Strategies for the Efficient Management of the Capacity in Fixed-Mobile Converged Networks

Doctoral Dissertation of:
Anna Buttaboni

Supervisor:
Prof. Massimo Tornatore

Tutor:
Prof. Antonio Capone

The Chair of the Doctoral Program:
Prof. Carlo Fiorini

2012/2015 – 28
Abstract

Nowadays customers can access telecom services via fixed-line networks or via mobile networks. Fixed-line broadband networks in Europe are currently dominated by ADSL technologies which provide up to 16 Mbit/s. Other solutions like Fiber to the Curb (FTTC) with VDSL are widely used with access speeds up to 50 Mbit/s. The technologies which will deliver fiber access directly to the home, referred to as Fiber to the Home (FTTH) solutions, are the next step and first deployments have been started. If the implementation of FTTH is based on a Passive Optical Network (PON) instead of a point-to-point solution, it is possible to reduce the network cost by eliminating the power supply through the installation of passive network elements, and by sharing a significant portion of the network cost among multiple users, while still enabling data rates of hundreds of Mbit/s. For the mobile area, different network technologies are available and widely used. 2G (e.g., GSM, GPRS) and 3G (e.g., UMTS, HSPA) networks have been already installed and in some countries the deployment of the 4G LTE technology (with data rates up to 100 Mbit/s) has been started. Future step is the LTE advanced (LTE-A) technology which can provide data rates up to 1 Gbit/s. With regard to wireless technologies, systems like Wi-Fi (standard IEEE 802.11) or WiMax (standard IEEE 802.16) are also widely used. So far fixed and mobile access networks have been optimised and evolved independently, with some contradicting trends. In fact there is a tendency to centralize fixed networks and to decentralize mobile
networks. A certain degree of convergence among the two network domains, typically referred as Fixed Mobile Convergence (FMC), has only been achieved at the service level with the introduction of all IP services (e.g., a practical case of FMC at service level can be found in smartphones and tablets which can access the same services through Wi-Fi and/or the cellular network). It is widely agreed that the development of a single convergent infrastructure for fixed and mobile networks would enable relevant savings in terms of Capex and Opex and would provide converged services to customers at reasonable costs in the years. The development of this FMC access network is driven by the requirement to combine optimal seamless quality of experience for end-users together with an optimised network infrastructure. Moreover, another motivation to merge together in a single and optimized structure both fixed and mobile traffic is related to the energy consumption of the current access network. In fact the access is the part of the network which is consuming the highest amount of energy. The convergence of fixed and mobile access networks in a single structure can help saving energy. In this sense, we can classify fixed mobile convergence referring to two aspects: functional convergence and the structural convergence. The functional convergence refers to the convergence of fixed and mobile network functions, while the structural convergence is the convergence of fixed and mobile infrastructures and equipment. The integration of functionalities and equipment are expected to enable relevant energy saving, e.g., lowering the number of nodes of the access network. One of the most promising network solutions to develop such FMC access networks are Next-Generation Passive Optical Network (NG-PONs). In particular, Long-Reach PONs which use both time-domain and wavelength-domain multiplexing, referred as LR WDM/TDM PON, are a suitable candidate to support a large number of different services (with different QoS requirements) originating from both fixed and mobile users.

The aim of this work is to propose, define and technically assess methods to efficiently manage the convergence of Fixed and Mobile traffics over the same infrastructure.

In particular, we first proposed Dynamic Bandwidth and Wavelength Allocation (DBWA) mechanisms for Long-Reach WDM/TDM Passive Optical Networks (LR WDM/TDM PONs). These proposed methods have shown to have a better performance, in terms of average packet delay, compared with other existent DBWAs designed for
LR WDM/TDM PONs. Moreover, in order to fairly allocate the bandwidth among all the users which transmit over multiple wavelengths in LR WDM/TDM PONs, we proposed a new equation. Such equation, given the length of the polling cycle time (in seconds), provides the maximum amount of bytes that each user can transmit in each polling cycle, according to the capacity assigned to each user. Then, we also studied the effect over the average packet delay of changing the length of the polling cycle. Regarding LR WDM/TDM PONs, we evaluated the multiplexing gain introduced when different number of wavelengths are used. We compared the performance of the different scenarios (with different number of wavelengths) in terms of average packet delay. Such results are also compared with the case where a single wavelength is used to transmit (LR TDM PON). From these results we can provide some preliminary consideration regarding the design of LR WDM/TDM PON which can be used for the FMC backhauling.

Then, we studied Dynamic Bandwidth and Wavelength Allocation (DBWA) mechanisms for LR WDM/TDM PONs used in a network scenario where the transmission technologies at the ONUs are all tunable lasers (i.e., uniform transmission technologies scenario). We started by evaluating DBWAs in this uniform transmission technologies scenario to consider which is the impact of the tuning time (TT) of the tunable lasers on the network performance in terms of delay. For this reason, we modified existing DBWAs in order to make them able to consider the TT of Tunable lasers. This modification consists in adding a delay to the transmissions allocated to give enough time to the ONUs to retune their lasers. Then, we evaluated the performance of these modified DBWAs. After this preliminary evaluation, we proposed a new TT-aware DBWA (called TAWA). The fact that a DBWA is TT-aware means that the scheduling is performed considering that the laser needs a certain amount of time to retune to another wavelength (i.e., as a $TT > 0$). Then, since the aim of this work is to study FMC network solutions to efficiently manage the different types of traffic present in such types of networks, we also studied DBWAs for LR WDM/TDM PONs in scenarios where different transmission technologies with different characteristics (i.e., different TTs) are used at the ONUs, considering that different services might be supported by different technologies (e.g., Tunable lasers, array of fixed tuned lasers, fixed tuned lasers, · · ·). Also in this case, we first evaluated the performance of some existing DBWAs, that we modified to make
them considering that different transmission technologies have different TTs, i.e., the modified DBWAS have different allocation policies are adopted for transmissions originated from ONUs with different technologies. In particular, in our work we considered that, in a LR WDM/TDM PON, three types of transmission technologies can coexist: i) Single fixed tuned lasers which cannot retune; ii) Array of fixed tuned lasers which can immediately tune their laser to another wavelength (TT = 0); iii) Tunable lasers which are characterized by a given TT (TT > 0). Then, to understand the importance of having a DBWA which is aware of the exact values of TT and is able to add different delays to different transmissions, we designed a novel version of the existing and already evaluated DBWAs where the OLT does not know the exact value of the TT of each ONUs, but it only knows if a particular transceiver is able to retune or not. These new versions of DBWAs can only exploit a simplified control information, therefore they only know if a transceiver is able to retune or not, but they do not know which is the value of TT that each transceiver needs to retune. For this reason, these new versions of DBWAs add a delay to all the transmissions of the ONUs which have to retune (i.e., Array of fixed tuned lasers and tunable lasers) considering a maximum value of TT (the same for all the ONUs). In this first part of the work, we assumed that the arrays of fixed tuned lasers are able to tune to every wavelength of the network. The number of wavelengths over which such arrays can transmit corresponds to the number of fixed tuned lasers installed in the array. However, we know that to have a cheap device the number of wavelengths where an array of fixed tuned lasers can transmit (i.e., number of lasers of the array) has to be limited. For this reason, we evaluated the performance of the previously studied DBWAs when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths. The results of this study show that using an array of lasers with a limited number of lasers provide an average packet delay which is only slightly higher (in the order of tenth of microseconds) than in the case where the arrays of tunable lasers have full tunability (i.e., they can transmit over every wavelength of the network). Moreover, we proposed a new DBWA that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength they have to retune to (i.e., higher distance between wavelengths requires higher TTs). To the best of our knowledge this is the first time that such a DBWA is proposed and analyzed.
Another part of the work concerns more the architecture of the FMC networks. Indeed, in this part of the research we considered a ring topology of an access optical network and we designed an ILP which is able to allocate the resources (i.e., wavelengths, or portion of wavelengths) to the ONUs according to their required amount of bandwidth (in bit/s) while, at the same time, minimize the total number of active wavelengths (i.e., minimize the power consumption at the OLT, minimizing the number of line cards used). Moreover, in our scenario the ONUs connected to the ring network provide traffic belonging to different service (e.g., fixed traffic, 3G, LTE, · · · ) with different requirements in terms of bandwidth. Since different types of traffic are managed through different network protocols and different network nodes, we add a constraint to our ILP that groups the transmissions of the ONUs providing the same type of traffic over the same wavelength (or the same set of wavelengths). We argue and we aim at demonstrating that physically separating the different types of traffic over different wavelengths provides a benefit in terms cost of traffic management (e.g., processing delay in the nodes). With this constraint, the outcome of the ILP is a set of Virtual-PONs (VPONs). VPONs are PONs where the ONUs belonging to a particular VPON are not necessarily connected to the same passive remote node. The ONUs belonging to a VPON can be connected to different branches of a PON or LR-PON (e.g., connected to different passive remote nodes of a ring optical access network), and their transmissions are physically separated from the transmissions of the other ONUs by using a dedicated wavelength (or set of wavelengths). The allocation of the transmissions of a VPON is managed by a single OLT. For this study, an ILP has been developed and solved with CPLEX.

All the results provided during these studies have been obtained through simulations.

The results of this research will be of impact for the future generations of high-speed broadband and mobile network infrastructure. An FMC access network will lead to significant network cost and energy-consumption reductions which will be key to address the profound transformations needed to face data traffic explosion in the medium to long term.
# Contents

1 Introduction
   1.1 Fixed Mobile Converged Networks .................................. 2
   1.2 PONs as a Backhauling infrastructure for FMC ...................... 4
   1.3 Acknowledgements ................................................... 6
   1.4 Work Organization ................................................... 6

2 Long-Reach TWDM Passive Optical Network...................... 9
   2.1 Passive Optical Network ............................................ 10
      2.1.1 First generation PON: EPON and GPON ...................... 16
      2.1.2 NG-PON1 ...................................................... 16
      2.1.3 NG-PON2 ...................................................... 17
   2.2 Long-Reach Passive Optical Network ................................. 19
   2.3 Long-Reach TWDM Passive Optical Network ......................... 21
   2.4 Research Challenges ................................................ 23
      2.4.1 Network Components ........................................... 23
      2.4.2 Dynamic Bandwidth Allocation Algorithms .................. 24
      2.4.3 Network Protection ............................................ 25

3 Dynamic Bandwidth and Wavelength Allocation Schemes.......... 27
   3.1 Multi-Point Control Protocol ..................................... 28
   3.2 Dynamic Bandwidth Allocation Algorithm ......................... 28
      3.2.1 Interleaved Polling with Adaptive Cycle Time .......... 31
      3.2.2 Multithread Algorithm ....................................... 33
3.3 Dynamic Bandwidth and Wavelength Allocation Algorithms ................. 34
  3.3.1 Earliest Finish Time Algorithm ........................................... 35
  3.3.2 Earliest Finish Time with Void Filling Algorithm .................. 36
3.4 DESL: Discrete Event Simulation Library ...................................... 38

4 Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs ................................................. 43
  4.1 Related works ......................................................................... 45
  4.2 Multi-thread implementation in LR WDM/TDM PONs .................... 46
  4.3 EFT-Partial-VF Multi-threaded ................................................. 47
  4.4 Definition of the Cycle Time ...................................................... 52
  4.5 Illustrative numerical results ..................................................... 54
    4.5.1 Generic Simulation Framework ............................................ 54
    4.5.2 Simulation Framework varying Network Parameters ............ 54
    4.5.3 Numerical Results and Discussion ....................................... 55
  4.6 Multiplexing Gain Evaluation .................................................. 60
  4.7 Novel LR WDM/TDM PON Architecture and Solutions for FMC Backhauling ......................................................... 62
  4.8 Conclusion ............................................................................... 63

5 Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs ................................ 67
  5.1 Contributions ........................................................................... 68
  5.2 Related works ......................................................................... 69
  5.3 TT-aware DBWAs with coexistence of transceivers ....................... 70
    5.3.1 Discovering Methods of the TT values .................................. 73
    5.3.2 Array of Fixed tuned lasers with limited tunability ............... 73
  5.4 TT-aware DBWAs with Wavelength Dependent Tuning Times .......... 73
  5.5 Results and discussion .............................................................. 75
    5.5.1 Simulative Scenario ............................................................. 75
    5.5.2 TT-aware DBWAs vs Simple DBWAs .................................... 77
    5.5.3 Average delay evaluation: Array of fixed tuned lasers with limited tunability ................................................................. 80
  5.6 Conclusion ............................................................................... 83

VIII
## Contents

6 Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

6.1 Related Works ........................................ 88
6.2 VPON assignment in FMC networks: Problem Definition and Modeling ................. 89
  6.2.1 VPON Assignment in FMC networks: the ILP Model .................................. 90
6.3 Illustrative Numerical Results ........................................ 93
  6.3.1 Network Scenarios .................................. 93
  6.3.2 Numerical Results and Discussion ............................... 95
6.4 Conclusion ........................................ 100

7 Multi-Operator Network Sharing ........................................ 105

7.1 Related Works ........................................ 105
7.2 Scenario ........................................ 106
7.3 Game Theoretic model for Multi-operator Network Sharing ................................ 108
7.4 Game theoretic Approach with Partial Accept ........................................ 113
7.5 Illustrative Numerical Results ........................................ 115
  7.5.1 Network scenario .................................. 115
  7.5.2 Numerical Results and Discussion ........................................ 116
7.6 Conclusion ........................................ 118

8 Conclusion ........................................ 123

8.1 Future Works ........................................ 126
The access-aggregation segment of today’s telecom networks is being pressured by the continuous bandwidth growth of fixed and mobile network services. Telecom operators are facing the big challenge to provide the best broadband services at the lowest cost. The operators need to evolve their networks in order to be more efficient and flexible, providing the best Quality of Service (QoS) or Quality of Experience (QoE) to their customers with the minimum cost. Novel solutions to perform cost-efficient backhaul of different services over this network segment are currently being investigated [1]. Among them, the concept of Fixed-Mobile Converged (FMC) networks is gaining attention [2]. An FMC network is a convergent access-aggregation network that provides concurrent backhauling of mobile and fixed traffic over a single infrastructure. The service provided over an FMC network have to use wireline and wireless/cellular networks in a seamless way, unifying these two technologies in a single access network for all kind of services. These two networks (i.e., fixed and mobile) have grown independently of each other and currently they are based on different technologies and protocols, and work in different ways. During the last years a certain degree of convergence
Chapter 1. Introduction

is appearing with the introduction of all IP services and IP Multimedia Subsystem (IMS) platforms [3]. The convergence of fixed and mobile network will generate savings in Capital Expenditure (CAPEX), in Operational Expenditure (OPEX), and in energy consumption. The network CAPEX can be reduced since a single infrastructure can be deployed to accommodate the traffic of both fixed and mobile users. The OPEX of the network can be lowered because, using a single unified network leads to a more efficient network utilization. Finally, the energy consumption in FMC networks is reduced because, having a single network infrastructure the number of active nodes is reduced compared to the case of two separated access networks.

