenario is based on a zero-bandwidth allocation to T2 and its effect on the performance of the average ONU upstream delay and the proportion of frames meeting the queueing delay requirement when compared to the case where the T2 bandwidth is overallocated in the IACG DBA, as discussed in section 4.4. The unused T2 bandwidth is shared between T3 and T4 in proportion to the number of ONUs used by T3 and T4, as shown in Table *5.2*. In this scenario, fronthaul traffic (T2) is carried by 9 ONUs, midhaul traffic (T3) is carried by 4 ONUs and backhaul traffic (T4) is carried by 3 ONUs, similar to scenario 1 in section 4.4, and depicted in Figure *5.3*.

Table 5.2: Bandwidth allocation in scenario 1 (T2 as fronthaul traffic) for proposed hybrid DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 0 bytes, SI_{max2} = 5	0 Mbit/s
T3	9 – 12	AB_{min3} = 27344 bytes, SI_{max3} = 5	350 Mbit/s
	9 – 12	AB_{sur3} = 391 bytes, SI_{min3} = 5	5 Mbit/s
T4	13 – 15	AB_{sur4} = 20859 bytes, SI_{min4} = 5	267 Mbit/s

5.3.2 Results and Discussion

The simulation results are presented in terms of the average ONU upstream delay, the proportion of frames meeting the queueing delay requirement and the amount of colorless grant (T5) used by T2, T3 and T4. The performance of the proposed hybrid DBA is evaluated taking into account the increased bandwidth to T3 and T4. Figure *5.6* shows the cumulative distribution function (CDF) of the average ONU upstream delay for the IACG DBA and the proposed hybrid DBA, at an offered load of 80%. It can be observed that 99.18% of T2 frames meet the 140 µs queueing delay requirement (as discussed in section 4.4) and this increases to 100% in the proposed hybrid DBA.

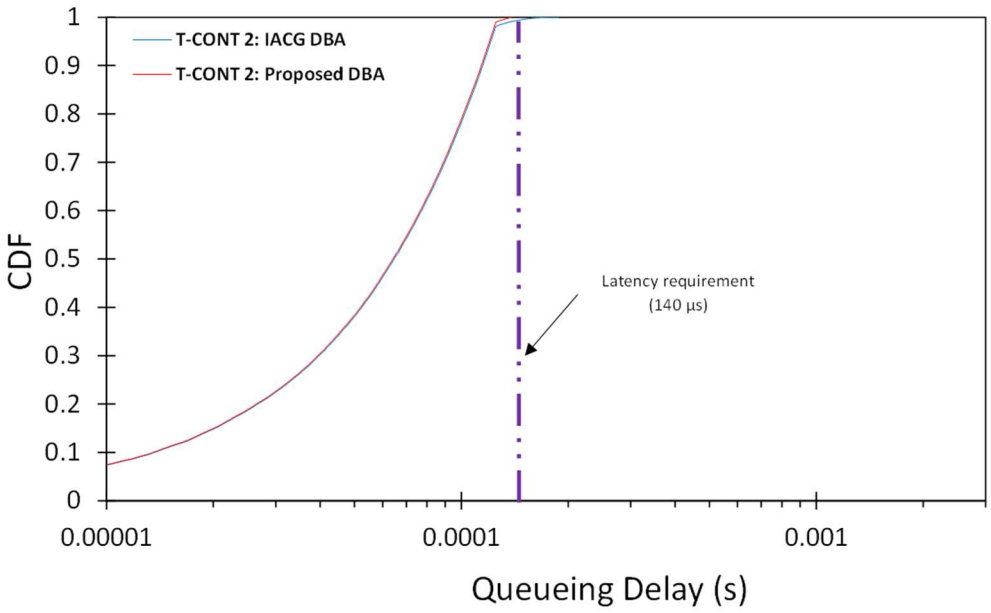


Figure 5.6: CDF of queueing delay at 80% traffic load in scenario 1 (T2 as fronthaul traffic) for the IACG DBA and proposed hybrid DBA.

In the proposed hybrid DBA, the reallocation of T2 bandwidth to T3 and T4 enables more T3 frames to utilise their own allocated bandwidth rather than using the colorless grant (T5). This is because T3 now becomes the first priority in grant allocation in the DBA process in place of T2 as was the case in the IACG DBA. This results in a 39% decrease (89% → 54%) in the amount of colorless grant bandwidth (T5) used by T3, when compared with the IACG DBA. With the increased bandwidth, the average ONU queueing delay of T3 and T4 reduces by 6% and 9% to 65.44 µs and 65.37 µs, respectively. The results for T3 and T4 also show that 100% of the frames continue to meet the 1 ms latency requirement for HLS [59].

5.3.3 Scenario 1 with increased number of ONUs

In the IACG DBA, the bandwidth for T2 was overallocated in order to obtain a better performance in terms of average ONU upstream delay and the proportion of frames

meeting the 140 µs queueing delay requirement. In this experiment, the number of ONUs used by T2 is increased in order to show that overallocation of bandwidth to T2 can be avoided and at the same time obtain improved results using the proposed hybrid DBA. Now 12 ONUs carry T2 traffic instead of 9, as was the case in section 5.3.1, while 3 ONUs and 1 ONU carry T3 and T4 traffic, respectively. The bandwidth in the upstream is shared in proportion to the number of ONUs carrying T2, T3 and T4 traffic, and are listed in Table *5.3*. This scenario uses the same simulation parameters listed in Table 4.1 in section 4.4.3.

Table 5.3: Bandwidth allocation in scenario 1 (T2 as fronthaul traffic) with increased ONUs for proposed hybrid DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8, 10 – 12	AB_{min2} = 0 bytes, SI_{max2} = 5	0 Mbit/s
T3	13 – 15	AB_{min3} = 35938 bytes, SI_{max3} = 5	460 Mbit/s
	13 – 15	AB_{sur3} = 469 bytes, SI_{min3} = 5	6 Mbit/s
T4	9	AB_{sur4} = 12188 bytes, SI_{min4} = 5	156 Mbit/s

5.3.4 Results and Discussion

The effect of increasing the number of ONUs used by T2 on the average ONU upstream delay and proportion of frames meeting the 140 µs queueing delay requirement is evaluated and compared to the IACG DBA. The results show that the proportion of frames meeting the 140 µs queueing delay requirement for T2 increased by 1.61 %, from 98.42% to 100 % in Figure *5.7*. There is also an improvement in the average ONU upstream delay (67.57 µs → 63.63 µs).

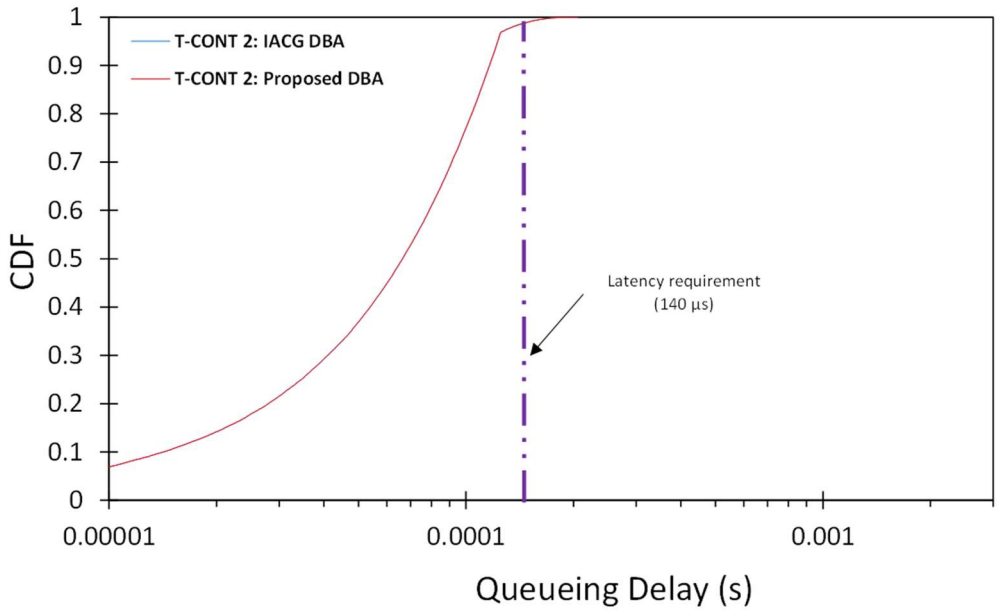


Figure 5.7: CDF of queueing delay at 80% traffic load in scenario 1 (T2 as fronthaul traffic) with increased ONUs for the IACG DBA and proposed hybrid DBA.

In this experiment, it is observed that the higher the number of ONUs used by T2 in the proposed hybrid DBA, the larger the proportion of frames that meet the 140 µs queueing delay requirement and the lower the average ONU upstream delay. This proves that the overallocation of bandwidth to T2 is not required to obtain increased performance when using the proposed hybrid DBA.