1.1 Fixed Mobile Converged Networks

Currently FMC is mainly based on the service level and operators have started to build a converged service control layer, but both fixed and mobile networks continue to work independently and they are composed of the same traditional network segments: access, aggregation and core in the case of fixed networks and radio access network (RAN) and mobile backbone network in the case of mobile segment. Fixed networks are currently based on very different access technologies such as xDSL, CATV and FTTx that can span over distances of a few kilometers, combined with aggregation networks that typically cover distances of tens or hundreds of kilometers using SONET/SDH, Ethernet and MPLS. Also mobile networks are based on different radio technologies such as LTE and LTE-Advanced that provide high bit rates such as 100 Mbit/s or 1 Gbit/s respectively. Moreover, mobile networks require a fixed connection to connect the RAN to the mobile backbone, this implies that a kind of convergence between fixed and mobile networks has taken place through a backhaul using point to point connections such as PDH, SDH and GbE.

Several traffic and economical drivers exist for FMC network deployment. First, the need for more capacity in the access network due to i) increasing number of subscribers, ii) users mobility and their need to access Internet everytime and everywhere, and iii) growing number of connected devices. Another important driver for the development of fixed and mobile converged networks is the increasing number of Over-The-Top (OTT) services that need high amount of bandwidth (e.g., online gaming, IP telephony, ...). The reference FMC network infrastructure is shown in Fig.
with respect to the different network segments of current fixed and mobile networks (i.e., customer premises, access network, aggregation network, and core network).

Figure 1.1: Reference topology of the FMC network architecture.

Basically, in FMC networks we can refer to two different types of convergence: i) Functional convergence, and ii) Structural convergence. Functional convergence is defined as the convergence of fixed and mobile network functions at Layer 3 (the network layer, e.g., IP protocol) or below. It will primarily impact the control plane of future networks through unified control mechanisms of fixed and mobile networks. Moreover, it will also impact the data plane through an optimization of protocol stack and a better distribution of data flows in the converged network. The functional convergence will improve QoS, QoE and flexibility for the end user. Functional convergence will benefit the customer by making the service independent from the access technology and the device used to connect to the network, by introducing natively converged technologies and protocols in the network domain, and not by adding other service control layers. Functional convergence should also give the customer the best access to the network for a particular service, in a transparent manner. For example, functional convergence would allow a WiFi service provider to implement fast handover for its customers between the WiFi plat-
form and the 3GPP network (of both same or different operator) thanks to unified control of heterogeneous networks and technologies as well as homogenization at functional level of authentication and subscriber management. Since the wireless portion of FMC networks can be composed by access nodes using different technologies (e.g., WiFi, 3G or LTE/LTE-A) and different cell sizes (e.g., macrocells, microcells, picocells, femtocells), functional convergence will be needed to provide interconnection of all the components of these heterogeneous networks and technologies. Structural convergence is defined as the mutualization of fixed and mobile access/aggregation network infrastructures and hardware (e.g. cable plants, cabinets, sites, equipment, buildings). Structural convergence will further optimize the use of the most expensive part of fixed and mobile networks, and drastically decrease cost and energy consumption. Moreover, structural convergence involves sharing the infrastructure and equipment of fixed and mobile networks.

1.2 PONs as a Backhauling infrastructure for FMC

A promising next-generation access technology to be used as a backhauling infrastructure for FMC networks is Passive Optical Networks (PONs) [4]. A PON is composed of an Optical Line Termination (OLT) located at the Central Office (CO) and several Optical Network Units (ONUs). A common feeder fiber connects the OLT to a passive optical splitter/combiner through which the signal is conveyed to each ONU in the network. The ONUs share the upstream channel in the time domain, with Time Division Multiple Access techniques (TDMA) whereas the downstream channel is broadcast to all ONUs. This basic implementation of PON can be also referred to as TDM-PON.

More recently, several solutions using WDM (called WDM-PONs) have been presented [5-7]. A simple version of WDM-PON, where a pair of wavelengths is assigned to each ONU, can provide a large amount of bandwidth to final users. However, this approach can lead to inefficiencies in the use of the available capacity because ONUs do not always transmit or use the entire channel capacity. This is why it is important to combine WDM with TDM in order to increase the channel utilization by sharing wavelengths among more than one ONU.

While typically Passive Optical Networks (PONs) support signal
reaches in the order of 10-20km, several studies have also investigated the concept of Long Reach PON [8], where the network span can be prolonged up to about 80-100km. This type of network is expected to significantly simplify the metro-access network architecture by reducing the number of network elements.

The evolution from LR TDM PON towards LR TDM/WDM PON is a quite recent research topic motivated by the need of such network architectures to support higher capacity. Recently, the Full Service Access Network (FSAN) forum selected as a primary solution for Next-Generation PON Stage 2 (NG-PON2) [9] a flavor of LR-PONs where the transmissions in the upstream channel are allocated through hybrid TDM/WDM access technique. These networks are defined as TWDM PONs.

Since in PONs the users share the same transmission medium it is necessary to have a mechanism that manages the transmissions of the users. In PONs, in order to manage the scheduling of the transmissions of the users, is used a Dynamic Bandwidth Allocation (DBA) algorithm. A DBA is a technique by which traffic bandwidth in a shared telecommunications medium can be allocated on demand and fairly among different users. Within this technique, the sharing of a link adapts to the instantaneous traffic demands of the ONUs connected to the network. In particular, a DBA schedules the transmissions of the ONUs in time, i.e., it assigns a point in time in which a ONU can start its transmission without colliding with any other transmission. A DBA takes advantage of the fact that i) the ONUs connected to the network are not always transmitting data, and ii) the traffic of the ONUs mainly occurs in bursts. The DBA dynamically allocates the amount of bandwidth for each ONU transmission according to the amount of traffic that an ONU need to transmit in that particular moment.

In TWDM PONs, the transmissions of the ONUs have to be allocated both in time and over a certain wavelength in order to avoid collisions. For these types of networks, in order to manage the scheduling of the transmissions of the users, are used Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithms. Similarly to a DBA, a Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithm dynamically assigns the amount of bandwidth that each ONU can transmit and also the wavelength over which each ONU will transmit.

The work presented in this thesis have been published in the
Chapter 1. Introduction

following papers:


1.3 Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7/2013-2015 under grant agreement no. 317762 COMBO project.

1.4 Work Organization

The rest of the work is organized as follows: In chapter 2 we will explain in details the evolution and the functionalities of the LR TWDM PONs, presenting also the state-of-the-art of such technology. Chapter 3 introduces the functionalities of DBA and DBWAs scheme such as the signalling protocol used and the evolution from DBAs designed for PONs to DBWAs for LR TWDM PONs. In chapter 4, we present our proposed DBWAs for LR TWDM PONs with
homogeneous transmission technologies. In chapter 5, we introduce our proposed DBWAs for LR TWDM PONs with heterogeneous transmission technologies. In chapter 6, we present our proposed method to assign different types of traffic to different Virtual PON (VPONs). Then, in chapter 7, we consider multi-operator network sharing strategies in FMC networks. In this chapter we introduce our proposed game theoretic model for the collaboration among different operators, and we compare this strategy with other benchmarking strategies. Finally in chapter 8, we draw the conclusion on the performance of the different scheduling disciplines, and also possible future works are discussed.
Chapter 2

Long-Reach TWDM Passive Optical Network

In this chapter we will explain in details the evolution of the Passive Optical Networks (PONs) towards the Long Reach TWDM PONs (LR TWDM PONs), and the features of these types of network. In particular, in Section 2.1 we introduce the concept of Passive Optical Network (PON), we explain the main functionalities of these networks, and we describe their evolution. In Section 2.1.1 we report the features of the two main standards PON systems. Section 2.1.2 introduce the first stage of PON evolution: the NG-PON1 system. In Section 2.1.3 we describe the main features and requirements of the second stage of PON evolution: the NG-PON2. Then, in Section 2.2 the reach extension of PONs, the Long Reach Passive Optical Network (LR-PON), is explained. Section 2.3 describes Long Reach TWDM PONs which are a particular flavor of the LR-PONs. Finally, 2.4 summarize the research challenges introduced by LR-PONs.
2.1 Passive Optical Network

Passive Optical Networks [10] are point-to-multipoint access network solutions based on passive components such as optical power splitters. Today’s PONs typically support 32-64 subscribers minimizing fiber deployments and providing a subsequent cost reduction.

Figure 2.1: TDM PON configuration.

A typical PON consists of an Optical Line Termination (OLT) located at the Central Office (CO) and several Optical Network Units (ONUs). A common feeder fiber that connects the OLT to a passive optical splitter/combiner, located at the Remote Node (RN), through which the signal is conveyed to each ONU in the network. The ONUs share the upstream channel in the time domain, with Time Division Multiple Access techniques (TDMA) whereas the downstream channel is broadcasted to all ONUs. This basic implementation, shown in Fig. 2.1 of PON can be also referred to as TDM PON. In a TDM PON the OLT time multiplexes packets and transmits them to the ONUs. In the downstream direction the signal is split among all ONUs through a splitter. Each ONU, which has a unique identifier, reads the destination address of each packet and selects the ones that match its address. In downstream the synchronization among transmissions
for different ONUs is straightforward because it is done directly by the OLT, which is the only transmitter in this direction. In the upstream direction, each ONU is assigned a time slot in which packets are transmitted to the OLT through the splitter, which now acts like a combiner, and where optical packets are passively multiplexed. Using more elaborate protocols the time slot duration can be dynamically adjusted according to the need of each ONU allowing for dynamic bandwidth allocation (DBA). DBAs can then be used to better utilize the capacity, which is a limited resource.

A drawback of TDM PONs consists in a lack of security in the downstream. Since the downstream transmission is broadcast, every user can get all the information. Eavesdropping is possible by decoding all the information which is not intended to a particular user. Though, there are advanced encryption techniques to prevent eavesdropping, nonetheless TDM PONs suffer from security. Moreover, since there is an enormous increase of bandwidth for services like IPTV, Video on Demand, TDM PON could not provide such high bandwidths because of the electronic components present on either side, and it is not highly scalable.

In the last years, in order to increase the split ratio, the number of network users, and the offered bandwidth, various novel PON solutions using Wavelength Division Multiplexing (called WDM PON) have been introduced [15]. A simple version of WDM PON, shown in Fig. 2.2, where a different pair of wavelengths, one for upstream and one for downstream transmission, is assigned to each ONU, can provide a large amount of bandwidth to each user. In WDM PONs the ONUs can be classified as colorless or colored. In the former one, the downstream carrier can be reused or remodulated at the ONUs such that no light source is required at the ONU [16]. These network components are called Reflective Semiconductor Optical Amplifiers (RSOAs). Another way to implement a colorless ONU is through installing a tunable laser (TL) that can be dynamically tuned to a particular wavelength according to the needs of the ONU. A colored ONU, instead, features a transmitter/receiver (or an array of fixed tuned laser) set to a fixed wavelength. In order to decrease the costs of operation, administration, and maintenance (OA&M) functions, in a WDM PON it is highly desirable to use colorless ONUs. Moreover, with these colorless transceivers, mass production becomes possible due to the high number of these components needed into a PON and production costs can be reduced.
In a WDM PON, the splitter, which was usually used in TDM PON, is replaced by a wavelength selective filter implemented with an arrayed waveguide grating (AWG) [17]. Besides other more complex functionalities like the wavelength routing function, this component allows to multiplex or demultiplex different wavelengths.

Positive features of the WDM PONs consist in the fact that, unlike TDM PONs, WDM PON provides more security, as each user has a particular wavelength where the data is sent. So, in WDM PONs the security is enhanced with respect to TDM PONs. Moreover, WDM PONs do not have issues with scalability. Indeed, in WDM PONs it is possible to add new wavelength for new connected ONUs when it is needed.

However, the use of a pure WDM PON can lead to inefficiencies in the use of the available capacity because ONUs do not always use the entire channel capacity of the wavelength. For this reason it is important to combine WDM with TDM in the access system in order to increase the channel utilization by sharing multiple channels among several ONUs. Another reason for hybrid TDM/WDM architectures is the need to have a system which is compatible with existing TDM PON architectures. The integration of TDM and WDM can ensure
the flexibility in bandwidth allocation, and a smooth transition from already placed TDM PONs to high bandwidth WDM PONs.

PON systems which combine TDM with WDM techniques are called TDM/WDM PONs. They are a compromise between the two multiplex schemes, which exploits the advantages of both techniques and combines network features of both TDM and WDM PONs.

Note that, in such networks, splitters and AWG components can be installed together. In this way it is possible to achieve a high split ratio since the signal is divided both by the power splitter and by the AWG. Figure 2.3 and Fig. 2.4 show two different ways to combine splitters and AWGs into a TDM/WDM PON. In these examples splitters have a split-ratio equal to 2. Moreover, this network configuration allows all the ONUs to share the physical medium in both time and wavelength domain.

Unlikely in WDM PONs, where each ONU sends and receives data over a different couple of wavelengths, in hybrid TDM/WDM PONs, ONUs can transmit over different wavelengths. For this reason, ONUs must be equipped with transceivers (TL or RSOA) which allow them to switch from a wavelength to another according to the assignment devised by the OLT.

TDM/WDM PONs provide a large amount of available bandwidth,
by using different wavelengths and dynamically sharing them, according to the needs of each user, using time-slot allocation. Moreover, as the bandwidth can be shared in both time and wavelength domains, the coordination among different transmissions is even more complicated for these network configurations. For this reason it is necessary to implement efficient schemes for Dynamic Bandwidth and Wavelength Allocation (DBWA).

However, the cost reduction provided by a PON might not be enough for future telecom networks. In fact, as an example, research shows that the realization of this kind of fiber access, throughout the United Kingdom, would cost around £15 billion \[18\]. For this reason, an alternative technology, called Long-Reach Passive Optical Network (LR-PON), was recently proposed as a more cost-effective architecture for future broadband access network.

Different PONs solutions, including Ethernet PON (EPON) \[11\] and Gigabit-PON (GPON) \[12\], have been standardized and are commercially available today. EPON and GPON are the most popular PON variations found in use today. The evolution of PON technology is classified into three generations: i) First generation (deployed PONs), ii) next generation stage 1 (NG-PON1), and iii)
next generation stage 2 (NG-PON2). This evolution of PON solutions and the corresponding capacities are shown in Fig. 2.5.

![Figure 2.5: PON generations.](image)

The first generation of PON is based on TDMA and provides an EPON downstream rate of 1 Gbit/s while for GPON the downstream bit rate is 2.4 Gbit/s. The NG-PON1 rate is 10 Gbit/s for both standards. In NG-PON2, the required bit rate is at least 40 Gbit/s for downstream transmission and 10 Gbit/s for upstream. Since, high bandwidth applications and Internet services are rapidly increasing the capacity provided by the currently deployed PONs will not be able to meet future demands. For this reason, it will become necessary to upgrade the network to NG-PON1. There are two main ways to achieve an upgrade: i) from already deployed EPON to XG-EPON, or ii) from existing GPON to XG-GPON. It is also possible to upgrade from existing GPON to XG-EPON but this type of migration will require physical and data link layer modifications. In order to find a proper upgrade scenario, the research community is investigating the possible technologies that might be used in NG-PON2, to meet the future requirements of users and network operators [13].
Chapter 2. Long-Reach TWDM Passive Optical Network

Table 2.1: Basic features of GPON and EPON standards

<table>
<thead>
<tr>
<th></th>
<th>GPON</th>
<th>EPON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>ITU G.984</td>
<td>IEEE 802.3 ah</td>
</tr>
<tr>
<td>Capacity Downstream (DS)</td>
<td>1244 / 2488 Mbps</td>
<td>1000 Mbps</td>
</tr>
<tr>
<td>Capacity Upstream (US)</td>
<td>155 / 622 / 1244 / 2488 Mbps</td>
<td>1000 Mbps</td>
</tr>
<tr>
<td>DS/US wavelength</td>
<td>1490 / 1310 nm</td>
<td>1490 / 1310 nm</td>
</tr>
<tr>
<td>Typical split ratios</td>
<td>1:32 / 1:64</td>
<td>1:16 / 1:32</td>
</tr>
<tr>
<td>Distance range</td>
<td>10 - 20 km</td>
<td>10 - 20 km</td>
</tr>
<tr>
<td>Maximum data rate</td>
<td>2.5 Gbps</td>
<td>1 Gbps</td>
</tr>
<tr>
<td>MAC (Framing)</td>
<td>GEM</td>
<td>Ethernet</td>
</tr>
<tr>
<td>Security</td>
<td>AES</td>
<td>Not guaranteed</td>
</tr>
<tr>
<td>QoS</td>
<td>Supported</td>
<td>Not supported</td>
</tr>
<tr>
<td>FEC</td>
<td>Optional RS(255,239)</td>
<td>Optional RS(255,239)</td>
</tr>
</tbody>
</table>

2.1.1 First generation PON: EPON and GPON

The EPON and the GPON standards have the same general principle in terms of framework and applications but their operation is different due to the implementation of the physical and data link layers. In both solutions, the equipment installed at the RN, such as optical splitters/combiners, is fully passive, with a network coverage of up to 20 km. Moreover, both EPON and GPON provide a wide bandwidth to the end users allowing also video broadcasting (digital and/or analogue). The split-ratio varies between 32 and 64 users.