5.4 Proposed Hybrid DBA with T2 and T3 as fronthaul traffic

In this section, the proposed hybrid DBA is implemented with both T2 and T3 as fronthaul traffic, separated as control and signalling in T2 and user data in T3. There is no overallocation of bandwidth for T2 and T3 in the proposed hybrid DBA, since they will not require the use of the IACG DBA process. The proposed hybrid DBA is compared with the IACG DBA in terms of average ONU upstream delay and the proportion of frames meeting the 140 µs queueing delay requirement using a simulation scenario with

increased loads for T2 and T3. Figure *5.8* shows the simulation model for the proposed hybrid DBA as was discussed in section 5.2, along with the simulation parameter settings. The T2 and T3 messages, which contain the mobile scheduling information, travel along the message path while the T2 and T3 packets, which contain the mobile data travel along the data path.

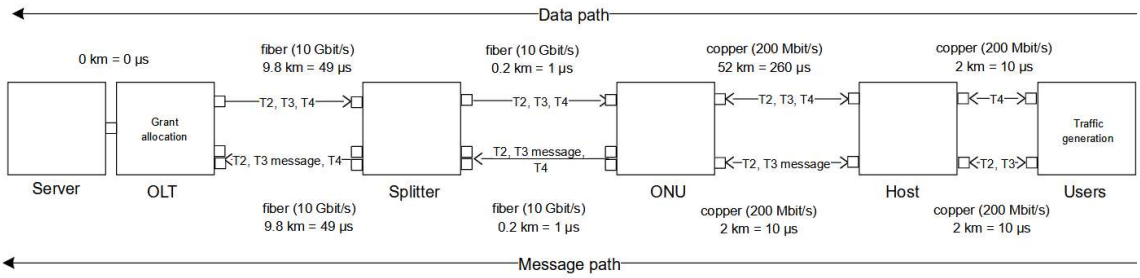


Figure 5.8: Network design diagram of simulation model in scenario 2 (T2 and T3 as fronthaul) for proposed hybrid DBA.

5.4.1 Scenario 2 (T2 and T3 as fronthaul)

This scenario is composed of 9 ONUs carrying the fronthaul traffic with control and signalling (T2) and user data (T3) in the ratio 22:78 (as discussed in section 4.4). The remaining 7 ONUs carry the midhaul traffic (T4). In the hybrid DBA, both T2 and T3 are not allocated bandwidth, so T4 is allocated the total bandwidth per ONU, as listed in Table *5.4*.

Table 5.4: Bandwidth allocation in scenario 2 (T2 and T3 as fronthaul traffic) for proposed hybrid DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 0 bytes, SI_{max2} = 5	0 Mbit/s
T3	0 – 8	AB_{min3} = 0 bytes, SI_{max3} = 5	0 Mbit/s
	0 – 8	AB_{sur3} = 0 bytes, SI_{min3} = 5	0 Mbit/s
T4	9 – 15	AB_{sur4} = 48594 bytes, SI_{min4} = 5	622 Mbit/s

5.4.2 Results and Discussion

The proportion of frames meeting the 140 µs queueing delay requirement for T2 remains unchanged at 100% as shown in Figure *5.9* and Figure *5.10*, while there is an increase, from 98.97% to 100%, for T3. The increased percentage is significant as some 5G use cases may require a packet reliability of 99.9% or higher, as described in section 2.2.3.

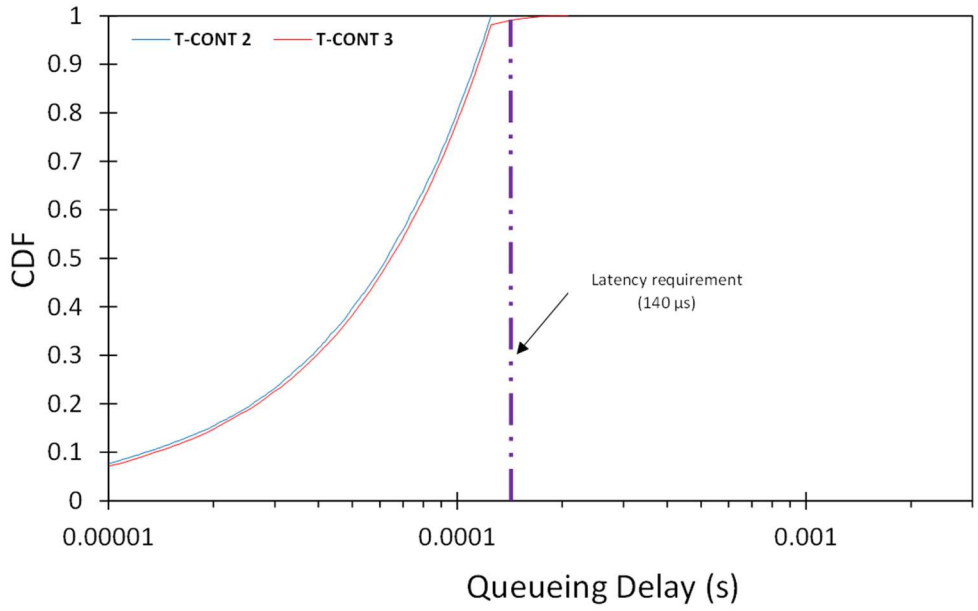


Figure 5.9: CDF of queueing delay at 80% traffic load in scenario 2 (T2 and T3 as fronthaul traffic) for IACG DBA.

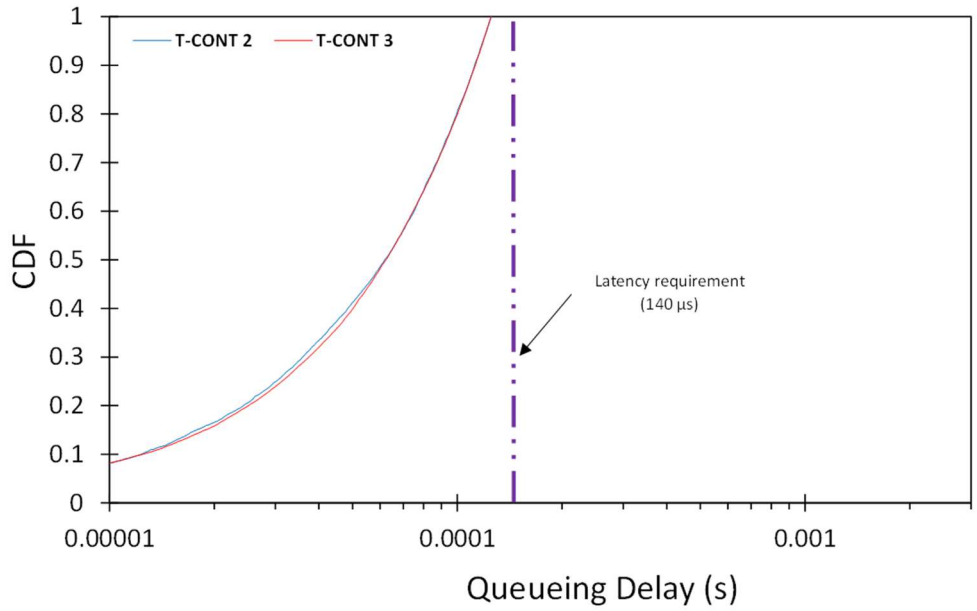


Figure 5.10: CDF of queueing delay at 80% traffic load in scenario 2 (T2 and T3 as fronthaul traffic) for proposed hybrid DBA.

The results obtained for the average ONU upstream delay show a 1% decrease (63.94 µs → 63.10 µs) for T2 and a 4% decrease (66.05 µs → 63.65 µs) for T3. Even though T2 carries a much smaller amount of the fronthaul traffic than T3, the delay values are similar. These results show that the overallocation of bandwidth for T2 and T3 in the IACG DBA can be discarded or eliminated in the proposed DBA while at the same time producing improved outcomes for fronthaul traffic in terms of the average ONU queueing delay and the proportion of frames meeting the 140 µs queueing delay requirement.

5.4.3 Scenario 2 with increased number of ONUs

This scenario is similar to the scenario in section 5.4.2, but with the number of ONUs carrying fronthaul traffic (T2 and T3) increased from 9 to 15 and 1 ONU carrying midhaul

traffic (T4). The bandwidth allocations for T2, T3 and T4 are listed in Table *5.4*. in section 5.4.1.

5.4.4 Results and Discussion

The results obtained are compared with the scenario for the IACG DBA with 9 ONUs carrying fronthaul traffic (T2 and T3) as was described in section 5.4.2. As the number of ONUs for fronthaul traffic is increased using the proposed DBA, the proportion of frames meeting the 140 µs queueing delay requirement remains unchanged at 100%. When this scenario is compared with the IACG DBA, there is no increase (unchanged at 100%) for T2 but a 2.18% increase (97.17% → 100%) for T3 in terms of the proportion of frames meeting the 140 µs queueing delay requirement as shown in Figure *5.11* and Figure *5.12*.

However, the average ONU upstream delay for T2 increases from 63.10 µs with 9 ONUs to 64.02 µs with 12 ONUs, a 1.45% increase. The average ONU upstream delay for T2 (64.02 µs) in the hybrid DBA results in a slight increase over that of the IACG DBA (63.77 µs) and a large decrease for T3 from 69.59 µs to 63.84 µs in the hybrid DBA.

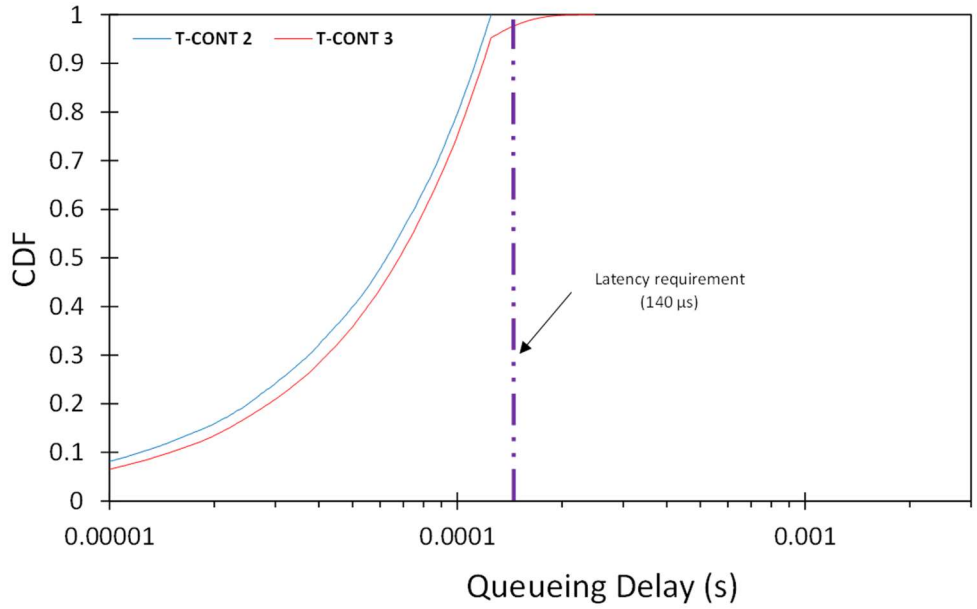


Figure 5.11: CDF of queueing delay at 80% traffic load in scenario 2 (T2 and T3 as fronthaul traffic) with increased ONUs for the IACG DBA.

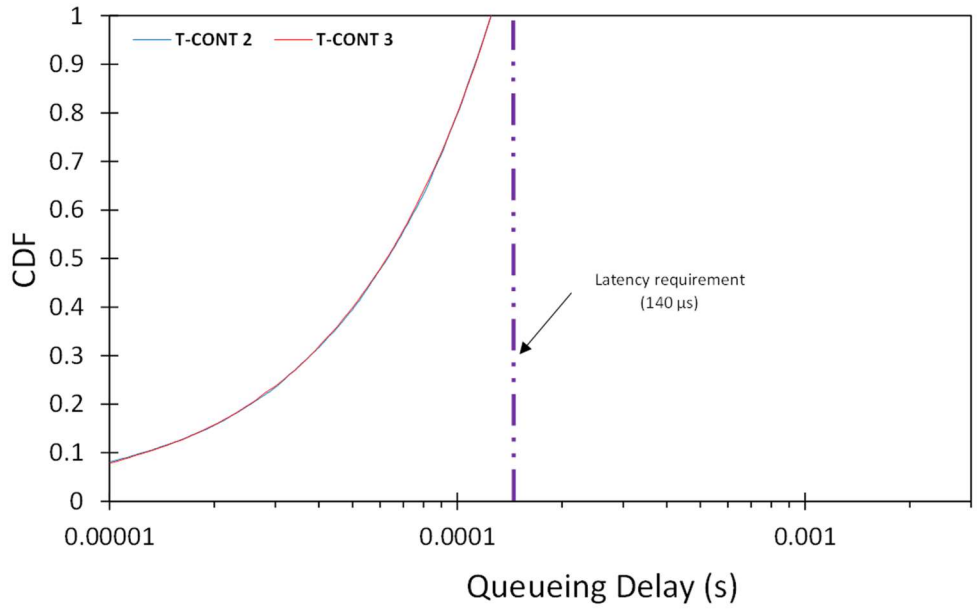


Figure 5.12: CDF of queueing delay at 80% traffic load in scenario 2 (T2 and T3 as fronthaul traffic) with increased ONUs for proposed hybrid DBA.

5.5 Experiments for Scenarios 1 and 2 at all traffic loads

For scenarios 1 and 2 in the proposed hybrid DBA, the same three experiments (experiment 1 – Bandwidth allocation proportional to maximum load, second

experiment 2 – Bandwidth allocation proportional to actual load and experiment 3 – bandwidth overallocation) described in section 4.6 are conducted but with different parameters for the T2, T3 and T4 bandwidth allocation. The fronthaul traffic, i.e., T2 (in scenarios 1 and 2) and T3 (in scenario 2), in the proposed hybrid DBA does not require the use of the IACG DBA process so its bandwidth allocation is zero, as shown in *Table 5.5* and *Table 5.6*. The performance of the hybrid DBA is evaluated in terms of the average ONU upstream delay and the percentage of frame loss meeting the 140 µs queueing delay requirement and compared with the IACG DBA.

Table 5.5: Bandwidth allocation proportional to maximum load in scenario 1 (T2 as fronthaul traffic) for proposed hybrid DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 0 bytes, SI_{max2} = 5	0 Mbit/s
T3	9 – 12	AB_{min3} = 11760 bytes, SI_{max3} = 5	151 Mbit/s
	9 – 12	AB_{sur3} = 390 bytes, SI_{min3} = 5	5 Mbit/s
T4	13 – 15	AB_{sur4} = 9113 bytes, SI_{min4} = 5	116 Mbit/s

Table 5.6: Bandwidth allocation proportional to maximum load in scenario 2 (T2 and T3 as fronthaul traffic) for proposed hybrid DBA.

T-CONT type	**ONU**	**Parameters**	**Bandwidth**
T2	0 – 8	AB_{min2} = 0 bytes, SI_{max2} = 5	0 Mbit/s
T3	0 – 8	AB_{min3} = 0 bytes, SI_{max3} = 5	0 Mbit/s
	0 – 8	AB_{sur3} = 0 bytes, SI_{min3} = 5	0 Mbit/s
T4	9 – 15	AB_{sur4} = 21260 bytes, SI_{min4} = 5	272 Mbit/s

5.5.1 Results and Discussion

Figure 5.13 and *Figure 5.14* shows the results for the percentage of frame loss for bandwidth overallocation in the IACG DBA and proposed hybrid DBA. It can be clearly

seen that the proposed hybrid DBA performs better than the IACG DBA and provides zero frame loss for T2 (in scenarios 1 and 2) and T3 (in scenario 2) at all traffic loads. The reason for this is that in the proposed hybrid DBA, there is no bandwidth waste at the ONU as the OLT knows the amount of bandwidth to allocate to T2 before the arrival of the frames at the ONU buffer in the upstream.

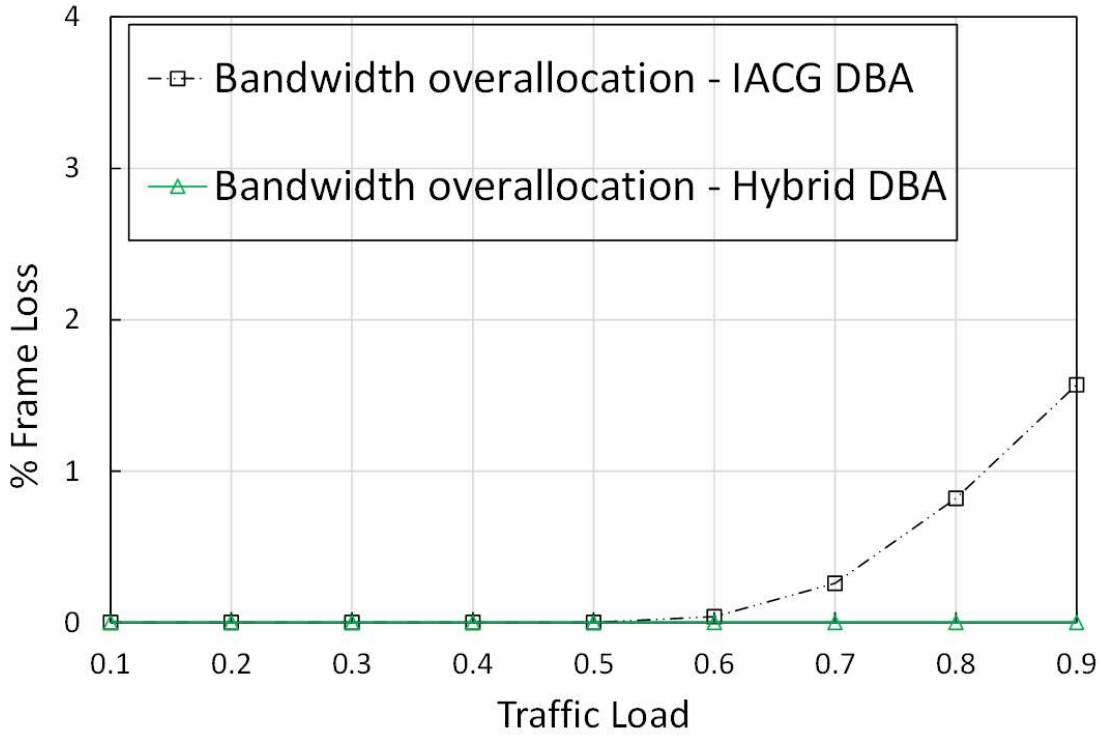


Figure 5.13: Percentage of T2 frame loss vs. traffic load for bandwidth overallocation in scenario 1 (T2 as fronthaul) for IACG DBA and Hybrid DBA.