EPON is defined by IEEE 802.3 and it is widely deployed in Asia while GPON is deployed in a number of other regions. GPON requirements were defined by the Full Service Access Network (FSAN) group that was ratified as ITU-T G.984 and is implemented in North America, Europe, Middle East, and Australasia. What makes them different is the MAC protocol and the data encapsulation scheme. While EPON carries bursts of pure Ethernet frames, GPON encapsulates data using Generic Encapsulation Method (GEM) also redefined as GPON Encapsulation Method in [12]. Table 2.1 summarizes the main differences between EPON and GPON.

2.1.2 NG-PON1

The upgrade from currently deployed PONs (EPON or GPON) to Next Generation PON stage 1 (NG-PON1) can be achieved in two main manners: i) from already deployed EPON to XG-EPON, or ii)
2.1. Passive Optical Network

From GPON to XG-GPON

XG-GPON has characteristics which are similar to the deployed GPON with some differences in the physical layer that provide considerable performance improvements. The XG-GPON is divided in two classes: i) XG-GPON1 provides asymmetrical transmission with 10 Gbit/s downstream and 2.5 Gbit/s upstream, and ii) XG-GPON2 which provides 10 Gbit/s symmetrical transmission. In particular, the XG-GPON1 physical layer has been defined in ITU-T G.987.2 [14], while the XG-GPON2 physical layers standard is still to be finalized.

According to the G.987.1 recommendation for XG-GPON1, two scenarios have been proposed to enable migration from GPON to XG-GPON1. The first scenario is a green-field migration which is the replacement of the copper connection into premises with an optical connection. The other option is the PON brown-field migration scenario which is an upgrade of the existing GPON system and this includes replacing or upgrading some of the network components such as ONU units or OLT modules if necessary.

Figure 2.6 shows the coexistence scenario, where the CO consists of two OLTs, one to carry the GPON connection and the other to carry the XG-GPON1 connection. Each OLT transmits its signal to a device that implements WDM to combine both signals in a common fiber. On the user side, a filter device such as a Wavelength Blocking Filter (WBF) is required in order to differentiate the PON data traffic.

2.1.3 NG-PON2

NG-PON2 provides a significant increase in the available bandwidth with a bit rate up to 40 Gbps. In April 2012, hybrid TDM and WDM
(TWDM-PON) technology was selected as the multiplexing technique for NG-PON2 by the FSAN community [20]. This decision was made on different factors such as: technology maturity, system performance, power consumption, and cost. TWDM technology combines the advantage of the high capacity provided by TDM and the large number of wavelengths of the WDM. Within this architecture, TDM frames are transmitted to several users over several wavelengths. Implementing a TWDM-PON requires that each ONU is equipped with a tunable transmitter that is tuned to any of the upstream wavelengths and a tunable receiver that can receive any of the downstream wavelengths. Moreover, this network infrastructure requires an optical amplifier at the OLT in order to amplify the downstream transmission and to pre-amplify the upstream signal. TWDM-PON is classified into static and dynamic approaches. In the static approach, both upstream and downstream wavelengths specified for ONUs are static and do not change during the process. In the dynamic approach, the wavelength can be tuned dynamically based on operation and communication needs. The dynamic approach has advantages over the static method since it allows load balancing, power savings, and resilience.

One of the major advantages of TWDM-PON is the ability to increase the capacity by adding more pairs of wavelengths and different stacking rates. Moreover, using several wavelengths enables
2.2 Long-Reach Passive Optical Network

to deploy TWDM-PON in a pay-as-you-grow manner by adding new pairs of wavelengths when it is needed. Furthermore, TWDM PON is useful for local loop unbundling (LLU) by installing different OLTs that support different sets of wavelengths for each vendor, and using a wavelength-selective device to multiplex the OLT ports over a single fiber. TWDM-PONs have the ability to integrate with the deployed PON in the same network infrastructure since the integration with the previous PON generations depends on ensuring that different wavelength bands are used by different PON generations. Despite the extensive research into developing NG-PON2 technologies, there are still several challenges that have to be addressed such as: increasing capacity, reducing cost, and extending reach. In order to increase the capacity it is possible to increase the number of wavelengths that are transmitted over the same fiber. To reduce the cost of the network exist a new network infrastructure, called Long Reach PON (LR-PON), is proposed. In the next section, the main features and functionalities of LR-PONs are presented.

2.2 Long-Reach Passive Optical Network

LR-PON [19] is a solution that extends PON coverage to enable a greater cost reduction by decreasing the number of the network plant locations. Therefore fewer control units need to be installed (CapEx reduction) and managed (OpEx reduction). The coverage span of PONs is increased from the traditional 20 km range to 100 km and beyond extending the physical reach of the access network to the core network, integrating then the access with the metro network. Figure 2.7 shows how the reach is extended towards the core network when LR PONs are introduced.

A general LR-PON architecture consolidates the multiple headend devices known as Optical Line Terminals (OLTs), which were located at the Local Exchange (LE) within PONs, and the Central Office (CO). An extended shared fiber is then deployed to the LE where optical splitters connect users to the shared fiber. In general, the LR-PON can simplify the network, reducing the number of equipment interfaces, network elements and nodes placement. Since each customer no longer requires a dedicated fiber from the local exchange to the premises, LR-PON induces also a cost per customer reduction. This allows the use of more expensive devices in the LE since the costs of these equipments are shared among a large number of users. On the
Figure 2.7: Reach extension of LR PON.

contrary, it is necessary that the ONU s are as simple as possible to be cost-effective. Furthermore, removal of the metro equipment would free space in local exchange sites, possibly allowing smaller sites to be removed, providing real-estate savings.

LR-PON architectures have two main topological configurations, the branch-and-tree structure, shown in Fig. 2.8, and the ring-and-spur topology, shown in Fig. 2.9. Branch-and-tree topology is composed of a feeder section, between the CO and the Remote Node (RN), and a distribution section where the optical signal is split among different users attached to the same RN. In the ring-and-spur configuration the feeder section is composed of a fiber ring where different trees networks are connected.

In summary, it is possible to make optical access networks more attractive and more feasible economically through a careful design of the network. LR-PON reduces the fiber requirements and, by increasing the split size, reduces the cost of shared devices. In such way, more expensive devices can be installed in the shared section as their cost can be divided among multiple users. Moreover, as expensive and more complex technologies can be used at the OLT or at the RN, at the ONU it is possible to install more simple and cost effective devices allowing a cost reduction of the ONU. However LR-
2.3 Long-Reach TWDM Passive Optical Network

The evolution from LR TDM PON towards LR WDM/TDM PON is a quite recent research topic motivated by the need of such network architectures to support higher capacity. Recently, the Full Service Access Network (FSAN) forum selected as a primary solution for Next-Generation PON Stage 2 (NG-PON2) [9] a flavor of LR-PONs where the transmissions in the upstream channel are allocated through hybrid TDM/WDM access technique. These networks are defined as TWDM PONs. LR TWDM PON system is an interesting high bandwidth, energy-efficient future network solution which combines the advantages of Long Reach, PON, and WDM/TDM systems.

TWDM PONs have several positive features [20]. First, they support a pay-as-you-grow provisioning. A TWDM PON could be deployed starting with a single wavelength pair and then it could be...
be upgraded adding more wavelength pairs, if a higher capacity is required. Also, TWDM PON is useful for local loop unbundling (LLU) by installing different OLTs that support different sets of wavelengths for each vendor, and using a wavelength-selective device to multiplex the OLT ports over a single fiber \cite{9}. This makes TWDM PON a promising option also in the context of FMC networks.

The development of LR TWDM PON architectures \cite{5} - \cite{6} poses a set of new research challenges both in terms of transmission technologies and control protocols. Among the latter, a challenging problem consists in developing new and efficient Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithms for LR TWDM PONs. These algorithms can provide an efficient utilization of the upstream channel and should be devised with the aim of minimizing the average packet delay between the OLT and each ONU.

Moreover, if we assume that different types of traffic are multiplexed over the same wavelengths of a TWDM PON, various drawbacks should also be considered: 1) a single, converged, layer-2 protocol must be used to transport different types of traffic (e.g., fixed traffic from residential users, fixed traffic from business users,
mobile 3G traffic, mobile 4G traffic, ...) which might have very heterogenous requirements (e.g., in terms of bandwidth, QoS, latency or synchronization); ii) the OLT will be required to look up each received frame in order to forward it towards the appropriate destination (as different types of traffic, e.g., 3G, 4G, fixed traffic, etc. might have to be forwarded towards different core nodes); iii) tunable lasers with a large tunability range might be needed, which are still very expensive [21].

2.4 Research Challenges

LR-PON architectures with high distance and high splitting factor introduce a number of new research challenges [23] which need to be addressed, also because these technologies have to be cost efficient in order to be adopted as an access solution.

2.4.1 Network Components

The first of these research challenges is related to the need of compensating the signal power due to the high attenuation introduced by high splitting factor and long reach. With this aim Optical Amplifiers (OAs) can be installed in some network configuration. Unfortunately, these components introduce Amplified Spontaneous Emission (ASE), a device dependent noise, which contributes to lower SNR that is already decreased by the high splitting factor of LR-PONs. Besides, due to burst mode transmission, OAs must be able to adjust their gain fast to output packet with uniform signal amplitude.

In order to lower ONU equipments costs, it is necessary to install optical sources which are cost efficient. A first solution to reach this aim is to use uncooled transmitters which are temperature-dependent and, for this reason, can produce wavelength drift of 20 nm. Wavelength drifts can cause problems when a DWDM system is used because a precise wavelength transmission is required; therefore other solutions like Reflective ONU (R-ONU) have to be investigated. An R-ONU, using Reflective SOAs (RSOAs) modulator [24], generates upstream signal from an external carrier which can be placed at the Remote Node or at the Central Office.

Due to burst mode transmission and different signal powers transmitted by ONUs placed at different distances from the OLT, burst mode-receivers [25] are needed in LR-PONs. Some implementations
of a burst-mode receiver already exist for actual access technologies but, as in LR-PON the speed of links and the number of customers supported is scaled up, it is necessary to investigate new burst-mode receivers with a higher sensitivity and a wider dynamic range.

2.4.2 Dynamic Bandwidth Allocation Algorithms

Considering challenges at levels higher than the physical layer, a wide research field consists of the development of new and efficient Dynamic Bandwidth Allocation (DBA) algorithms for LR-PON [26]. Existing DBA algorithms, used in TDM PON, are not efficient if used in LR-PON due to high round trip time (RTT) introduced by the long distances covered by this technology. Such feature of LR-PON systems causes long delays in the signaling procedure, affecting in this way the whole network performance. Therefore it is necessary to implement new DBA algorithms which:

- avoid performance decay due to a long signaling message delay;
- support different classes of services;
- are scalable in terms of the number of efficiently supported users.

Moreover, with the introduction of LR WDM/TDM PONs, it is also introduced the issue of how to efficiently assign to each ONU not only the time but also the wavelength to be used in order to improve the channel utilization. Furthermore, the average packet delay must be kept small in order to improve performance and to allow the network to support new Internet media services like symmetrical HD real time applications. Another issue of Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithms is to fairly share the available bandwidth among all the ONUs also providing Quality of Service (QoS) to the end users. Nowadays several DBAs for LR PON have been presented, while only few DBWAs for LR WDM/TDM PON [27] have been proposed. Basically, all these algorithms can be grouped in two categories: QoS-aware and QoS-unaware schemes. In particular, in this thesis, we investigate QoS-unaware DBWA methods with the objective of developing new algorithms that efficiently exploit the available bandwidth while reducing the end-to-end traffic delay.
2.4.3 Network Protection

Another challenge is protection of the LR-PONs. As LR-PONs utilize the large transmission capacity of optical technology, and it is oriented to serve a large number of end users, any network failure may cause a significant loss of data. Therefore it is necessary to implement efficient protection schemes in LR-PONs \cite{28}. A basic protection system is naturally implemented in the ring-and-spur topology \cite{29} exploiting the two different transmission directions of the ring in case of fiber or node failures. In branch-and-tree topologies, instead, a dual homing technique can be used in the feeder fiber section in order to recover from fiber failures. The dual homing technique consists in connecting the same LR-PON to different CO/OLT to ensure that, in case of single node failure, at least one of these two provides connection to the network. Protection schemes for LR-PON need further investigation.
In this chapter we present Dynamic Bandwidth Allocation (DBA) schemes for LR-PONs and Dynamic Bandwidth and Wavelength (DBWA) schemes for LR TWDM PONs. First of all, in Section 3.1 we explain the signalling protocol used by the allocation algorithms that we propose and evaluate. In Section 3.2 we explain in details the different characteristics of DBAs. Then, in Section 3.2.1 we explain the functionalities of the Interleaved Polling with Adaptive Cycle Time which is one of the first solutions to dynamically allocate the bandwidth in PONs. After introducing this algorithm, we present in Section 3.2.2 another allocation scheme, based on IPACT, which is instead designed specifically for LR-PONs. This algorithm, named Multithread Polling, is implemented to reduce the average packet delay of LR-PON and to achieve a better upstream channel utilization. Then, we present algorithms from the literature designed for LR TWDM PON (i.e., DBWAs) and used in this thesis as benchmarks. These DBWAs, called Earliest Finish Time (EFT) and Earliest Finish Time with Void Filling (EFT-VF), are based on the functionalities of
Chapter 3. Dynamic Bandwidth and Wavelength Allocation Schemes

IPACT algorithm. The EFT is presented in Section 3.3.1 while the EFT-VF is presented in Section 3.3.2. Finally, we devote Section 3.4 to the description of the simulation tool that we used to evaluate all the DBAs and DBWAs studied in this work.