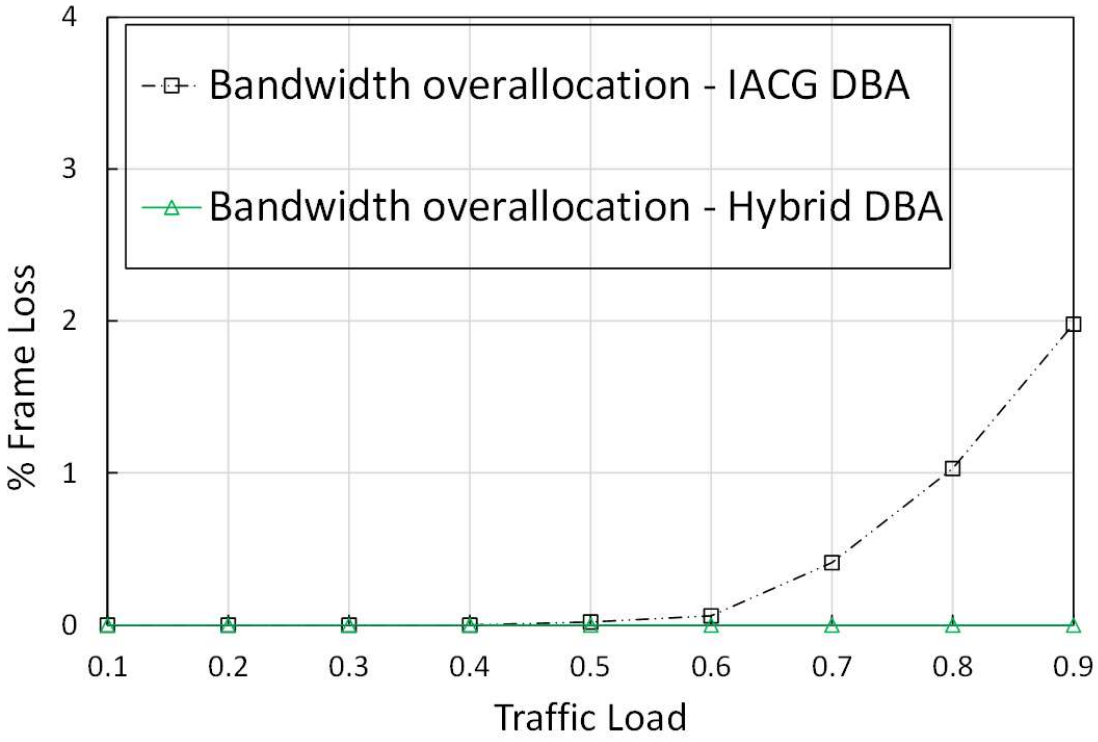


Figure 5.14: Percentage of T3 frame loss vs. traffic load for bandwidth overallocation in scenario 2 (T2 and T3 as fronthaul) for IACG DBA and Hybrid DBA.

In scenario 1 (T2 as fronthaul), the average ONU upstream delay for fronthaul traffic (T2) at 90% traffic load in the IACG DBA is the least (approximately 68 µs) when there is bandwidth overallocation, as shown in *Table 5.7*. For the hybrid DBA, the average ONU upstream delay for T2 is lower and remains the same at approximately 64 µs for all three experiments.

In scenario 2 (T2 and T3 as fronthaul), the average ONU upstream delay at 90% traffic load for T2 (control and signalling – fronthaul traffic) in the IACG DBA remains at approximately 64 µs for the three experiments, while it is approximately 64 µs in the hybrid DBA. As shown in *Table 5.8,* the average ONU upstream delay at 90% traffic load for T3 (user data – fronthaul traffic) for the hybrid DBA remains constant at 64 µs in the

three experiments and this value is lower than the smallest value for the IACG DBA, obtained for the bandwidth overallocation experiment.

Table 5.7: Comparison of T2 average ONU upstream delay in IACG DBA and Hybrid DBA for Scenario 1 (T2 as fronthaul) at 90% traffic load.

Experiment	**IACG DBA**	**Hybrid DBA**
Bandwidth allocation proportional to maximum load	71 µs	64 µs
Bandwidth allocation proportional to actual load	71 µs	64 µs
Bandwidth overallocation	68 µs	64 µs

Table 5.8: Comparison of T3 average ONU upstream delay in IACG DBA and Hybrid DBA for Scenario 2 (T2 and T3 as fronthaul) at 90% traffic load.

Experiment	**IACG DBA**	**Hybrid DBA**
Bandwidth allocation proportional to maximum load	76 µs	64 µs
Bandwidth allocation proportional to actual load	75 µs	64 µs
Bandwidth overallocation	68 µs	64 µs

In order to further test the performance of the proposed hybrid DBA, a fourth experiment (Increased number of ONUs) is conducted, in which the number of ONUs carrying fronthaul traffic for scenarios 1 and 2 is increased from 9 to 15 ONUs with 1 ONU carrying midhaul traffic in scenario 2 or both midhaul and backhaul/fixed access traffic in scenario 1. Scenario 1 uses one ONU, with T3 carrying midhaul traffic and T4 carrying backhaul/fixed access traffic in the ratio 57:43 (T3:T4). This ratio is used to represent the number of ONUs used by T3 and T4 in the third experiment – bandwidth overallocation (4 out of 16 ONUs for T3 = 57% and 3 out of 16 ONUs for T4 = 43%). The number of ONUs used by fronthaul traffic is increased in order to show that overallocation of bandwidth to T2 can be avoided while at the same time obtaining improved results using the proposed hybrid DBA.

In the hybrid DBA, for both scenarios 1 and 2, the frame loss of the fronthaul traffic (T2) remains at 0% while the average ONU upstream delay remains unchanged at approximately 64 µs when using either 9 ONUs or 15 ONUs for transmission. The average ONU upstream delay for midhaul traffic (T3) in scenario 1, which is transmitted using only one ONU, decreases from 66 µs to 63 µs, compared to the case when 4 ONUs was used for transmission while the backhaul traffic (T4), decreases from 66 µs (when using 3 ONUs) to 64 µs (when using 1 ONU). In scenario 2, the average ONU upstream delay for midhaul traffic (T4) remains the same at 66 µs.

Increasing the number of ONUs for fronthaul traffic in the proposed hybrid DBA shows that the performance for frame loss and average ONU upstream delay remains unchanged in both scenarios. However, for non-fronthaul traffic (midhaul, backhaul and fixed access) there is a slight improvement in the average ONU upstream delay.

5.6 Conclusion

In this chapter, a hybrid DBA operating with both the CO-DBA and the IACG DBA for a converged fronthaul was proposed. The hybrid DBA used an XGS-PON system model developed using the OMNeT++ discrete event network simulator. In the proposed hybrid DBA, the CO-DBA transmitted fronthaul traffic while the IACG DBA, using its colorless grant phase, transmitted non-fronthaul traffic (midhaul, backhaul and fixed access traffic) in the network. There is no bandwidth allocation for fronthaul traffic (T2) in both scenarios thereby allowing for much higher loads to be implemented in T2.

The performance of the proposed hybrid DBA was evaluated using the same four experiments in two deployment scenarios as the IACG DBA, described in Chapter 4. The hybrid DBA was shown to give enhanced performance over the IACG DBA for the percentage of frame loss meeting the 140 µs queueing delay requirement for fronthaul traffic and the average ONU upstream delay. It also showed that the midhaul traffic (T3) in scenario 1 relies less on the IACG DBA colorless grant. This is significant in achieving low-latency transmission in the upstream for fronthaul.

When the number of ONUs carrying fronthaul traffic is increased, the IACG DBA shows an increased frame loss and average ONU upstream delay while the hybrid DBA maintained a zero-frame loss for fronthaul traffic. This shows that the overallocation of bandwidth to fronthaul traffic is not required to obtain improved performance for frame loss and average ONU upstream delay when there is an increase in the number of ONUs used to carry fronthaul traffic.

6. Conclusion and Future Work

6.1 Summary

This thesis has presented a detailed description of the research work focused on the use of a passive optical network dynamic bandwidth allocation scheme in meeting the transport requirements in a converged fronthaul network. The transport requirements were based on a threshold of 250 µs for the latency requirement and 99.9% (eMBB use cases) to 99.999% (URLLC use cases) for the reliability requirement.

- The first goal of this research was to investigate the use of the IACG DBA in a converged network by extending and making improvements to the existing model.

- The second goal was to evaluate the IACG DBA using simulations and to analyse its performance. The analysis focused on the average ONU upstream delay and the proportion of frames meeting the fronthaul latency requirement.

- The third goal was to propose a hybrid DBA, operating with the IACG DBA and cooperative DBA, that handles a mix of fronthaul traffic and non-fronthaul traffic in a converged network while satisfying the strict latency requirement for fronthaul traffic.

- The fourth goal was to use the proposed hybrid DBA to effectively reduce the ONU upstream queueing delay and efficiently utilise the bandwidth for fronthaul traffic.

6.2 Conclusions

The following conclusions can be learnt from the main results in this thesis:

- A modified version of the IACG DBA is suitable for meeting the latency requirements in a converged network. This can be achieved by allocating a higher proportion of bandwidth to fronthaul traffic as compared to other types of traffic in the network.

- The average upstream delay of the fronthaul traffic may not reveal whether all the T-CONT frames being sent in the upstream direction are meeting the latency requirement. In order to obtain a realistic view of this, further investigation by examining the delay distribution of the T-CONT frames is necessary.

- An analysis of the delay distribution of T-CONT frames was carried out and it was observed that the IACG DBA is suitable for meeting the latency requirement in a converged network provided an adequate allocation of bandwidth is made available for fronthaul traffic.

- Higher upstream line rates beyond 10 Gbit/s will show a higher proportion of frames meeting the latency requirement and the delays are likely to be lower.

- A higher percentage of frames meet the latency requirement and the average upstream delay decreases as the frame size decreases. This is because smaller frame sizes avoid the need for fragmentation and are likely to be transmitted in one upstream transmission cycle leaving fewer frame fragments waiting for subsequent upstream cycles to be transmitted.

- The proposed hybrid DBA was successfully implemented using the OMNeT++ network simulator. The simulation model was modified for a converged network that transmits a mix of different traffic types and uses the ITU-T cooperative DBA and status reporting DBA methods.

- The proposed hybrid DBA was designed to operate with the cooperative DBA which transmits latency-sensitive fronthaul traffic and IACG DBA, which has a colorless grant phase, and transmits non-fronthaul traffic with less stringent latency requirements.

- The proposed hybrid DBA is shown to provide improved results for the average ONU upstream delay and proportion of frames meeting the latency requirement for fronthaul traffic.

6.3 Future Work

A hybrid DBA and model operating with the IACG DBA and cooperative DBA has been proposed in Chapter 5. The proposed hybrid DBA can be extended in various ways, which is discussed below.

Adjusting the burst allocation frequency is one way to improve the existing model of the hybrid DBA. In the hybrid DBA, the transmission of non-fronthaul traffic is handled by the IACG DBA where a single burst is assigned to each ONU every 125 µs frame i.e., serving an ONU every 125 µs. The burst allocation frequency can be increased to 4 bursts per 125 µs frame (serving an ONU every 31.25 µs) and even 16 bursts per 125 µs frame (serving an ONU every 7.8 µs). Different burst allocation frequencies can be used for different ONUs and for different traffic types in the same PON to investigate its impact on the upstream latency in a converged network.

ONU activation and ranging is a process that is performed by the OLT when a new ONU wants to join a PON and before the ONU is permitted to transmit any upstream data. This process can cause traffic disruptions for ONUs carrying fronthaul traffic in the network. These traffic disruptions can be reduced by modifying the hybrid DBA model to include a second PON system in the network that handles the ONU activation and ranging process and then the newly activated ONU can join the first PON system afterwards. This is a possible way to improve the hybrid DBA model and further reduce upstream latency for fronthaul traffic.

A fixed frame size of 1500 bytes was used in the proposed hybrid DBA and model. An upstream transmission of mixed frame sizes with URLLC (short frame size) and eMBB (long frame size) traffic can be investigated to determine the impact of URLLC traffic on eMBB reliability and data rate.

Different scheduling approaches for transmitting URLLC and eMBB traffic in a converged network based on their respective requirements can be studied and compared. This is achieved by using the different T-CONT types in an ONU for URLLC and eMBB traffic.

A study of having different frames sizes being transmitted using different T-CONTs can be performed. This study can be compared with the existing model that uses a fixed frame size for transmission of T-CONTs. This is to determine the impact on the performance of the network in terms of latency.

Loss of synchronisation is a challenge for mobile traffic that will be transmitted through an asynchronous Ethernet network. Synchronisation over Ethernet can be achieved through the exchange of time-stamped packets using the IEEE 1588 Precision Time Protocol (PTP). The model can be extended by partitioning the time on the network to ensure contention-free transmission of certain packets. A study of the Packet Delay Variation (PDV) can be performed to determine its effect on the network.

The model can be extended for use in high-speed PONs (hsp), e.g., 50 Gbit/s PON, currently undergoing standardisation by the ITU-T. The performance of the fronthaul

network can be studied in terms of the latency, latency variation, bandwidth utilisation efficiency and reliability.

The improvements to the hybrid DBA model stated above for future work are important to further reducing the upstream latency and increasing the proportion of frames that meet the fronthaul latency requirements which is needed for also meeting the reliability requirements for different 5G use cases such as eMBB, URLLC and mMTC.

Bibliography

[1] "Cisco Annual Internet Report (2018–2023)," San Jose, CA, USA, Mar. 2020.

[2] "Ericsson Mobility Report," Stockholm, Sweden, Jun. 2021.

[3] "ETSI GR F5G 001: Fifth Generation Fixed Network (F5G); F5G Generation Definition Release #1," Sophia Antipolis, France, Dec. 2020.

[4] "ITU-R M.2083-0: IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Geneva, Switzerland, Sep. 2015.

[5] "Deliverable D3.6: Final architectural recommendations for FMC networks," COMBO EU Project, Sep. 2016.

[6] "TR-470: 5G Wireless Wireline Convergence Architecture," Broadband Forum, Fremont, CA, USA, Aug. 2020.

[7] "TR-456: AGF Functional Requirements," Broadband Forum, Fremont, CA, USA, Aug. 2020.

[8] "TR-181 Issue 2 Amendment 14: Device Data Model," Broadband Forum, Fremont, CA, USA, Nov. 2020.

[9] "3GPP TS 23.316: Wireless and wireline convergence access support for the 5G System (5GS)," Geneva, Switzerland, Sep. 2021.

[10] D. Goovaerts, "A common core for fiber, 5G could be just around the corner," *Fierce Telecom*, Oct. 05, 2021. https://www.fiercetelecom.com/tech/a-common-core-for-fiber-5g-could-be-just-around-corner (accessed Oct. 15, 2021).

[11] "ITU-T G.984.3: Gigabit-capable passive optical networks (G-PON): Transmission convergence layer specification," Geneva, Switzerland, Jan. 2014.

[12] "ITU-T G.989.3 Amendment 2: 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification," Geneva, Switzerland, Nov. 2018.

[13] R. A. Butt, "Energy and bandwidth efficient medium access layer solution for ITU PONs (PhD thesis)," Universiti Teknologi Malaysia, Johor, Malaysia, 2017.

[14] A. M. Zin, S. M. Idrus, A. Ramli, R. A. Butt, F. M. Atan, and N. A. Ismail, "Performance evaluation of XG-PON with DBA based-watchful sleep mode," in *2018 IEEE 7th International Conference on Photonics (ICP)*, Apr. 2018, pp. 1–3.

[15] R. A. Butt, S. M. Idrus, K. N. Qureshi, P. M. A. Shah, and N. Zulkifli, "An energy efficient cyclic sleep control framework for ITU PONs," *Opt. Switch. Netw.*, vol. 27, pp. 7–17, Jan. 2018.

[16] S. O. Edeagu, R. A. Butt, S. M. Idrus, and N. J. Gomes, "Performance of PON Dynamic Bandwidth Allocation Algorithm for Meeting xHaul Transport Requirements," in *25th International Conference on Optical Network Design and Modelling (ONDM 2021)*, Jun. 2021, pp. 1–6.

[17] J. Kani, S. Kuwano, and J. Terada, "Options for future mobile backhaul and fronthaul," *Opt. Fiber Technol.*, vol. 26, pp. 42–49, Dec. 2015.

[18] "3GPP TS 23.002: Network architecture," Sophia Antipolis, France, Mar. 2021.

[19] "3GPP TS 36.300: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Overall description Stage 2," Sophia Antipolis, France, Mar. 2021.

[20] "3GPP TR 21.915: Release Description; Release 15," Sophia Antipolis, France, Sep. 2019.