3.1 Multi-Point Control Protocol

All the algorithms presented in this work use as a signalling protocol the Multi-Point Control Protocol (MPCP), specified in the IEEE 802.3ah standard [11], in which five types of control frames are defined: REGISTER REQUEST, REGISTER, REGISTER ACK, REPORT and GATE. While the first three are used by the ONU during the discovery and registration procedure, the last two are used to plan the data transmissions. In particular the Report message is used by the ONU to communicate to the OLT the status of the queue. The Gate instead is used by the OLT, which is the network element which manages the scheduling policy, for granting the upstream timeslot to each ONU. MPCP is not concerned with a particular bandwidth allocation scheme but it is a supporting mechanism that can facilitate the implementation of different bandwidth allocation algorithms in PONs and LR PONs.

3.2 Dynamic Bandwidth Allocation Algorithm

A Dynamic Bandwidth Allocation (DBA) algorithm is a technique by which the bandwidth in a shared medium can be allocated on demand and fairly among different users. Within this technique, the sharing of a link adapts to the instantaneous traffic demands of the ONUs connected to the network. In particular, a DBA schedules the transmissions of the ONUs in time, i.e., it assigns a point in time, called transmission start time, in which a ONU can start its transmission without colliding with any other transmission. A DBA takes advantage of the fact that i) the ONUs connected to the network are not always transmitting data, and ii) the traffic of the ONUs mainly occurs in bursts. The DBA dynamically allocates the amount of bandwidth for each ONU transmission according to the amount of traffic that an ONU need to transmit in that particular moment. The scheduling decision is taken by the OLT for each ONU in each polling cycle (also called cycle time). The cycle time ($T_{cycle}$) is the time interval between successive requests from the same ONU, where
every ONU transmits exactly once.

The scheduling problem, solved by a DBA algorithm, is described by the authors in [30] as a layered scheme. The first layer is defined scheduling framework, and the second layer is the scheduling policy. The scheduling policy is a method for the OLT to produce the schedule whereas the scheduling framework is a logistical framework that determines when the OLT makes scheduling decisions.

The scheduling framework can be divided in three main categories: Online scheduling, offline scheduling and just-in-time scheduling. This last method is a new scheduling framework defined by McGarry et al. in [30]. Simulation results [31] comparing Online and Offline scheduling show that the former significantly outperform the latter with respect of average queue length. The Online scheduling framework allocates the transmission of each ONU as soon as the Report message is received. For this reason, the average delay of the transmissions and the average queue length are lower if compared with the Offline policy. The Offline scheduling framework collects for each polling cycle all the Report messages of the ONUs and, according to the information received, the OLT allocates all the transmissions of the ONUs at the same time, at the end of the polling cycle. With this method, the OLT can perform the optimal allocation since it knows exactly all the requests of the ONUs. The drawback of this policy is that it introduces delay to the transmissions. This happens because the ONUs have to wait at least until the end of the current polling cycle to have a scheduling. Just-In-Time (JIT) algorithm can be viewed as a compromise between the offline scheduling and the online scheduling approaches. In this scheduling framework, ONUs are added to a scheduling pool as their Report messages are received by the OLT. When a channel becomes available, all the ONUs in the same pool are scheduled together according to a given scheduling policy. ONUs scheduled such that their transmissions will start shortly after the time when their scheduling is produced are classified as "imminent" and, for these ONUs, the current schedule is considered firm and the OLT transmits then Gate messages to inform them of their granted transmission window. For the other ONUs in the pool the scheduling is classified as "tentative" and it could be rearranged in the next scheduling round when new Report messages from other ONUs arrive, providing new information to the OLT. An alternative approach is to firmly schedule the transmissions of all ONUs in each scheduling round. Authors refer to this second method as online JIT
while the first case, where the tentative ONUs participate in future scheduling rounds, is classified as online JIT Tentative. The online JIT scheduling framework has the ability to make better scheduling decisions if compared with the online scheduling framework because JIT accumulates a set of Report messages before choosing a scheduling. Such behavior gives the possibility to the OLT to better organize the upstream transmissions in order to have a better channel utilization. Moreover, the JIT framework gives a lower average transmission delay if compared with the offline framework, as it does not wait for all the Report messages of all ONUs, but it collects only the Report transmitted until the moment when a channel becomes available.

The scheduling policy is the algorithm used by the OLT to schedule the transmission of the ONUs. In this work we only consider as a scheduling framework the Online scheduling, and we focus on the design of the scheduling policy.

Moreover, the DBWA problem can be divided in grant sizing and grant scheduling problem. The grant sizing problem denotes the amount of bandwidth to be assigned $B_i$ to ONU$_i$ in each polling cycle. Considering the grant sizing problem, exist different policies which can be used to assign the amount of bandwidth to be assigned $B_i$ to each ONU in each polling cycle. $B_i$ is the amount of bandwidth in bytes that an ONU$_i$ is allowed to transmit in each polling cycle $T_{cycle}$. $B_{max}$ is the maximum amount of bandwidth that each ONU is allowed to transmit in a polling cycle.

- **Fixed service** does not consider the size of the transmission requested by the ONU in its Report message, and it always grant a fixed maximum window $B_{max}$. As a result it has a constant cycle time $T_{cycle}$. This method corresponds to the fixed TDMA PON system.

- **Limited service** grants the requested number of bytes, but no more than $B_{max}$. This type of service is the most conservative scheme and has the shortest cycle of all the grant sizing schemes.

- **Constant credit** is a scheme that adds a constant credit to the requested window size. the idea behind adding the credit is: assume $x$ bytes arrived between the times an ONU sent a Request and received a Grant. If the granted window size equals requested window + $x$ (i.e., it has a credit size of $x$), these $x$
3.2. Dynamic Bandwidth Allocation Algorithm

bytes will not have to wait for the next Grant message to arrive; they will be transmitted with the current Grant, and the average packet delay will be shorter.

- The Linear credit policy uses an approach similar to the Constant credit scheme. However, the size of the credit is proportional to the requested window. The main idea here is that: network traffic possesses a certain degree of predictability [32]. If we observe a long burst of data, this burst is likely to continue for some time into the future. Correspondingly, the arrival of more data during the last cycle may signal that we are observing a burst of packets.

- In the Elastic service mechanism, the only limiting factor is the maximum cycle time \( T_{\text{cycle},\text{max}} \). The maximum window is granted in such a way that the accumulated size of last \( N \) Grants (including the one being granted) does not exceed \( N \times B_{\text{max}} \) bytes (where \( N \) is the number of ONUs). Thus, if only one ONU has data to send, it may get a Grant of size up to \( NxB_{\text{max}} \).

The grant scheduling problem stands for the time and wavelength to transmit data in an efficient resource-wise manner, for example by minimizing the packet delay within the network. The main issue of the bandwidth distribution is the grant scheduling problem.

In this thesis, we mainly investigate the problems related to the grant scheduling, and we use as grant sizing policy the Limited service.

In the following, we will explain the main functionalities of two DBAs that we consider in our work: the Interleaved Polling with Adaptive Cycle Time (IPACT), and the Multithread algorithms.

3.2.1 Interleaved Polling with Adaptive Cycle Time

In this section we will give an overview on the Interleaved Polling with Adaptive Cycle Time (IPACT) algorithm [33]. This scheduling scheme has been designed for PONs, more precisely for EPONs. Based on the main principle and functionalities of this algorithm, a large number of other allocation schemes have been proposed after IPACT. For this reason it is important to clarify in this section what is the basic idea on which IPACT is designed. We assume that we have a PON with three ONUs as shown in Fig. 3.1.
Chapter 3. Dynamic Bandwidth and Wavelength Allocation Schemes

1. Assume that at time $t_0$ OLT knows exactly how many bytes buffered at each ONU and the round trip time (RTT) to each ONU. OLT stores this information on its polling table and starts sending a Gate message to $ONU_1$. The Gate message should contain the ID of $ONU_1$ as well as the size of the granted window.

2. Once $ONU_1$ receives its Gate message and it starts sending its buffered data up to the granted window size which is 8000 bytes in this example. At the end of the transmission window, $ONU_1$ will generate and send its Report control message, which allows the OLT to know exactly the newly requested window size for the next cycle. In this example $ONU_1$ asks for a transmission window of 600 bytes.

3. Since the OLT knows exactly the RTT of $ONU_1$ and how many bytes this ONU will send, it can schedule the control Gate message of $ONU_2$ so that there will be no data collision.

4. Upon receiving data and Report message from $ONU_1$, the OLT updates its polling table for the next polling cycle.

5. Similarly, the OLT can schedule the transmission of $ONU_3$ Gate message as it knows the RTT of $ONU_2$ and how many bytes $ONU_2$ will send.

If the OLT authorizes each ONU to send its entire buffer contents in one transmission, ONUs with high data volume might monopolize the entire bandwidth. To avoid this, the OLT will limit the maximum transmission size to a maximum limit $B_{max}$. Thus, every ONU will be allowed to send as many bytes as it has requested in a previous polling cycle, but no more than some maximum limit $B_{max}$. Actually many other types of service can be applied within this allocation scheme, however is shown in [33] that the limited service discipline is the most conservative scheme and has the shortest cycle of all the other schemes. In this work, the limited service implemented in IPACT is then used in all the proposed allocation schemes to choose the length of data that each ONU can send in each polling cycle.

As shown in many works, IPACT has degraded performance when it is applied in LR-PONs. This is due to the large network span and to the consequently increased RTT of LR-PONs with respect to PONs. For this reason the Multithread Polling has been proposed for LR-PON systems.
3.2. Dynamic Bandwidth Allocation Algorithm

Figure 3.1: Basic idea of the IPACT algorithm.

3.2.2 Multithread Algorithm

Multithread (MT) algorithm [34] has been proposed as a solution to overcome the problem of the increased RTT in a LR TDM PON, with respect to TDM PONs, which leads in general to an increased average packet delay. Therefore, to achieve better performance in terms of packet delay in a LR-PON, the basic idea of the Multithread algorithm is to allow an ONU to send its Report before the previous Gate message is received. Practically, this allocation scheme exploits the benefits of having multiple polling processes running simultaneously. Users are then enabled to send bandwidth requests before receiving acknowledgement from the OLT for the previously requested data. The basic idea of the Multithread polling where two ONUs and two threads (denoted by red and black) are used is shown in Fig. 3.3. The black polling process (also called thread) is, in this example, the same shown in Fig. 3.2, where a single-thread polling scheme is exemplified. Before the GRANT message for this first thread is received, the ONUs start a new polling process (denoted by red) sending to the OLT a new message. In such way, the ONUs do not have to wait until the
end of data transmission of the first thread to send a new Report message asking for a new transmission opportunity.

### 3.3 Dynamic Bandwidth and Wavelength Allocation Algorithms

Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithms dynamically assigns both the wavelength over which an ONU can transmit and the transmission start time. This types of algorithms are used in TWDM PONs where the transmissions of the ONUs have to be allocated both in time and over a certain wavelength in order to avoid collisions. Also DBWAs, like generic DBAs, can be viewed as consisting of grant sizing and grant scheduling problems [41].

In the following, we will explain the main functionalities of two DBWAs that we consider in our work: the Earliest Finish Time (EFT), and the Earliest Finish Time with Void Filling (EFT-VF) algorithms. In Table 3.1 we define the variables used hereafter.
3.3. Dynamic Bandwidth and Wavelength Allocation Algorithms

3.3.1 Earliest Finish Time Algorithm

The Earliest Finish Time (EFT) algorithm [35] is the most straightforward and less complex algorithm to dynamically allocate the bandwidth and the wavelength in a LR WDM/TDM PON. Within this algorithm the scheduling is obtained according the rule that no other transmission in the same wavelength should have been programmed after the selected schedule, while the selected wavelength should have the earliest finish time among all channels. In particular, wavelength assignment and the scheduling, according to EFT algorithm, are calculated as:

\[ w_{eft} = \arg \min_h (L_h), \quad h \in W_i \]  \hspace{1cm} (3.1)

\[ t_{eft} = \arg \max_j (F_{w, j} + t_{guard}) \]  \hspace{1cm} (3.2)

In Fig. 3.4 we show an example of the functionalities of the EFT. Figure 3.4 shows an upstream channel with two wavelengths (\( W_1 \) and \( W_2 \)) where some transmissions are already allocated (in blue). The figure shows the start time (ST) chosen by the EFT algorithm.

Despite its simplicity and multiplexing gain due to the possibility of solving upstream contention in the wavelength domain, in the EFT algorithm the upstream channels are not utilized as efficiently as possible. In fact it is expected to create scheduling voids between two subsequent scheduled transmissions on the same wavelength. This issue is more accentuated when the ONUs have distances from the OLT which are very different from one another. In such way EFT leads to a waste in the upstream bandwidth and resulting in increased frame queuing delay.
3.3.2 Earliest Finish Time with Void Filling Algorithm

The Earliest Finish Time with Void Filling (EFT-VF) algorithm, which is an enhancement of the EFT algorithm [35], is based on the observation that in a PON and in LR-PONs different distances between OLT and ONUs may lead to very diverse propagation delays creating what authors define as scheduling voids. A scheduling void is a period of time between two subsequent transmissions where there is no scheduled transmissions on the channel. The Void Filling part of the EFT-VF algorithm aims at filling these voids by scheduling other transmissions during the time when the channel is unused. A void must be long enough to enable transmissions; a time period with this feature is called eligible void and its length in bytes is equal to the length of data requested by the ONU plus the Report message. An eligible void is defined according to Eqn. 3.3, where $S_{h,j+1}-F_{h,j}$ is the time distance between the reception by the OLT of the first bit of the $(j+1)$-th transmission on the $h$-th wavelength ($S_{h,j+1}$), and the reception of the last bit of the $j$-th transmission on that same wavelength ($F_{h,j}$). $t_i + R_i + t_c$ is the minimum allocation time for ONU$_i$, depending on the round trip time ($R_i$) and on the time needed to send the gate message to the ONU ($t_c$). $S_{h,j+1}-(t_i + R_i + t_c)$ is the time distance between the reception at the OLT of the first bit of the $(j+1)$-th transmission on the $h$-th wavelength, and the time when the first bit of the transmission of ONU$_i$ can reach the OLT, according to its round trip time $R_i$.

$$V_i = \{F_{h,j} | S_{h,j+1} - F_{h,j} \geq t_g + t_c\}, \quad h \in W_i \tag{3.3}$$

It has been proven that the EFT-VF algorithm yields better performance than EFT. The EFT-VF algorithm adds to the functionalities of the EFT defined in Eqns. 3.1, 3.2 the Void Filling procedure that is defined in Eqns. 3.4 and 3.5.

$$w_{vf} = \arg\min_{h}(F_{h,j} | F_{h,j} \in V_i), \quad h \in W_i \tag{3.4}$$

$$t_{vf} = \arg\min_{j}(F_{w,j} | F_{w,j} \in V_i) + t_{guard} \tag{3.5}$$
3.3. Dynamic Bandwidth and Wavelength Allocation Algorithms

The wavelength and time scheduling for EFT-VF algorithm are finally assigned according to the equation:

\[ w = \min(w_{eft}, t_{vf}), \quad t = \min(t_{eft}, t_{vf}) \]  

(3.6)

In Fig. 3.5 we show an example of the functionalities of the EFT-VF algorithm. Figure 3.5 shows an upstream channel with two wavelengths (W₁ and W₂) where some transmissions are already allocated (in blue). The eligible voids for the transmission of ONUᵢ that have to be allocated are shown in red. In the first case (Case 1) we show an example on a scenario where the Void filling part of the algorithm is used. In this first case, the first available transmission start time for the ONUᵢ that have to be allocated is in a void. In the second case (Case 2), we show that the start time (ST) is chosen according to the EFT algorithm. In this case, the eligible voids for the transmission of ONUᵢ allow to assign a start time which is not the minimum according to this scenario.