[21] “ITU-R M.2150: Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2020 (IMT-2020),” Geneva, Switzerland, Feb. 2021.
[22] “3GPP TR 38.912: Study on New Radio (NR) access technology,” Sophia Antipolis, France, Jul. 2020.
[23] “3GPP TR 38.813: New frequency range for NR (3.3-4.2 GHz),” Sophia Antipolis, France, Mar. 2018.
[24] “3GPP TR 38.814: New frequency range for NR (4.4-5.0 GHz),” Sophia Antipolis, France, Mar. 2018.
[25] “3GPP TR 38.815: New frequency range for NR (24.25-29.5 GHz,” Sophia Antipolis, France, Jun. 2018.
[26] “3GPP TS 22.261: Service requirements for the 5G system,” Sophia Antipolis, France, Apr. 2021.
[27] “C-RAN: The road toward green RAN (Version 2.5),” China Mobile Research Institute, Beijing, China, Oct. 2011.
[28] T. Pfeiffer, “Next generation mobile fronthaul and midhaul architectures [Invited],” *IEEE/OSA J. Opt. Commun. Netw.*, vol. 7, no. 11, pp. B38–B45, Nov. 2015.
[29] A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, “Things you should know about fronthaul,” *J. Light. Technol.*, vol. 33, no. 5, pp. 1077–1083, Mar. 2015.
[30] A. Checko *et al.*, “Cloud RAN for mobile networks - A technology overview,” *IEEE Commun. Surv. Tutorials*, vol. 17, no. 1, pp. 405–426, Sep. 2015.
[31] “5G RAN CU - DU network architecture, transport options and dimensioning, Version 1.0,” NGMN Alliance, Frankfurt, Germany, Apr. 2019.
[32] F.-L. Luo and C. J. Zhang, Eds., *Signal Processing for 5G: Algorithms and Implementations*. Chicester, UK: John Wiley & Sons, 2016.
[33] K. M. S. Huq and J. Rodriguez, Eds., *Backhauling/Fronthauling for Future Wireless Systems*. Chicester, UK: John Wiley & Sons, 2017.
[34] “ITU-T GSTR-TN5G: Transport network support of IMT-2020/5G,” Geneva, Switzerland, Oct. 2018.
[35] A. De La Oliva *et al.*, “Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks,” *IEEE Wirel. Commun.*, vol. 22, no. 5, pp. 32–40, Oct. 2015.
[36] “3GPP TR 38.801: Study on new radio access technology: Radio access architecture and interfaces,” Sophia Antipolis, France, Apr. 2017.
[37] “O-RAN.WG9.XTRP-REQ-v01.00: Xhaul transport requirements,” O-RAN Alliance, Alfer, Germany, Nov. 2020.
[38] “CPRI Specification V7.0: Common Public Radio Interface (CPRI); Interface Specification,” CPRI Cooperation, Oct. 2015.
[39] “Reference Point 3 Specification, Version 4.2,” Open Base Station Architecture Initiative (OBSAI), Mar. 2010.
[40] “ETSI GS ORI 001: Open Radio equipment Interface (ORI); Requirements for Open Radio equipment Interface (ORI) (Release 4),” Sophia Antipolis, France, Oct. 2014.
[41] S. Ahmadi, *5G NR: Architecture, Technology, Implementation and Operation of 3GPP New Radio Standards*. London, UK: Elsevier, 2019.
[42] S. Abrate *et al.*, “An overview of the CPRI specification and its application to C-RAN-based LTE scenarios,” *IEEE Commun. Mag.*, vol. 9, no. 1, pp. 1–5, May 2017.
[43] “Deliverable D2.1: iCIRRUS – intelligent C-RAN architecture,” iCIRRUS EU Project,

Jul. 2015.
[44] D. Samardzija, J. Pastalan, M. MacDonald, S. Walker, and R. Valenzuela, “Compressed transport of baseband signals in radio access networks,” *IEEE Trans. Wirel. Commun.*, vol. 11, no. 9, pp. 3216–3225, Sep. 2012.
[45] B. Guo, W. Cao, A. Tao, and D. Samardzija, “LTE/LTE-A Signal compression on the CPRI interface,” *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 117–133, Sep. 2013.
[46] D. Wubben *et al.*, “Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN,” *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.
[47] U. Dötsch, M. Doll, H. P. Mayer, F. Schaich, J. Segel, and P. Sehier, “Quantitative analysis of split base station processing and determination of advantageous architectures for LTE,” *Bell Labs Tech. J.*, vol. 18, no. 1, pp. 105–128, Jun. 2013.
[48] “Deliverable D3.1: Verification of Ethernet as a transport protocol for fronthaul/midhaul,” iCIRRUS EU Project, Jan. 2016.
[49] “IEEE Std 1914.3-2018: IEEE Standard for radio over ethernet encapsulations and mappings,” New York, New York, USA, Oct. 2018.
[50] “eCPRI Specification V2.0: Common Public Radio Interface: eCPRI interface specification,” CPRI Cooperation, May 2019.
[51] N. J. Gomes, P. Chanclou, P. Turnbull, A. Magee, and V. Jungnickel, “Fronthaul evolution: From CPRI to ethernet,” *Opt. Fiber Technol.*, vol. 26, pp. 50–58, Dec. 2015.
[52] “ITU-R M.2410-0: Minimum requirements related to technical performance for IMT-2020 radio interface(s),” Geneva, Switzerland, Nov. 2017.
[53] H. J. Son and S. M. Shin, “Fronthaul size: Calculation of maximum distance between RRH and BBU,” *Netmanias Tech-Blog*, 2014. https://www.netmanias.com/en/post/blog/6276/c-ran-fronthaul-lte/fronthaul-size-calculation-of-maximum-distance-between-rrh-and-bbu (accessed Jul. 28, 2020).
[54] “Laser wavelength standards,” *National Physical Laboratory, UK*. https://www.npl.co.uk/products-services/dimensional/laser-wavelength-standards.
[55] “ITU-T G.652: Characteristics of a single-mode optical fibre and cable,” Geneva, Switzerland, Nov. 2016.
[56] “ITU-T G.655: Characteristics of a non-zero dispersion-shifted single-mode optical fibre and cable,” Geneva, Switzerland, Nov. 2009.
[57] J. S. Wey, “The outlook for PON standardization: A tutorial,” *J. Light. Technol.*, vol. 38, no. 1, pp. 31–42, Jan. 2020.
[58] “eCPRI Transport Network V1.2: Requirements for the eCPRI transport network,” CPRI Cooperation, Jun. 2018.
[59] “IEEE Std 1914.1-2019: IEEE Standard for packet-based fronthaul transport networks,” New York, New York, USA, Apr. 2020.
[60] “Fronthaul requirements for C-RAN,” NGMN Alliance, Frankfurt, Germany, Mar. 2015.
[61] “SCF159: Small cell virtualization functional splits and use cases,” Small Cell Forum, London, UK, Jan. 2016.
[62] “ITU-T G.Sup66: 5G wireless fronthaul requirements in a passive optical network context,” Geneva, Switzerland, Sep. 2020.

[63] “ITU-T G.8271: Time and phase synchronization aspects of telecommunication networks,” Geneva, Switzerland, Mar. 2020.
[64] D. Hagarty, S. Ajmeri, and A. Tanwar, *Synchronizing 5G Mobile Networks*. San Jose, CA, USA: Cisco Press, 2021.
[65] “ITU-T G.8261: Timing and synchronization aspects in packet networks,” Geneva, Switzerland, Sep. 2019.
[66] D. P. Venmani, O. L. Moult, F. Deletre, Y. Lagadec, and Y. Morlon, “On the role of network synchronization for future cellular networks: an operator’s perspective,” *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 58–64, Sep. 2016.
[67] “3GPP TS 38.104: NR; Base Station (BS) radio transmission and reception,” Sophia Antipolis, France, Apr. 2021.
[68] “3GPP TR 36.819: Coordinated multi-point operation for LTE physical layer aspects (Release 11),” Sophia Antipolis, France, Sep. 2013.
[69] “ITU-T G.8260: Definitions and terminology for synchronization in packet networks,” Geneva, Switzerland, Mar. 2020.
[70] “3GPP TS 38.133: NR; Requirements for support of radio resource management,” Sophia Antipolis, France, Mar. 2021.
[71] “IEEE Std 1588-2019: IEEE Standard for a precision clock synchronization protocol for networked measurement and control systems,” New York, New York, USA, Jun. 2020.
[72] “ITU-T G.8262: Timing characteristics of a synchronous equipment slave clock,” Geneva, Switzerland, Nov. 2018.
[73] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view,” *IEEE Access*, vol. 6, pp. 55765–55779, Sep. 2018.
[74] P. Sehier *et al.*, “Transport evolution for the RAN of the future,” *J. Opt. Commun. Netw.*, vol. 11, no. 4, pp. B97–B108, Apr. 2019.
[75] “3GPP TS 38.401: NG-RAN; Architecture description,” Sophia Antipolis, France, Apr. 2021.
[76] “ITU-T G.8300: Characteristics of transport networks to support IMT-2020/5G,” Geneva, Switzerland, May 2020.
[77] “3GPP TS 38.300: NR; NR and NG-RAN Overall description; Stage-2,” Sophia Antipolis, France, Mar. 2021.
[78] “ITU-T X.200: Information technology - Open Systems Interconnection - Basic Reference Model: The basic model,” Geneva, Switzerland, Jul. 1994.
[79] “3GPP TS 38.201: NR; Physical layer; General description,” 3GPP, Sophia Antipolis, France, Jan. 2020.
[80] “3GPP TS 38.321: NR; Medium Access Control (MAC) protocol specification,” Sophia Antipolis, France, Mar. 2021.
[81] “3GPP TS 38.322: NR; Radio Link Control (RLC) protocol specification,” Sophia Antipolis, France, Jan. 2021.
[82] “3GPP TS 38.323: NR; Packet Data Convergence Protocol (PDCP) specification,” Sophia Antipolis, France, Mar. 2021.
[83] “3GPP TS 38.331: NR; Radio Resource Control (RRC); Protocol specification,” Sophia Antipolis, France, Mar. 2021.
[84] “3GPP TS 24.501: Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3,” Sophia Antipolis, France, Apr. 2021.