The EFT-VF algorithm gives a large improvement in the average transmission delay, if compared with the EFT scheme, only when it is used in a network where the distances between OLT and ONUs have
Chapter 3. Dynamic Bandwidth and Wavelength Allocation Schemes

a very large distribution, e.g., between 500 m and 100 km. Instead, EFT-VF is applied in a network where the ONUs-OLT distances are not so spread, the improvement provided by these two algorithms is limited. Basically, the EFT-VF algorithm tries to solve the same problem addressed by the Multi-thread algorithm, using different strategies. Multi-thread algorithm aims at reducing the waste of bandwidth in each polling cycle adding two or more threads in the middle of the first polling cycle. Essentially, multiple transmission threads are maintained between each ONU and the OLT.

3.4 DESL: Discrete Event Simulation Library

In this section we present the Discrete Event Simulation Library (DESL) which is the simulation tool used to study all the algorithms in this work. DESL [36] is a C++ library developed under Microsoft Visual C++.NET 2003. It contains classes and routines that facilitate creating simulation models and performing an analysis. DESL is a collection of classes that allow creation of a Pending Event Set (PES) engine as well as creation of network elements (objects) that are able to generate or process various simulated events. The original version of DESL which can be downloaded from [36] implements the IPACT algorithm designed for Ethernet Passive Optical Networks. The connection of the elements of the DESL used to implement the IPACT project is shown in Fig. 3.6. This simulation tool only implement the upstream transmissions. This means that in downstream are transmitted only the signalling messages, while in upstream both data packet and control messages are transmitted.

In order to implement the DBWAs studied in this work, we extend this tool to simulate a LR TWDM PON. Therefore, several changes have been apportioned to the original version from [36]. First, the reach of the simulated network is extended to a maximum of 100km. Then new virtual network elements are added to simulate the hybrid WDM/TDM transmission. These virtual network elements are added only for simulation purposes: they do not have any relation with devices of the real networks, and they do not add any kind of delay to the processing or transmission of the packets that they forward. These virtual network elements are the Virtual OLTs and Virtual ONUs. These nodes are needed to simulate the transmissions of the ONUs over multiple wavelengths: one Virtual ONU is added for each real ONU and for each wavelength. This means that if an ONU
3.4. DESL: Discrete Event Simulation Library

Figure 3.6: Connections of the elements for the original IPACT project [36].

can transmit over $W$ wavelengths, $W$ Virtual ONUs are connected to this ONU. Regarding Virtual OLTs, one of these nodes is added for each wavelengths used in the network. A scheme representing the setup of our simulation is presented in Fig. 3.7. Figure 3.7 shows the upstream channel from the traffic source to the OLT. The traffic source generated the traffic for each ONU. In particular, in our setting to reflect the property of the real Internet traffic, the traffic source generate self-similar traffic by aggregating multiple sub-streams, each consisting of alternating Pareto-distributed ON/OFF periods, with a Hurst Parameter of 0.8. The simulator generates Ethernet frames with a length distributed between 64 and 1518 bytes. Then, the traffic generated by the traffic source is transmitted to the ONUs that buffer the packet received in a queue. In our settings, the buffer size of each ONU has a length of 10 Mbytes. The ONUs communicate with the OLT to ask for a transmission opportunity. When the OLT assigns to an ONU a transmission start time over a particular wavelength and transmission length ($B_i$), the ONU send a set of packets (up to $B_i$) to the Virtual ONU corresponding to the assigned wavelength. The Virtual ONU receives and forward the packet to the Virtual OLT. In the Virtual OLT the arrival time of the packet is registered and the delay of each packet is computed subtracting to the arrival time the value of the time stamp assigned to the packet when it is generated. The Report message is transmitted after each burst of
Figure 3.7: Connections of the elements for the modified simulator used in this work.

packets and, after reaching the Virtual OLT, is forwarded to the OLT. The information contained in this Report message is used by the OLT to perform a new assignment for the following polling cycle. In fact, the DBWA algorithm is implemented in the OLT.

The overall delay experienced by a packet is composed by the propagation delay, the transmission delay, and the queuing delay. The queuing delay depends on the time a packet has to wait in the queue of the ONU. The propagation delay and the transmission delay are added by the link connecting Virtual ONUs and Virtual OLTs. The propagation delay is the amount of time needed to transmit a packet over a fiber of length \( l \). The propagation speed of the fiber is \( 2 \times 10^8 m/s \). Therefore, the propagation delay added by a link with length \( l = 100km \) is \( (100 \times 10^3)/2 \times 10^8 \) which is equal to 0.5ms. The transmission delay is computed by dividing the length of a packet in bits by the transmission speed of the link in bit/s. For example, if the packet is 1000 bits and the link has a capacity of 10 Mbit/s, then the resulting transmission delay is 0.8 ms. The only link that adds delay to the transmissions is the link connecting Virtual ONUs and Virtual OLTs. In fact, the link connecting the ONU to the Virtual ONU, and the Virtual OLT to the OLT are virtual network elements in a sense that these links only forward the packets to the subsequent node, but they do not add any delay to the transmissions.
3.4. DESL: Discrete Event Simulation Library

All the parameter used in our simulations are specified in each chapter.
### Table 3.1: Definition of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>Time when the Report from $ONU_i$ arrives at the OLT.</td>
</tr>
<tr>
<td>$t_{guard}$</td>
<td>Guard band time between two subsequent transmissions of different ONUs.</td>
</tr>
<tr>
<td>$R_i$</td>
<td>Round-trip propagation time between OLT and $ONU_i$.</td>
</tr>
<tr>
<td>$t_c$</td>
<td>Time needed for the transmission of Report or Gate frame.</td>
</tr>
<tr>
<td>$W_i$</td>
<td>Set of wavelength supported by $ONU_i$.</td>
</tr>
<tr>
<td>$S_{h\text{j}}$</td>
<td>Arrival time at the OLT of the first bit of the $j$-th scheduled transmission on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$F_{h\text{j}}$</td>
<td>Finish time of the $j$-th transmission on the $h$-th wavelength corresponding to the reception of the last bit of the control frames which is piggy-backed to the data packets.</td>
</tr>
<tr>
<td>$L_h$</td>
<td>Finish time of the last transmission scheduled on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$V_i$</td>
<td>Set of eligible voids for $ONU_i$ calculated according to its $R_i$ and its requested length of data, which must be granted.</td>
</tr>
<tr>
<td>$t_{g,i}$</td>
<td>Time needed to transmit the length of data which $ONU_i$ asks to transmit. Can be partitioned.</td>
</tr>
<tr>
<td>$w_{e\text{eft}}$</td>
<td>Upstream wavelength which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$w_{e\text{ef}}$</td>
<td>Upstream wavelength which could be chosen by the Void Filling algorithm.</td>
</tr>
<tr>
<td>$t_{e\text{eft}}$</td>
<td>Transmission start time within the $w_{e\text{eft}}$ wavelength which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$t_{v\text{ef}}$</td>
<td>Transmission start time within the $w_{v\text{ef}}$ wavelengths, which could be chosen by the Void Filling algorithms.</td>
</tr>
<tr>
<td>$w$</td>
<td>Chosen upstream wavelength.</td>
</tr>
<tr>
<td>$t$</td>
<td>Transmission start time within the $w$ wavelength.</td>
</tr>
</tbody>
</table>
Dynamic Bandwidth Allocation (DBA) is a major research challenge in the migration of Passive Optical Networks (PON) systems towards Long Reach PON, especially when hybrid WDM/TDM is used (LR WDM/TDM PON). New solutions for Dynamic Bandwidth and Wavelength Assignment (DBWA) are sought to address two main critical issues of such PON systems: how to schedule the transmissions over multiple wavelengths and how to efficiently exploit the bandwidth in presence of long propagation delays of long reach scenarios. Among the already proposed DBWA algorithms, one of most promising is the Earliest Finish Time with void filling (EFT-VF) algorithm for LR WDM/TDM EPON (Ethernet-based PON) [35].

The contribution of this chapter is threefold: i) we investigate the behaviour of the Multi-thread polling algorithm, by extending in the case of a LR WDM/TDM PON. To the best of our knowledge, the Multi-thread polling has never been applied to a hybrid WDM/TDM access network. ii) We propose a new algorithm based on the idea of EFT-VF
and Multi-thread schemes, designed to improve the performance of these two allocation solutions. iii) We extend the definition of polling cycle to a hybrid WDM/TDM PON and we define an equation to calculate the maximum amount of bytes that each ONU can transmit in a cycle time, given the cycle time length. It is important to be able to calculate this maximum amount of traffic per cycle to guarantee a certain bandwidth to each ONU. Moreover, we aim at evaluating the improvement introduced by the statistical multiplexing when LR TWDM PON are used. In particular, we study the performance of the already proposed EFT algorithm in different scenarios. Based on such evaluations, we discuss some possible guideline to efficiently accommodate different types of traffic (e.g., data traffic, voice traffic) in combined structures as Fixed and Mobile converged (FMC) access networks.

The algorithms proposed in this chapter use as a signalling protocol the Multi-Point Control Protocol (MPCP) [11]. In this work we evaluate only DBAs for EPON [11]. The possible extension of these ideas to GPON has not been investigated and is out of the scope and objectives of this research.

The rest of the chapter is organized as follows: in Section 4.1 we present a quick survey of existing DBWA solutions. Section 4.2 describes the Multi-thread polling implementation in a LR WDM/TDM PON. In Section 4.3 we introduce our proposed algorithm. In Section 4.4 we extend the definition of polling cycle to a hybrid WDM/TDM network and we propose an equation to calculate the cycle time, given the network parameters. In Section 4.5 we provide a simulation analysis and illustrative results of the previously presented algorithm implementations, over different network scenarios. Then, in Section 4.6 we evaluate the already proposed EFT algorithm in different network scenario in order to evaluate the gain introduced by statistical multiplexing. Section 4.7 is devoted to the discussion of novel possible network solutions for the convergence of fixed and mobile (FMC) traffic backhauling. Finally, in Section 4.8 we give concluding remarks on the performance of the different scheduling disciplines.

The work presented in this chapter was published in:

4.1. Related works


### 4.1 Related works

A number of DBWA for both online and offline scheduling has been presented in the last years, some of which are architecture dependent [27], i.e. they are designed for particular network solutions. The first of these is specifically designed for the STARGATE EPON [37]. Another existent architecture dependent DBWA is designed for SPON [38]. In [39] is presented a DBA for the SARDANA network. Other proposed solutions can be used for generic network architectures. Among them, there is GATE Driven DBA (GD-DBA) [40].

The Latest-Finish Time with Void Filling (LFT-VF), Distanced Based Grouping (DBG) and Earliest-Finish Time with Void Filling (EFT-VF) algorithms [35] take advantage of the distribution of the distances from the OLT to the ONUs, trying to remedy the inefficiencies in the utilization of the upstream channel given by this distribution. Even if the DBG algorithm gives satisfactory performance, this solution has the drawback of reducing the WDM multiplexing gain within each group of ONUs due to the lower amount of wavelengths assigned to each of these groups. LFT-VF algorithm, instead, chooses the channel with the latest horizon. In particular the selected wavelength must have the latest finish time among all channels. The void filling part keeps track and tries to fill the voids left on the upstream channel. The EFT-VF algorithm chooses the channel where the previously scheduled transmission will end first. Both the LFT-VF and the EFT-VF algorithms, having a similar performance, give a large improvement in the average transmission delay, if compared with the EFT scheme, only when they are used in a network where the distances between OLT and ONUs have a very large distribution, e.g., between 500 m and 100 km. Instead, when LFT-VF and EFT-VF are applied in a network where the ONUs-OLT distances are not so spread, the improvement provided by these two algorithms is limited, like
shown later in this chapter. Basically, these last three algorithms try to solve the same problem addressed by the Multi-thread algorithm [34], using different strategies. Multi-thread algorithm aims at reducing the waste of bandwidth in each polling cycle adding two or more threads in the middle of the first polling cycle. Essentially, multiple transmission threads are maintained between each ONU and the OLT.

In this work we focus on the decision of when to schedule the transmission of ONUs with the aim of reducing the average packet delay. We assume that the grant sizing has already been solved during a preprocessing phase where it is calculated the maximum amount of data that can be transmitted by each ONU in each cycle time in order to guarantee certain bandwidth to each ONU.

### 4.2 Multi-thread implementation in LR WDM/TDM PONs

In this section, we introduce our proposed extension of the Multi-thread algorithm over LR WDM/TDM PONs to which we refer as WDM-MT. To the best of our knowledge, this is the first time that the Multi-thread is applied in LR WDM/TDM PONs. When MT is implemented in a WDM/TDM network, the basic idea of having multiple polling processes running simultaneously remains the same and furthermore the benefits of having more wavelengths are exploited. However, the transmission of the a generic thread \( t \), for ONU \( i \), must be scheduled after the end of the transmission of thread \( t - 1 \), because a single ONU can only send one transmission at a time. As a consequence, it may happen that on a particular wavelength a transmission is not scheduled at the earliest available time of the channel, but it is forced to wait till the end of the previously granted transmission for the same ONU. We refer to this scheduling constraint as thread coordination. This constraint introduces an additional delay and it may lead to inefficiencies in the utilization of the channel. By using multiple wavelengths, the average time between two subsequent transmissions of the same ONU is significantly decreased. However, note that, the greater the number of wavelengths used, the more the thread coordination constraint affects the algorithm performance. The fact of having more wavelengths gives a higher possibility to allocate more transmissions at the same time. Unfortunately, the thread coordination constraint limits this possibility avoiding that different transmissions of the same ONU are allocated during the same period. If two different transmissions of the same ONUs can not
be scheduled during the same time period, may happen that one of these two transmissions will be delayed until the end of the previous one. In such way, more voids are created in the channel. In the following section we present an allocation strategy which combines the positive features of the EFT-VF and Multi-thread algorithms, to improve the overall performance of the allocation scheme.

### 4.3 EFT-Partial-VF Multi-threaded

In this section, we try to exploit the positive features of Multi-thread and EFT-VF algorithms, by combining them appropriately. Therefore, in this section we present our proposed allocation scheme called EFT-Partial-VF Multi-threaded (EFT-partial-VF MT). In Tab. 7.1 we define the notation used hereafter.

We assume that the laser tuning time is negligible in considered architecture. Note that this assumption reasonably holds for advanced technologies such as fast tunable lasers and for the multi-wavelength lasers [42].

**Case 1: VF Algorithm used**

![Diagram of EFVVF Algorithm with Partial Void Filling, Multi-threaded](image)

- **Void chosen ≥ Data Length requested + Report Length**

**Case 2: EFT Algorithm used**

- **Start:** Start Time
- **Wj:** Wavelength
- **R:** Report transmission
- **Vj:** Eligible void

**Figure 4.1:** Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded.
Chapter 4. Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs

Figure 4.2: Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded when the Void Filling solution is used to schedule both first and second transmission.

At a high level, this algorithm can work under two modes of operation namely the Single Thread (ST) operation and the Multi-thread (MT) operation. The algorithm stays in the ST operation if a transmission can be completed without partitioning, for example in a wavelength channel where the horizon is idle at the time when data are ready for transmission or a void large enough is found; instead the algorithm moves to the MT operation if the transmission needs to be partitioned into P trunks. Furthermore, an ONU returns in ST operation if the REPORT messages of at least \(P - 1\) trunks report an amount of data equal to zero as extensively explained in the following. The scheme switches from the ST operation to the MT operation independently for each ONU, according to the traffic load provided by each user. This means that at the same time some ONU can be in MT operation while some other ONU can be in ST operation. More specifically, in the ST operation the decision on where to schedule the transmission is made following the same strategy of the EFT-VF algorithm. Differently than EFT-VF, in the EFT-Partial-VF MT, an eligible void is defined, according to Eqn. (4.1), as a time period between two already scheduled transmissions which is long enough to schedule at least a fraction \(P\) of the requested length of data.

\[
V_i = \{F_{hrj} | S_{hrj+1} - \max(F_{hrj}, t_i + R_i + t_c) \geq t_{gi}/P + t_c\},
\]

where \(S_{hrj+1} - F_{hrj}\) is the time distance between the reception by the OLT of the first bit of the \((j+1)\)-th transmission on the \(h\)-th wavelength,
and the reception of the last bit of the \( j \)-th transmission on that same wavelength. \( t_i + R_i + t_c \) is the minimum allocation time for \( \text{ONU}_i \), depending on the round trip time and on the time needed to send the gate message to the ONU. \( S_{h, i+1} - (t_i + R_i + t_c) \) is the time distance between the reception at the OLT of the first bit of the \((j + 1)\)-th transmission on the \( h \)-th wavelength, and the time when the first bit of the transmission of \( \text{ONU}_i \) can reach the OLT, according to its \( R_i \). In such way, a transmission for a single \( \text{ONU}_i \) can be divided into multiple blocks of data, each of which having a different \( T_{i,rj} \). These \( T_{i,rj} \) are all allocated through the equations listed below. If a transmission is partitioned in more portions, an additional constraint is needed in order to schedule the blocks of data following the first one. Such condition must guarantee that the different transmissions do not overlap since an ONU can not usually send multiple transmissions at the same time.

**Figure 4.3:** Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded when the first part of the transmission is scheduled through the Void Filling algorithm, while for the second the EFT strategy is used.

The equations used to define the wavelength assignment and the scheduling, according to EFT algorithm, are:

\[
w'_{eft} = \arg \min_h L_h, \quad h \in W_i
\]  \( \quad (4.2) \)

\[
n'_{eft} = \arg \max_j F_{w_{rj}} + 1
\]  \( \quad (4.3) \)
Chapter 4. Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs

Whereas, for the Void Filling part of the algorithm the wavelength assignment and the scheduling are chosen according:

$$w_{vf} = \arg \min_h (F_h | F_h \in V_i), \ h \in W_i$$  \hspace{1cm} (4.4)

$$n_{vf} = \arg \min_j (F_w | F_w \in V_i) + 1$$  \hspace{1cm} (4.5)

Then, the wavelength and time scheduling finally assigned for a transmission are chosen according to the equation:

$$w = \min(w_{eft}, w_{vf}), \ n = \min(n_{eft}, n_{vf})$$  \hspace{1cm} (4.6)

The additional constraint to schedule subsequent transmission blocks for ONU$_i$ is defined by Eqn. (4.7), using the notation:  

$n_{ip}$: Position index of the reservation within the wavelength $w$ for the p-th block of data for the ONU$_i$, with $p \in [1,P]$ integer.  

t$_p$: Time needed by the ONU to transmit the length of data granted in the p-th transmission block.

$$n_{ip} \geq n_{ip-1} + t_{p-1};$$  \hspace{1cm} (4.7)

An example of the basic operation of this algorithm, where all the data are transmitted in a single block, is shown in Figure 4.1. In Figure 4.1 we show how the $T_{ij}$ is assigned by choosing the earliest between the one chosen by Void Filling algorithm (case 1) and the one chosen by EFT algorithm (case 2). Examples of the EFT-Partial-VF MT operations, where the transmission is divided into different blocks, are shown in Figure 4.2 and Figure 4.3. Figure 4.2 shows a situation where the partial void filling algorithm is applied for both first and second transmission. Conversely, Figure 4.3 shows the case where the first transmission is scheduled through the partial void filling algorithm while the second transmission is assigned using EFT rule. In all these examples, we use $P = 2$, which means that at most a transmission will be split in two blocks of data. In ST operation, when the transmission for the ONU$_i$ is divided into different portions of data, the algorithm switches to MT operation. The number of threads that are created for a particular ONU corresponds to the number of chunks in which a single transmission is partitioned. Therefore, at most $P$ threads are assigned to this same ONU$_i$. The number of threads assigned to an ONU corresponds to the number of chunks in which its transmission is partitioned. The maximum number in which
a transmission can be partitioned \((P)\) must be chosen considering that it exists a tradeoff between dividing a transmission in a large amount of small "sub-blocks" of data (large value of \(P\)) and dividing the same transmission in a small amount of large "sub-blocks" of data (small value of \(P\)). In fact, when the "sub-blocks" are small it is easier to fill also very small voids in the channel, but at the same time this condition leads to a proliferation of control messages which increases the total amount of transmission overhead, finally leading to a poor performance of the algorithm. Conversely, by partitioning the transmission in few "sub-blocks", we limit the total amount of transmission overhead due to control messages, but we also give less opportunities to the algorithm to find a void which can be filled with a longer "sub-block" of data. However, to guarantee a certain capacity to each ONU, in EFT-Partial-VF MT, the sum in bytes of all the chunks in which a transmission can be partitioned must be at most equal to \(B_{\text{max},i}\). Switching from ST to MT operation is possible because, in EFT-Partial-VF MT algorithm, the ONU is allowed to send a REPORT message after each "sub-block" transmission. Each of these REPORT messages, when received by the OLT, initiates a new thread. In fact, for each REPORT received, the OLT sends a new GATE message to the ONU indicating when it can start the transmission for this new thread. As an example, when the \(p\)-th REPORT associated to the \(p\)-th "sub-block" \((p \in 1, 2, ..., P)\) is received, the OLT generates a new scheduling for the length of data requested by the ONU\(_{i}\). Then, the OLT sends the GATE message which starts the \(p\)-th thread of ONU\(_{i}\). When an ONU is in MT operation, its transmissions are scheduled using the same strategies of the EFT-VF algorithm described by (4.2), (4.3), (4.4), and (4.5). This means that the scheduling for each thread is done choosing the \(S_{h_{r},j}\) between the earliest \(S_{h_{r},j}\) assigned using the EFT algorithm and the \(S_{h_{r},j}\) chosen by the VF rule. During the MT operation, an eligible void is defined through the following equation:

\[
V_{i} = \{F_{h_{r},j} | S_{h_{r},j+1} - \max(F_{h_{r},j}, t_{i} + R_{i} + t_{c}) \geq t_{g_{i}} + t_{c} \}, \\
\quad h \in W_{i}
\] (4.8)

Where \(S_{h_{r},j+1} - F_{h_{r},j}\) is the time distance between the reception by the OLT of the first bit of the \((j+1)\)-th transmission on the \(h\)-th wavelength, and the reception of the last bit of the \(j\)-th transmission on that same wavelength. \(t_{i} + R_{i} + t_{c}\) is the minimum allocation time for ONU\(_{i}\), depending on the time round trip time and on the time needed to
send the gate message to the ONU. $S_{hrj+1} - (t_i + R_i + t_c)$ is the time distance between the reception at the OLT of the first bit of the $(j+1)$-th transmission on the $h$-th wavelength, and the time when the first bit of the transmission of ONU$_i$ can reach the OLT, according to its $R_i$. As a result, the requested length of data for a single thread is not divided into multiple transmissions. This feature is motivated by the fact that it is necessary to prevent an uncontrolled proliferation of threads. Moreover it is expected that, since there are more threads, each REPORT of each thread requests to transmit a smaller length of data. This happens because, with more threads, data and REPORT packets are sent more often with respect to ST operation. Then the length of data in the ONU queue decreases faster. Consequently, this feature of having smaller REPORT requests allows the algorithm to fill also small voids without the need to cut the requested length $G_i$ into multiple transmissions. When the traffic load of the ONU$_i$ decreases, the ONU$_i$ sends REPORT messages to the OLT stating that since the queue is becoming smaller, less data is required in the next cycle. Therefore, if the number of bytes required in a REPORT is zero this REPORT is ignored by the OLT. In this way, if the ONU was transmitting in $P$ threads it now uses only $P - 1$ threads. When the OLT receives $P - 1$ REPORT messages reporting zero bytes in the ONU queue, the algorithm switches again to ST operation.

### 4.4 Definition of the Cycle Time

The cycle time ($T_{cycle}$) is the time interval between successive requests from the same ONU, where every ONU transmits exactly once. For the Multi-thread algorithm the definition of $T_{cycle}$ is slightly different and is the time interval between successive requests of the same thread from the same ONU, where every ONU sends one transmission for each thread. It is important to provide a new definition for $T_{cycle}$ in order to guarantee a certain capacity to each ONU. In fact, when the duration of the maximum $T_{cycle}$ is defined, it is possible to calculate the $B_{max,i}$, which is the maximum amount of bytes that each ONU can transmit in each $T_{cycle}$. Concerning TDM PONs, a definition of $T_{cycle}$ is given in [43]. In this section we extend the definition of $T_{cycle}$ to a hybrid WDM/TDM system and we propose an equation to calculate $B_{max,i}$. In order to calculate the maximum $T_{cycle}$, we consider the network at high load, when all the ONUs are transmitting at their maximum capacity and there are no voids on the channel. We
consider the average number of ONU transmitting on a wavelength in each $T_{cycle} \cdot N/W$. We define the cycle time as the time period between two subsequent transmissions of the same ONU, where all ONUs transmit once, without regard to the wavelength they are transmitting on. According to this definition we identify an equation to calculate the value of $B_{max,i}$:

$$B_{max,i} = \left\lfloor \frac{(T_{cycle} - T \cdot (N/W) \cdot T_{guard}) \cdot \phi_i \cdot C_w}{8} \right\rfloor \quad (4.9)$$

In this equation $\phi_i$ is a weight, computed in Eqn. 4.10, which represents the fraction of the total capacity of a wavelength required by ONU$_i$. $T$ is the number of threads, where $T = 1$ corresponds to the single thread case. $T_{guard}$ is the guard band time between two subsequent transmissions and $C_w$ is the capacity of each wavelength. $(T_{cycle} - T \cdot (N/W) \cdot T_{guard})$ is the total duration of the cycle time where we subtract all the $T_{guard}$ of all the $N/W$ ONUs which are transmitting, in average, over the same wavelength. This time period is the time dedicated to the transmission of the data of the ONUs. When this amount of time is multiplied by $C_w$ we obtain the total amount of bits which can be transmitted in a cycle time. Finally, multiplying this value by $\phi_i$ we have the maximum amount of bits that ONU$_i$ can transmit according to its transmission rate $C_i$.

$$\phi_i = \frac{C_i}{\sum_{1 \leq k \leq N/W} C_k} = \frac{C_i}{C_w} \quad (4.10)$$

To compute the $\phi_i$, we must consider a constraint on the capacity of the ONUs, shown in Eqn. 4.11, which states that the sum of the capacity required by the different ONUs must be less or equal to the total capacity of the network, $C_{pon}$. The $C_{pon}$ is basically the sum of the capacity of the single wavelengths.

$$\sum_{1 \leq k \leq N} C_k \leq C_{pon} \quad (4.11)$$

In such way, the previous equations guarantee that each ONU is able to transmit at its required transmission rate.
Chapter 4. Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs

4.5 Illustrative numerical results

4.5.1 Generic Simulation Framework

To evaluate the performance of the proposed algorithms, we implemented a network simulator based on the Discrete Event Simulation Library (DESL) \[36\], modified to simulate a LR WDM/TDM PON. This simulation tool only implements the upstream transmissions. To reflect the property of the real Internet traffic, we generate self-similar traffic by aggregating multiple sub-streams, each consisting of alternating Pareto-distributed ON/OFF periods, with a Hurst Parameter of 0.8. The simulator generates Ethernet frames with a length distributed between 64 and 1518 bytes. The buffer size of each ONU has a length of 10 Mbytes. The polling cycle time is 2 ms, and accordingly the $B_{\text{max},i}$ calculated through Eqn. 4.9 is 7687 bytes for Multi-thread and 15500 bytes for the other schemes, where the $B_{\text{max},i}$ is the same for each ONU. The guard band time between two subsequent transmissions is 1 $\mu$s. Our topology includes 128 ONUs all transmitting over all the 8 wavelengths. Each channel, consisting of a single wavelength, has a bit rate of 1 Gbit/s which gives a total capacity of 8 Gbit/s. As we simulate a Long-Reach scenario, the distances from the OLT to the ONUs are uniformly distributed between 80 km and 100 km. The load offered to the network by the ONUs varies during the simulation from 0.05 to 1. At load 1 the bit rate of each ONU is 62.5 Mbit/s. For the Multi-thread algorithm, we use $T = 2$, while for EFT-Partial-VF MT we use $P = 2$ that we consider a reasonable value in which to partition a transmission in order to fill also very small voids while avoiding a proliferation of control messages.

4.5.2 Simulation Framework varying Network Parameters

In this section, we present different simulation scenarios, obtained varying network parameters, in which we evaluated the algorithms considered in our work. We consider that all the ONUs transmit over all the available wavelengths and then we consider that $W_i = W \forall i$, where $W$ is the set of wavelengths of the network. We also define $N$ to be the total number of ONUs of the LR WDM/TDM PON and $N/W$ is the average number of ONUs per wavelength. $C_i$ is capacity per ONU. We consider two different network scenarios where the parameters $W, N$ and $N/W$ have been varied as shown in Case 1 and Case 2 of Tab. 5.2, Case 3 in Tab. 5.2 summarize the parameters used...
4.5. Illustrative numerical results

to simulate a scenario where different sets of ONUs transmit at a different rate.

4.5.3 Numerical Results and Discussion

Figure 4.4: Average packet delay comparison between EFT, EFT-VF, Multi-thread over LR WDM/TDM PON, and EFT-Partial-VF MT algorithms.

Generic Simulation Framework

We first compare our results with pure EFT algorithm to see the improvement due to the implementation of the Multi-thread and to our proposed allocation scheme. Then we compare the results given by EFT-VF, Multi-thread over LR WDM/TDM PON, and EFT-Partial-VF MT.

In Figure 4.4 we plot the average packet delay versus the offered load. We can observe that the EFT-P-VF MT and the Multi-thread algorithms introduce a significant improvement regarding the packet delay, compared to EFT. Conversely the gain of the EFT-VF solution with respect to the EFT algorithm is limited (less than 5%). We can notice the same result also in Figure 4.5 where we show the percentage of average delay reduction with respect to EFT algorithm. Besides this limited gain, we observe that the computational complexity of the EFT-VF is $O(N(N+W))$ while the complexity of the EFT is
O(N * W). Since the value of W \ll N we can consider that the complexity of the EFT-VF is O(N^2) and the complexity of EFT is O(N), so the first one is linear in N while the second is exponential in N, and the difference is not relevant if N is limited. Regarding the Multi-thread over LR WDM/TDM PON, this improvement is significantly relevant for low loads whereas for medium loads it starts to decrease. Unfortunately, for high loads the average delay provided by this scheme becomes very high. This behavior is due to the thread coordination effect introduced in Section 4.2 which does not allow to efficiently exploit multiple wavelengths. Note that, for low/medium loads, the average delay in the two MT-based solutions tends to become smaller for increasing traffic. This behavior is counter-intuitive, but it can be reasonably explained considering the transmission of control messages. In fact, at low loads, the length (in byte) of the transmissions of the ONU tends to be short, and so control messages are sent very often, negatively impacting algorithm performance. Conversely, when the load increases and the length of data packets becomes higher, the number of control messages is smaller, and less time is wasted in the control phase, allowing us to take more advantage from multiple and simultaneous polling processes.