[85] “3GPP TS 37.324: Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Service Data Adaptation Protocol (SDAP) specification,” Sophia Antipolis, France, Oct. 2020.
[86] N. J. Gomes *et al.*, “Boosting 5G through ethernet: How evolved fronthaul can take next-generation mobile to the next level,” *IEEE Veh. Technol. Mag.*, vol. 13, no. 1, pp. 74–84, Mar. 2018.
[87] “O-RAN.WG4.CUS.0-v06.00: Control, user and synchronization plane specification,” O-RAN Alliance, Alfer, Germany, Mar. 2021.
[88] “3GPP TR 38.816: Study on Central Unit (CU) - Distributed Unit (DU) lower layer split for NR,” Sophia Antipolis, France, Jan. 2018.
[89] “SCF225: 5G nFAPI Specification,” Small Cell Forum, London, UK, Apr. 2021.
[90] M. Maier and N. Ghazisaidi, *FiWi Access Networks*. Cambridge, UK: Cambridge University Press, 2011.
[91] P. Chanclou, A. Cui, F. Geilhardt, H. Nakamura, and D. Nesset, “Network operator requirements for the next generation of optical access networks,” *IEEE Netw.*, vol. 26, no. 2, pp. 8–14, Mar. 2012.
[92] J. R. Stern, J. W. Ballance, D. W. Faulkner, S. Hornung, D. B. Payne, and K. Oakley, “Passive optical local networks for telephony applications and beyond,” *Electron. Lett.*, vol. 23, no. 24, pp. 1255–1256, Oct. 1987.
[93] J. R. Stern, C. E. Hoppitt, D. B. Payne, and K. Oakley, “TPON - A passive optical network for telephony,” in *14th European Conference on Optical Communication (ECOC 1988)*, Sep. 1988, pp. 203–206.
[94] C. E. Hoppitt and D. E. A. Clarke, “Provision of telephony over passive optical networks,” *Br. Telecom Technol. J.*, vol. 7, no. 2, pp. 100–114, Apr. 1989.
[95] D. W. Faulkner, D. B. Payne, J. R. Stern, and J. W. Ballance, “Optical networks for local loop applications,” *J. Light. Technol.*, vol. 7, no. 11, pp. 1741–1751, Nov. 1989.
[96] M. S. Goodman, H. Kobrinski, and K. W. Loh, “Application of wavelength division multiplexing to communication networks architectures,” in *1986 IEEE International Conference on Communications (ICC)*, Jan. 1986, pp. 931 – 933.
[97] H. Kobrinski *et al.*, “Demonstration of high capacity in the LAMBDANET architecture: A multiwavelength optical network,” *Electron. Lett.*, vol. 23, no. 16, pp. 824–826, Jun. 1987.
[98] S. S. Wagner, H. Kobrinski, T. J. Robe, H. L. Lemberg, and L. S. Smoot, “Experimental demonstration of a passive optical subscriber loop architecture,” *Electron. Lett.*, vol. 24, no. 6, pp. 344–346, Feb. 1988.
[99] N. Ansari and J. Zhang, *Media Access Control and Resource Allocation - For Next Generation Passive Optical Networks*. New York, New York, USA: Springer, 2013.
[100] “Full Service Access Network (FSAN) Group.” https://www.fsan.org/.
[101] “ITU-T G.983.1: Broadband optical access systems based on Passive Optical Networks (PON),” Geneva, Switzerland, Jan. 2005.
[102] “ITU-T G.984.1: Gigabit-capable passive optical networks (GPON): General characteristics,” Geneva, Switzerland, Mar. 2008.
[103] “ITU-T G.987.1: 10-Gigabit-capable passive optical networks (XG-PON): General requirements,” Geneva, Switzerland, Mar. 2016.
[104] “ITU-T G.987.2: 10-Gigabit-capable passive optical networks (XG-PON): Physical media dependent (PMD) layer specification,” Geneva, Switzerland, Feb. 2016.

[105] “ITU-T G.987.3: 10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification,” Geneva, Switzerland, Jan. 2014.
[106] “ITU-T G.987.4: 10-Gigabit-capable passive optical networks (XG-PON): Reach extension,” Geneva, Switzerland, Jun. 2012.
[107] “ITU-T G.9807.1: 10-Gigabit-capable symmetric passive optical network (XGS-PON),” Geneva, Switzerland, Jun. 2016.
[108] “ITU-T G.989: 40-Gigabit-capable passive optical networks (NG-PON2): Definitions, abbreviations and acronyms,” Geneva, Switzerland, Oct. 2015.
[109] “ITU-T G.9804.1: Higher speed passive optical networks - Requirements,” Geneva, Switzerland, Nov. 2019.
[110] “ITU-T G.9804.2: Higher speed passive optical networks: Common Transmission Convergence Layer Specification,” Geneva, Switzerland, Sep. 2021.
[111] “ITU-T G.9804.3: 50-Gigabit-capable passive optical networks (50G-PON) Physical media dependent (PMD) layer specification,” Geneva, Switzerland, Sep. 2021.
[112] “IEEE Std 802.3ah-2004: IEEE Standard for information technology-local and metropolitan area networks-Part 3: CSMA/CD access method and physical layer specifications.,” New York, New York, USA, Sep. 2004.
[113] “IEEE Std 802.3av-2009: IEEE Standard for Information technology-Local and metropolitan area networks-Specific requirements-Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 1: Physical Layer Specifications and Management Par,” New York, New York, USA, Oct. 2009.
[114] “IEEE Std 802.3ca-2020: IEEE Standard for Ethernet Amendment 9: Physical Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks,” New York, New York, USA, Jul. 2020.
[115] C. Desanti, L. Du, J. Guarin, J. Bone, and C. F. Lam, “Super-PON: An evolution for access networks,” *J. Opt. Commun. Netw.*, vol. 12, no. 10, pp. D66–D77, Oct. 2020.
[116] H. Roberts, “Status of ITU-T Q2/15: New higher speed PON projects,” *IEEE Commun. Stand. Mag.*, vol. 4, no. 1, pp. 57–59, Mar. 2020.
[117] Y. Luo, S. Yin, N. Ansari, and T. Wang, “Resource management for broadband access over time-division multiplexed passive optical networks,” *IEEE Netw.*, vol. 21, no. 5, pp. 20–27, Sep. 2007.
[118] J. Zheng and H. T. Mouftah, “Adaptive scheduling algorithms for Ethernet passive optical networks,” *IEE Proc. - Commun.*, vol. 152, no. 5, pp. 643–647, Oct. 2005.
[119] J. Zheng and H. T. Mouftah, “A survey of dynamic bandwidth allocation algorithms for Ethernet Passive Optical Networks,” *Opt. Switch. Netw.*, vol. 6, no. 3, pp. 151–162, Jul. 2009.
[120] O. Haran and A. Sheffer, “The importance of dynamic bandwidth allocation in GPON networks,” PMC-Sierra, Inc., Jan. 2008.
[121] A. Shami, M. Maier, and C. Assi, *Broadband Access Networks: Technologies and Deployments*. Dordrecht, Netherlands: Springer, 2009.
[122] “ITU-T G.989.3 Amendment 3: 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification,” Geneva, Switzerland, Mar. 2020.
[123] M. Radivojević and P. Matavulj, *The Emerging WDM EPON*. Cham, Switzerland: Academic Mind and Springer, 2017.
[124] J. D. Angelopoulos, H.-C. Leligou, T. Argyriou, and S. Zontos, “Prioritized