![Figure 4.5: Average delay reduction of EFT-VF, Multi-thread over LR WDM/TDM PON, and EFT-Partial-VF MT algorithms compared to EFT scheme.](image-url)
4.5. Illustrative numerical results

The EFT-VF algorithm shows an average packet delay slightly lower with respect of the EFT scheme. Such difference strictly depends on the characteristics of the access network where this allocation is applied. In fact the authors in [35] apply their algorithm in a very particular architecture, similar to SARDANA and SUCCESS, where the distances from ONUs to the OLT are very diverse. Conversely, we tested the EFT-VF scheme in a LR WDM/TDM PON where all the ONUs are at a very large distance from the OLT. The solution which gives the best performance is the EFT-Partial-VF MT proposed in this work. At low loads, this scheme provides results similar to those provided by EFT-VF algorithm. This happens because, at these loads, the lengths of data requested by ONUs are small if compared to the length of voids and, for this reason, it is possible to fill the voids on the channel without slice transmissions. When the load increases, the EFT-Partial-VF algorithm makes better use of the voids formed on the channel as it is allowed to cut transmissions in more portions. In such way it is able to fill also smaller voids. Moreover, if an ONU requires to send a huge amount of data, it can switch to MT operation achieving a better performance. In contrast, when the load of an ONU in MT operation decreases, the ONU will enter again in ST operation. Through this strategy it possible to avoid the performance degradation, observed instead when Multi-thread is applied over LR WDM/TDM PONs, due to the constant presence of control messages to maintain the different threads for each ONU. This feature is very important to avoid unnecessary control messages especially given the bursty self-similar nature of Internet traffic. Finally, we can observe that the EFT-P-VF MT algorithm provide a higher gain in the average delay with respect to the EFT-VF, while both these two schemes have a computational complexity of $O(N^2)$. Also in this case we consider that the values of $P$ and $W$ are negligible with respect to $N$.

Performance evaluation varying network parameters

In this section we evaluate the performance of the algorithms in different simulation scenarios, defined in Tab. 5.2 where different network parameters have been varied. Note that we show the performance only for our proposed EFT-P-VF MT algorithm, but the same considerations apply also for the other algorithms evaluated in this work. Moreover, all the considerations that we make in this section are valid when the $C_i$ is maintained constant for each scenario. In the first scenario, we vary the number of ONUs $N$ and we fix
the number of wavelengths $W$. According to these parameters, the value of $N/W$ changes. In this simulation scenario, we can notice that the performance given by each algorithm is the same for each value of $N$, for low and medium loads. Figure 4.6 shows that, the more the value of $N$ increases and therefore increasing $N/W$, the more the average packet delay assumes very high values at a low value of traffic load. This counter intuitive behaviour will be explained later in this section. In the second simulation scenario, we vary the number of wavelengths $W$ and we fix the number of ONUs $N$. Figure 4.7 shows that, also in this second scenario, the performance of the algorithm is the same under all configurations for low and medium loads. For high loads, when the number of wavelengths $W$ decreases and, consequently, $N/W$ increases, the delay assumes very high values at a lower traffic loads. Having noticed that the performance depends on the value of $N/W$, we also tested our algorithm in a scenario where we vary both $N$ and $W$ in order to have always the same $N/W$. We can then affirm that, the performance variation highlighted in first and second scenarios it does not depend on the values of $N$ and $W$ but on the ratio $N/W$. In fact, we noticed that when the value of $N/W$ remains the same, the algorithm has the same average delay for all.
4.5. Illustrative numerical results

loads, even if the values of $N$ and $W$ vary. In the third simulation

![Figure 4.7: Average packet delay comparison of EFT-Partial-VF MT algorithm varying the value of $W$.](image)

scenario we compare the case where different set of ONUs transmit at a different rate, called unbalanced case, with the situation where all the ONUs transmit at the same rate, the balanced case. In Fig. 4.8 are shown the results of this comparison. In the unbalanced case we notice that, even if a set of ONUs is transmitting at a high rate, the performance of the ONUs transmitting at a lower rate is not affected to much. This means that our proposed algorithm is able to manage an unbalanced scenario providing a reasonable average delay to each set of ONUs. The performance degradation noticed when the value of $N/W$ increases is due to the $T_{guard}$. We can observe from Eq. 4.9 that, given a certain $T_{cycle}$, when the number of $N/W$ increases, the impact of the guard band time in Eqn. 4.9 increases, being it multiplied by $N/W$. Therefore, the percentage of guard band time with respect to the data transmission time increases when $N/W$ increases. A solution to avoid the performance degradation which follows the increase of the value of $N/W$ is to define a longer $T_{cycle}$ for networks with a higher $N/W$. In such way, the percentage of guard band time remains constant. In fact, as the length of the $T_{guard}$ depends on technological constraints which define a minimum value of this parameter [44], it is not possible to set a lower $T_{guard}$ when
Chapter 4. Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs

$N/W$ increases. There is a limit on the length of the $T_{cycle}$. The more the $T_{cycle}$ become longer, the more the transmissions of the ONUs are delayed. Then the average cycle time increases. So, we can conclude that exists a trade off between having a small cycle time, which allow to reach small average packet delay, and having a long cycle time, that allows to have a small overhead due to the guard band time per $T_{cycle}$. Further study is needed to identify the right balance.

4.6 Multiplexing Gain Evaluation

In this section, we evaluate the performance of the EFT algorithm in network scenarios with different number of wavelengths. The network parameters used for this evaluation are presented in Table 4.3. Figure 4.9 shows the average delay of the EFT algorithm in network configurations with different number of wavelengths. From Fig. 4.9 we note that by increasing the number of wavelengths it is possible to decrease the average network delay. Moreover, we can also notice...
that the gain introduced by doubling the number of wavelengths decreases when the number of wavelengths increases. In other words, the gain provided by a network scenario with \( W = 2 \) with respect to a scenario with \( W = 1 \) is much higher than the gain achieved by a scenario with \( W = 4 \) with respect to a case with \( W = 2 \).

In Fig. 4.9 we can also notice that congestion (i.e., when the average delay measured tends to assume very high values) in the different network scenarios occurs at different loads. This is due to the time per polling cycle which cannot be used to transmit data due to the presence of guard band times which must be allocate between consecutive transmissions of different ONUs. In fact, increasing the number of wavelengths the average number of ONUs per cycle that transmit over a single wavelength decreases. Consequently, the average amount of guard band time per cycle, over each wavelength, decreases.

Based on these observations, we can conclude that while in a LR
WDM/TDM PON it is possible to improve the network performance by increasing the number of wavelengths, however it is not very convenient to add a high number of wavelengths since the improvement of the performance saturates soon. Note that this statement applies if the bandwidth assigned to each ONU ($B_i$) is dimensioned in order to saturate the total network capacity.

4.7 Novel LR WDM/TDM PON Architecture and Solutions for FMC Backhauling

Based on the previous results in our work, we discuss some guidelines for the implementation of a optical access system which can efficiently serve different types of traffic, as in the case of an FMC architecture, by using a LR WDM/TDM PON.

In our work we have studied the performance of different bandwidth and wavelength allocation algorithms, namely the EFT and the WDM-MT algorithms, in scenarios with different number of wavelengths. By doing so, we have evaluated how the network performance in term of average delay varies if a LR TDM PON is used or a hybrid WDM/TDM PON with different numbers of wavelengths is exploited. In this study we have not included the analysis of the pure WDM PON for two main reasons: i) in WDM PONs each ONU transmits on a separate wavelength and it is provided a large bandwidth. Due to the bursty nature of the traffic, each ONU is not always transmitting, and the wavelength capacity is not efficiently exploited. Therefore, to install a pure WDM acces network can be very inefficient. ii) In our work we aim at evaluating the gain provided by the statistical multiplexing which can not be considered in a WDM PON. To provide such kind of analysis we use DBWAs, which are not used in WDM PON since there is no risk of collision between transmission of different ONUs. The results provided by our simulations show that with respect to a pure TDM PON (with a single wavelength) if a WDM/TDM PON is adopted we can improve the performance of the traffic, in term of average packet delay. Moreover, we noticed that the gain obtained by introducing more wavelengths decreases when the number of wavelengths increases. Based on this observation we can state that using a hybrid WDM/TDM PON can significantly improve the performance of the network even without dedicating an entire wavelength to each ONU. Based on Fig. 4.9 we can state that it is not necessary to add a high amount of wavelengths
4.8. Conclusion

In order to improve the performance of the network. This result can lead to some interesting preliminary observations. First of all, if we are able to obtain a limited average delay allowing ONUs to transmit over a limited number of wavelengths we can design the access network dividing the entire spectrum of wavelength into subgroups in order to create a set of virtual PONs. Each virtual PON could be devoted to serve different sets of ONUs by dividing them, for example, on the basis of the types of service. In such way, the LR WDM/TDM PON could become an attractive solution to implement a FMC system, which by definition, has to serve different services with different performance requirements (e.g., average delay, maximum delay). The second reason is related to the cost of the network. The cost of a tunable transceiver depends also on the width of its tuning range (i.e., the number of wavelengths that a transceiver must be able to tune): the higher the tuning range, the higher the cost of a device \[45,46\]. So, if we are able to achieve satisfying network performance with reasonably narrow tuning range of the lasers, we could be able to install multiple LR WDM/TDM PON with a small amount of wavelengths (also called virtual PONs), which can have a limited cost, instead of a single large WDM/TDM PON requiring transceivers with wide tunability.

4.8 Conclusion

In this work we have presented, analyzed and evaluated different strategies for LR WDM/TDM PON to perform upstream scheduling and wavelength assignment.

Regarding the implementation of a pure Multi-thread polling strategy in LR WDM/TDM PONs, our results show that, while at low/medium loads it achieves large improvement with respect to EFT algorithms, for medium/high loads it suffers from a sudden increase of delay. So, in a LR WDM/TDM PON, the pure application of Multi-thread algorithm replicates the same gain achieved in LR TDM PONs with respect to single thread polling, at least at low and medium loads. Therefore, in this work we found an effective solution to combine the positive features of the EFT-VF algorithm with the Multi-thread algorithm, in order to provide a new and efficient allocation scheme for LR WDM/TDM PON. In fact, simulation results show that the EFT-partial-VF Multi-threaded algorithm provides lower average packet delay, for a wide range of traffic load, compared to the other
schemes. Simulation results underline that the performance of the algorithms depends on the average number of ONU transmitting on each wavelength in a cycle time \((N/W)\). We also investigated the effect of the overhead due to guard band time on the performance of the algorithms. We can conclude that, for a fixed value of \(T_{\text{cycle}}\), when the value of \(N/W\) increases, the overhead increases, and in turn the average packet delay becomes higher. Future works in the field of DBWA algorithm for hybrid WDM/TDM PON include the investigation of the impact of the lasers tuning time on the performance of such allocation schemes.
4.8. Conclusion

Table 4.1: Definition of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>Time when the REPORT from ONU$_i$ arrives at the OLT.</td>
</tr>
<tr>
<td>$R_i$</td>
<td>Round-trip propagation time between OLT and ONU$_i$.</td>
</tr>
<tr>
<td>$t_c$</td>
<td>Time needed for the transmission of Report or Gate frame.</td>
</tr>
<tr>
<td>$T_{irj}$</td>
<td>Start Time for the $j$-th transmission of the ONU$_i$, which is the time when an ONU can start its transmission according to its $R_i$.</td>
</tr>
<tr>
<td>$W_i$</td>
<td>Set of wavelength supported by ONU$_i$.</td>
</tr>
<tr>
<td>$S_{hrj}$</td>
<td>Arrival time at the OLT of the first bit of the $j$-th scheduled transmission on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$F_{hrj}$</td>
<td>Finish time of the $j$-th transmission on the $h$-th wavelength corresponding to the reception of the last bit of the control frames which is piggy-backed to the data packets.</td>
</tr>
<tr>
<td>$L_h$</td>
<td>Finish time of the last transmission scheduled on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$B_{max,i}$</td>
<td>Maximum length for a transmission of ONU$_i$, in each cycle time, in order to ensure fairness among the ONUs and grant the maximum capacity achievable from each of them.</td>
</tr>
<tr>
<td>$V_i$</td>
<td>Length of data, in bytes, requested by an ONU$_i$.</td>
</tr>
<tr>
<td>$t_{g,i}$</td>
<td>Time needed to transmit the length of data which ONU$_i$ asks to transmit.</td>
</tr>
<tr>
<td>$p$</td>
<td>Maximum number of portions in which a transmission can be partitioned.</td>
</tr>
<tr>
<td>$w_{eft}$</td>
<td>Upstream wavelength which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$w_{ef}$</td>
<td>Upstream wavelength which could be chosen by the Void Filling algorithm.</td>
</tr>
<tr>
<td>$n_{eft}$</td>
<td>Transmission turn within the $w_{eft}$, which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$n_{ef}$</td>
<td>Transmission turn within the $w_{ef}$, which could be chosen by the Void Filling algorithms</td>
</tr>
<tr>
<td>$w$</td>
<td>Chosen upstream wavelength.</td>
</tr>
<tr>
<td>$n$</td>
<td>Transmission turn within the $w$ wavelength.</td>
</tr>
</tbody>
</table>
Chapter 4. Dynamic Bandwidth and Wavelength Allocation Schemes for Long Reach WDM/TDM PONs

Table 4.2: Network parameters for different network scenarios

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>N</th>
<th>N/W</th>
<th>Capacity per ONU</th>
<th>Capacity per λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>8</td>
<td>128</td>
<td>64</td>
<td>62.5 Mbit/s</td>
<td>8 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>64</td>
<td>32</td>
<td>8 Gbit/s</td>
<td>4 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>64</td>
<td>32</td>
<td>128 Gbit/s</td>
<td>2 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>64</td>
<td>32</td>
<td>2 Gbit/s</td>
<td>1 Gbit/s</td>
</tr>
<tr>
<td>Case 2</td>
<td>2</td>
<td>128</td>
<td>64</td>
<td>62.5 Mbit/s</td>
<td>8 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>64</td>
<td>32</td>
<td>4 Gbit/s</td>
<td>4 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>32</td>
<td>16</td>
<td>2 Gbit/s</td>
<td>2 Gbit/s</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>16</td>
<td>8</td>
<td>1 Gbit/s</td>
<td>1 Gbit/s</td>
</tr>
<tr>
<td>Case 3</td>
<td>8</td>
<td>128</td>
<td>64</td>
<td>32 ONU@250 Mbit/s</td>
<td>8 Gbit/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>96 ONU@62.5 Mbit/s</td>
<td>1 Gbit/s</td>
</tr>
</tbody>
</table>

Table 4.3: Network parameters of different network scenarios for the evaluation of the EFT algorithm

<table>
<thead>
<tr>
<th>W</th>
<th>N</th>
<th>Capacity per ONU</th>
<th>Capacity per λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>7.8125 Mbit/s</td>
<td>1 Gbit/s</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>15.625 Mbit/s</td>
<td>2 Gbit/s</td>
</tr>
<tr>
<td>4</td>
<td>128</td>
<td>31.25 Mbit/s</td>
<td>4 Gbit/s</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>62.5 Mbit/s</td>
<td>8 Gbit/s</td>
</tr>
</tbody>
</table>
Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs

Passive Optical Networks (PONs) are continuously evolving. One important requirement of these systems is the possibility to smoothly upgrade the network infrastructure as technologies evolve [48]–[49]. During the process of network upgrade, it is expected a coexistence of new and old end–users, i.e., Optical Network Units (ONU), with diverse transceiver technologies.