multiplexing of traffic accessing an FSAN-compliant GPON," in *International Conference on Research in Networking*, May 2004, vol. 3042, pp. 890–901.
[125] J. D. Angelopoulos, H. C. Leligou, T. Argyriou, S. Zontos, E. Ringoot, and T. Van Caenegem, "Efficient transport of packets with QoS in an FSAN-aligned GPON," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 92–98, Feb. 2004.
[126] H.-C. Leligou, C. Linardakis, K. Kanonakis, J. D. Angelopoulos, and T. Orphanoudakis, "Efficient medium arbitration of FSAN-compliant GPONs," *Int. J. Commun. Syst.*, vol. 19, no. 5, pp. 603–617, Jun. 2006.
[127] M.-S. Han, H. Yoo, B.-Y. Yoon, B. Kim, and J.-S. Koh, "Efficient dynamic bandwidth allocation for FSAN-compliant GPON," *J. Opt. Netw.*, vol. 7, no. 8, pp. 783–795, Aug. 2008.
[128] M. S. Han, H. Yoo, and D. S. Lee, "Development of efficient dynamic bandwidth allocation algorithm for XGPON," *ETRI J.*, vol. 35, no. 1, pp. 18–26, Feb. 2013.
[129] V. Sales, J. Segarra, and J. Prat, "An efficient dynamic bandwidth allocation for GPON long-reach extension systems," *Opt. Switch. Netw.*, vol. 14, no. 1, pp. 69–77, Aug. 2014.
[130] J. Segarra, V. Sales, and J. Prat, "GPON scheduling disciplines under multi-service bursty traffic and long-reach approach," in *12th International Conference on Transparent Optical Networks (ICTON 2010)*, Jun. 2010, p. Tu.A1.4.
[131] J. Jiang and J. M. Senior, "A New efficient dynamic MAC protocol for the delivery of multiple services over GPON," *Photonic Netw. Commun.*, vol. 18, no. 2, pp. 227–236, Oct. 2009.
[132] J. Jiang, M. R. Handley, and J. M. Senior, "Dynamic bandwidth assignment MAC protocol for differentiated services over GPON," *Electron. Lett.*, vol. 42, no. 11, pp. 653–655, May 2006.
[133] K. Kanonakis and I. Tomkos, "Offset-based scheduling with flexible intervals for evolving GPON networks," *J. Light. Technol.*, vol. 27, no. 15, pp. 3259–3268, Aug. 2009.
[134] R. A. Butt, M. Waqar Ashraf, M. Faheem, and S. M. Idrus, "A survey of dynamic bandwidth assignment schemes for TDM-based passive optical network," *J. Opt. Commun.*, vol. 41, no. 3, pp. 279–293, Apr. 2020.
[135] M. S. Han, "Dynamic bandwidth allocation with high utilization for XG-PON," in *16th International Conference on Advanced Communication Technology*, Feb. 2014, pp. 994–997.
[136] H. C. Leligou, C. Linardakis, K. Kanonakis, J. D. Angelopoulos, and T. Orphanoudakis, "Efficient medium arbitration of FSAN-compliant GPONs," *Int. J. Commun. Syst.*, vol. 19, no. 5, pp. 603–617, Jun. 2006.
[137] "ITU-T G.983.4: A broadband optical access system with increased service capability using dynamic bandwidth assignment," Geneva, Switzerland, Nov. 2001.
[138] M. P. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. S8-15, Aug. 2004.
[139] R. A. Butt, S. M. Idrus, K. N. Qureshi, N. Zulkifli, and S. H. Mohammad, "Improved dynamic bandwidth allocation algorithm for XGPON," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. 87–97, Jan. 2017.
[140] J. S. Wey, Y. Luo, and T. Pfeiffer, "5G wireless transport in a PON context: An

overview," *IEEE Commun. Stand. Mag.*, vol. 4, no. 1, pp. 50–56, Mar. 2020.
[141] J. S. Wey and J. Zhang, "Passive optical networks for 5G transport: Technology and standards," *J. Light. Technol.*, vol. 37, no. 12, pp. 2830–2837, Jun. 2019.
[142] T. Tashiro *et al.*, "A novel DBA scheme for TDM-PON based mobile fronthaul," in *2014 Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2014, p. Tu3F.3.
[143] S. Hatta, N. Tanaka, and T. Sakamoto, "Feasibility demonstration of low latency DBA Method with high bandwidth-efficiency for TDM-PON," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2017, p. M3I.2.
[144] D. Hisano, H. Uzawa, Y. Nakayama, H. Nakamura, J. Terada, and A. Otaka, "Predictive bandwidth allocation scheme with traffic pattern and fluctuation tracking for TDM-PON-based mobile fronthaul," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2508–2517, Nov. 2018.
[145] D. Hisano and Y. Nakayama, "Two-stage optimization of uplink forwarding order with cooperative DBA to accommodate a TDM-PON-based fronthaul link," *J. Opt. Commun. Netw.*, vol. 12, no. 5, pp. 109–119, May 2020.
[146] J. A. Arokkiam, X. Wu, K. N. Brown, and C. J. Sreenan, "Experimental evaluation of TCP performance over 10Gb/s passive optical networks (XG-PON)," in *33rd IEEE Global Communications Conference (GLOBECOM 2014)*, Dec. 2014, pp. 2223–2228.
[147] P. Alvarez, N. Marchetti, D. Payne, and M. Ruffini, "Backhauling mobile systems with XG-PON using grouped assured bandwidth," in *19th European Conference on Networks and Optical Communications (NOC 2014)*, Jun. 2014, pp. 91–96.
[148] A. M. Mikaeil, W. Hu, T. Ye, and S. B. Hussain, "Performance evaluation of XG-PON based mobile front-haul transport in Cloud-RAN architecture," *J. Opt. Commun. Netw.*, vol. 9, no. 11, p. 984, Nov. 2017.
[149] D. Eugui and J. A. Hernández, "Analysis of a hybrid fixed-elastic DBA with guaranteed fronthaul delay in XG(s)-PONs," *Comput. Networks*, vol. 164, p. 106907, Dec. 2019.
[150] A. Zaouga, A. De Sousa, M. Najjar, and P. P. Monteiro, "Self-adjusting DBA algorithm for Next Generation PONs (NG-PONs) to support 5G fronthaul and data services," *J. Light. Technol.*, vol. 39, no. 7, pp. 1913–1924, Apr. 2021.
[151] S. Das, F. Slyne, A. Kaszubowska, and M. Ruffini, "Virtualized EAST-WEST PON architecture supporting low-latency communication for mobile functional split based on multiaccess edge computing," *J. Opt. Commun. Netw.*, vol. 12, no. 10, pp. D109–D119, Oct. 2020.
[152] "O-RAN.WG4.CTI-TCP.0-v02.00: Cooperative transport interface transport control plane specification," O-RAN Alliance, Alfer, Germany, Mar. 2021.
[153] D. Zhang, D. Liu, X. Wu, and D. Nesset, "Progress of ITU-T higher speed passive optical network (50G-PON) standardization," *J. Opt. Commun. Netw.*, vol. 12, no. 10, p. D99, Oct. 2020.
[154] D. Van Veen and V. Houtsma, "Strategies for economical next-generation 50G and 100G passive optical networks [Invited]," *J. Opt. Commun. Netw.*, vol. 12, no. 1, pp. A95–A103, Jan. 2020.
[155] J. Zhang, Y. Xiao, H. Li, and Y. Ji, "Performance analysis of optical mobile fronthaul for cloud radio access networks," *J. Phys. Conf. Ser.*, vol. 910, no. 1, p. 012053, Oct. 2017.

[156] S. Bidkar, R. Bonk, and T. Pfeiffer, “Low-Latency TDM-PON for 5G Xhaul,” in *22nd International Conference on Transparent Optical Networks (ICTON 2020)*, Jul. 2020, p. Tu.A2.2.
[157] “OMNeT++ discrete event simulator.” https://omnetpp.org/ (accessed Jun. 25, 2020).
[158] “Further study on critical C-RAN technologies v1.0,” NGMN Alliance, Frankfurt, Germany, Mar. 2015.
[159] G. S. Birring, P. Assimakopoulos, and N. J. Gomes, “An ethernet-based fronthaul implementation with MAC/PHY split LTE processing,” in *36th IEEE Global Communications Conference (GLOBECOM 2017)*, Dec. 2017, pp. 1–6.
[160] “ITU-T G.989.3: 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification,” Geneva, Switzerland, Oct. 2015.
[161] G. Kramer, B. Mukherjee, and G. Pesavento, “IPACT: a dynamic protocol for an Ethernet PON (EPON),” *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 74–80, 2002.
[162] H. Uzawa *et al.*, “Practical mobile-DBA scheme considering data arrival period for 5G mobile fronthaul with TDM-PON,” in *43rd European Conference on Optical Communication (ECOC 2017)*, Sep. 2017, p. M.1.B.2.
[163] H. Uzawa *et al.*, “First demonstration of bandwidth allocation scheme for network-slicing-based TDM-PON toward 5G and IoT era,” in *2019 Optical Fiber Communication Conference (OFC)*, Mar. 2019, p. W3J.2.
[164] H. Uzawa *et al.*, “Dynamic bandwidth allocation scheme for network-slicing-based TDM-PON toward the beyond-5G era,” *J. Opt. Commun. Netw.*, vol. 12, no. 2, pp. A135 – A143, Feb. 2020.
[165] S. Zhou, X. Liu, F. Effenberger, and J. Chao, “Mobile-PON: a high-efficiency low-latency mobile fronthaul based on functional split and TDM-PON with a unified scheduler,” in *2017 Optical Fiber Communication Conference (OFC)*, Mar. 2017, p. Th3A.3.
[166] S. Zhou, X. Liu, F. Effenberger, and J. Chao, “Low-latency high-efficiency mobile fronthaul with TDM-PON (Mobile-PON),” *J. Opt. Commun. Netw.*, vol. 10, no. 1, p. A20, Jan. 2018.
[167] H. Nomura, H. Ujikawa, H. Uzawa, H. Nakamura, and J. Terada, “Novel DBA scheme integrated with SR- and CO-DBA for multi-service accommodation toward 5G beyond,” in *45th European Conference on Optical Communication (ECOC 2019)*, Sep. 2019, pp. 1–4.