In this work, we focus on a scenario where the network is upgraded from a TDM PON to a hybrid WDM/TDM PON (both short-reach and long-reach), and where the transceivers used in the old TDM PON can coexist with the new transceivers adopted in the hybrid WDM/TDM PON. Another scenario of technology coexistence over PON and long-reach PON (LR-PON) [8] is the fixed and mobile converged (FMC) access/aggregation network, where both fixed and mobile traffic are backhauled through a common network infrastructure. In this scenario, different transceivers could be used to transmit the traffic originated by mobile and fixed access points due to the different
performance requirements of the two types of services. Therefore, also in the FMC scenario different transmission technologies might need to coexist.

In our scenario, the transceivers of the ONUs differ according to: i) their capacity to retune their lasers to different wavelengths (i.e., tunable or not tunable), and ii) the time needed for tunable transceivers to retune its laser to a different wavelength. We refer to this time interval as Tuning Time (TT) of a laser. In our work, we identify three families of transceivers: i) Single fixed tuned lasers which can not retune; ii) Array of fixed tuned lasers \[50\] which can immediately tune their laser to another wavelength \((TT = 0s)\); iii) Tunable lasers \[21\] which are characterized by a given tuning time \((TT > 0s)\). In this scenario, we have to manage the traffic transmitted by different types of transceivers coexisting on the same network infrastructure, without deteriorating the performance (mainly in terms of average packet delay) of any particular technology.

The rest of the chapter is organized as follows. In Section 5.1, we detail the contributions provided by our work. In Section 5.2, we present related works on existing DBWAs. Section 5.3 introduces our proposed TT-aware DBWAs. In Section 5.3.1, we propose methods that can be used by the OLT to retrieve the exact value of the TT of each transceiver of the network. Section 5.4 presents the new DBWA that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength that they have to retune. In Section 5.5 we introduce the simulative scenario and we discuss some numerical results. In Section 5.6, we draw the main conclusions of this work.

The work presented in this chapter was published in:


5.1 Contributions

The investigation in the work presented in this chapter has multiple objectives:

1. We propose new Dynamic Bandwidth and Wavelength Allocation (DBWA) algorithms (based on the functionalities of some
5.2. Related works

existing non TT-aware DBWA) which support transceivers characterized by different values of TTs. We remark that studies of DBWAs dealing with transceivers with not negligible TTs already exist but not for the case that we are studying, i.e., DBWAs for transceivers with different TTs. The proposed DBWAs are able to jointly manage transmissions of fixed tuned lasers, array of fixed tuned lasers (i.e., with a \( TT = 0 \)), and tunable transceivers with \( TT > 0 \) (i.e., the DBWA is TT-aware).

2. We evaluate the importance, in terms of performance (i.e., average packet delay), of having DBWAs that support transceivers with different values of TTs. Therefore, as comparison term we use DBWAs that are blind to the fact that different transceivers can have different values of TT.

3. We propose some method that can be used by the OLT to obtain the exact value of the TT of each transceiver of the network, in case tunable lasers at the ONUs have different TT values (and not one single TT for all ONUs, as typically assumed).

4. We propose a new DBWA that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength they have to retune to (i.e., higher distance between wavelengths requires higher TTs). To the best of our knowledge this is the first time that such a DBWA is proposed and analyzed.

5. Finally, in all the previous evaluations we assume the number of fixed tuned lasers installed in the array is equal to the number of wavelengths in the network. However, to have a cheap device, the number of fixed tuned lasers in an array has to be limited. For this reason, we also evaluate the performance of the previously proposed TT-aware DBWAs when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths.

We compare the performance of our proposed solutions in both short-reach (up to 20 km of network span) and long-reach scenarios (up to 100 km of network span) when various types of transceiver coexist over the same network infrastructure.

5.2 Related works

In this section, we provide an overview of existing DBWAs that take into account the TT delay. In general, we can classify the DBWAs in
two classes: i) TT-aware, and ii) without TT. The TT-aware DBWAs are the algorithms which are optimized for taking advantage of the presence of the TT delay of the transceivers. The DBWAs without TT do not consider any TT. These solutions are designed for scenarios where the transceivers can immediately retune, i.e., they have a $TT = 0$.

TT-aware DBWAs have been proposed in [51, 52]. In [51] the authors proposed a scheduling policy to exploit the benefit introduced by the laser tunability for both upstream and downstream transmissions. This work focused on offline scheduling and just-in-time scheduling [31], while in our work we study and propose DBWA with online scheduling. In [52] the authors extended the Interleaved Polling with Adaptive Cycle Time (IPACT) scheme with additional TT considerations and investigated the effect of the laser TT on the network performance. However, this work does not cover Long Reach access network or mixed transmission technologies. In [53], it is presented an automatic load-balancing DBWA algorithm for WDM/TDM PONs. In this solution a tunable device (with very long TT, i.e., 10 ms or more) is tuned to a different wavelength when load-balancing is needed due to high load conditions.

To the best of our knowledge, our proposed TT-aware DBWA is the first that performs online bandwidth allocation, considers also a long reach scenario, and is designed to be applied in a scenario where the coexistence of several transmission technologies has to be managed. In [54], we firstly introduced the concept of DBWA able to support transceivers with different values of TTs. In this work, we extend the algorithms proposed in [54] considering also the case where tunable lasers have a variable TT depending on the distance of the wavelength to be retuned, and the case where array of fixed tuned lasers have a limited number of lasers (i.e., smaller than the number of wavelengths in the network).

### 5.3 TT-aware DBWAs with coexistence of transceivers

In this section, we describe the main functionalities of the proposed DBWAs, that solve the problem of allocating the transmissions originated by transceivers with different TTs.

**Background.** All the proposed DBWAs are based on the Multi Point Control Protocol (MPCP) [11]. In this work, we focus on the decision of when to schedule the transmission of ONUs with the aim
of reducing the average packet delay. We assume that the grant sizing has already been solved during a preprocessing phase where it is calculated the maximum amount of data that can be transmitted by each ONU in each cycle time in order to guarantee certain bandwidth to each ONU. In particular, we use as a grant sizing policy the limited sizing \[55\].

Without loss of generality, we consider here three different families of transceivers: i) Single fixed tuned lasers which can not retune; ii) Array of fixed tuned lasers \[50\] which can immediately tune their laser to another wavelength \((TT = 0s)\); iii) Tunable lasers \[21\] which are characterized by a given tuning time \((TT > 0s)\). We refer to \(N_1\) as the set of ONUs that use tunable lasers to transmit and therefore they can tune to any wavelength in the network with a \(TT > 0\). \(N_2\) is the set of ONUs using array of lasers, which can transmit over each wavelength in the network with a \(TT = 0\). Finally, \(N_3\) are the ONUs that use a single fixed tuned laser, and so they can only transmit over their nominal wavelength. The definition of these ONU groups is presented in Table 5.1.

Our algorithms are developed extending the Earliest Finish Time (EFT) and the Earliest Finish Time with Void Filling (EFT-VF) \[35\], and introduced in chapter 3.

**Proposed TT-aware DBWAs.** Since we apply all the studied DBWAs in a scenario where different technologies have different tuning times we first modify the previous algorithms to take into account the laser TT delay. Particularly, for the ONUs in \(N_1\), the modified algorithms calculates the interval between the instant when the Gate message is received at the ONU and the start time of the transmission at the same ONU. During this interval the ONU can retune its laser, if needed. If this interval is shorter than the laser TT, the algorithm adds an additional delay \((\delta)\). In particular, \(\delta\) is added if \(t - t_{rx} < TT\) where \(t\) is the start time computed with the algorithm (EFT or EFT-VF), and the value assigned to \(\delta\) is \(\delta = TT - (t - t_{rx})\). Therefore, in this case the EFT algorithm adds the computed \(\delta\) to the start time computed in Eqn. 4.3 and the EFT-VF algorithm adds \(\delta\) to the start time computed in Eqn. 4.6. For the ONUs in \(N_2\) the original versions of EFT and EFT-VF are directly applied. ONUs in \(N_3\) transmit in the first available start time (considering also voids when it is applied the EFT-VF algorithm) over a fixed pre-assigned wavelength (and, since TT=0, there is no need to add TT-related delays). These new algorithms are referred here as \(EFT+TT\) and \(EFT-VF+TT\), where TT indicates that these DBWAs take
Chapter 5. Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs

Figure 5.1: Example of the transmission allocation of an ONU in N1 where a retuning happens and there is a need to add delay to retune the laser.

\[ W_j: \text{Wavelength}; \quad V_i: \text{Eligible voids for ONU}_i; \quad T_x_{t+1}: \text{ONU}_i \text{ transmission in } (t)\text{-th cycle} \]
\[ t_{rx}: \text{Time when the Gate message is received by ONU}_i; \quad TT_i: \text{Tuning Time of ONU}_i \text{ laser} \]
\[ t_{cur}: \text{Time when the OLT decides the scheduling for ONU}_i \quad Rtt_i: \text{RTT of ONU}_i \]
\[ St_{t+TT}: \text{Minimum possible StartTime considering the TT of laser for ONU}_i \]
\[ St^{eft}: \text{StartTime chosen with EFT algorithm for ONU}_i; \quad t_c: \text{Time to transmit gate message} \]
\[ St^{eft+d}: \text{StartTime actually allocated with EFT algorithm to ONU}_i \text{ where a delay is added to allow the laser to retune.} \]

The allocation procedure for an ONU in N1 is shown in Fig. 5.1. In Fig. 5.1, the first available wavelength to allocate a transmission is \( W_3 \), then ONU needs to retune its laser. In this case the algorithm calculates the interval between the instant when the Gate message is received at the ONU (\( t_{rx} \)) and the start time of the transmission (\( t_{eft} \)), and then it adds a delay in order to give time to the laser of ONU to retune. Therefore, the transmission start time will be delayed (\( t_{eft} + \delta \)).

Simplified version of TT-aware DBWAs. For the sake of comparison, we considered also simplified versions of EFT and EFT-VF where the OLT does not know the exact value of the TT of each ONU, but it only knows if a particular transceiver is able to retune or not, referred as Simple EFT+TT and Simple EFT-VF+TT. We implemented these versions to evaluate the importance of having a DBWA which is aware of the exact values of TT and is able to add different delays to different transmissions (i.e., like EFT+TT and EFT-VF+TT do). For this reason, Simple EFT+TT and Simple EFT-VF+TT add a delay to all the transmissions of the ONUs which have to retune (i.e., N1 and N2) considering a maximum value of TT (the same for all the ONUs).

We expect that both the EFT+TT and the EFT-VF+TT provide a lower average packet delay with respect to the simple versions of the same
5.4. TT-aware DBWAs with Wavelength Dependent Tuning Times

algorithms. Our aim is to evaluate the gain in terms of average packet delay using EFT+TT or EFT-VF+TT instead of Simple EFT+TT or Simple EFT-VF+TT, and then understand the importance of a DBWA that is aware of the type of transceivers used by each ONU.

5.3.1 Discovering Methods of the TT values

For the previous algorithms, except the Simplified versions, we assume that OLT knows the exact value of the TT of all the transceivers in the network. Three methods can be implemented to inform the OLT about the TT values: i) Through an initial setting provided by the network operator; ii) With a calculation during the ranging phase, when the round-trip-time between the ONUs and the OLTS can be calculated using the packet delay of the ONUs. Once the round-trip-time is estimated, the OLT can run the ranging mechanism another time over a different wavelength, and then it will be able to compute also the value of the TT of each ONU. iii) Each TT value can be transmitted within the Report message which is sent by each ONU.

5.3.2 Array of Fixed tuned lasers with limited tunability

We also evaluate the performance of some of the previously studied DBWAs (i.e., EFT+TT and EFT-VF+TT) when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths. In fact, in all the previous evaluation we assumed that the array of tunable lasers are able to tune to every wavelength of the network, i.e., the number of lasers in an array is equal to the number of wavelengths. However, we know that to have a cheap device the number of fixed tuned lasers in the array has to be limited.

In this case, the algorithm allocates the transmissions of the ONUs through Eqns.\[4.2-4.6\] where for Eqn.\[4.2\] \(W_i\) is not equal to the entire set of wavelengths of the network.

5.4 TT-aware DBWAs with Wavelength Dependent Tuning Times

In the previous sections, we considered that the tunable lasers of ONUs in \(N_1\) can retune from a wavelength to each other wavelength with the same TT. In this section we introduce our proposed TT-aware DBWAs for the case when tunable lasers have different TTs according to the wavelength they have to retune (i.e., the higher the spectral
Chapter 5. Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs

distance between two wavelengths, the higher the TT). In this scenario, the lasers of the ONUs in N1 have a different value of TT according to the spectral distance between the wavelength \( i \) where they currently transmit and the wavelength \( j \) they have to retune to. We refer to this distance as wavelength gap, \( w_{ij} \). The wavelength gap is computed through Eqn. 5.1 as the difference in terms of number of wavelengths between the wavelength where an ONU is currently transmitting (\( w_i \)) and the wavelength the ONU has to retune to transmit (\( w_j \)).

\[
w_{ij} = |w_i - w_j|, \quad i \neq j
\]  

(5.1)

The TT of ONUs lasers in N1 is proportional to the wavelength gap, and it is computed through Eqn. 5.2 where \( TT_{\text{base}} \) is the TT needed to retune between two adjacent wavelengths (i.e., when \( w_{ij} = 1 \)). \( TT_{ij} \) is the time needed to retune from a wavelength to any other wavelength in the network. This \( TT_{ij} \) is proportional to \( w_{ij} \).

\[
TT_{ij} = w_{ij} \times TT_{\text{base}}, \quad i \neq j
\]  

(5.2)

We first tested the EFT-VF+TT algorithm proposed in section 5.3 in this scenario where the tunable lasers of ONUs in N1 have different TTs according to the number of wavelengths they have to retune. In this scenario, the EFT-VF+TT algorithm has the same functionalities explained in section 5.3 but for the ONUs in N1 the considered TT is computed through Eqn. 5.2. We refer to this variation of the EFT-VF+TT algorithm as EFT-VF+\( TT_{ij} \), where "+\( TT_{ij} \)" refers to the fact that the TT of the tunable lasers changes according to the distance between two wavelengths. Then, we propose a new DBWA technique specifically designed for this scenario where the tunable lasers of ONUs in N1 have different TTs according to the number of wavelengths they have to retune.

We refer to this new proposed DBWA as Earliest start-time with Void Filling +\( TT_{ij} \) (EsT-VF+\( TT_{ij} \)), where "+\( TT_{ij} \)" refers to the fact that the TT of the tunable lasers changes according to the distance between two wavelengths. Also the new proposed DBWA is based on the functionalities of EFT-VF but the policy used to allocate the transmissions of ONUs in N1 is designed to minimize the effect of the different \( TT_{ij} \) on the packet delay. For ONUs in N2 and N3 this algorithm use the same allocation procedure defined for EFT-VF+TT. For the ONUs in N1, this algorithm first computes for each wavelength and for each void the start time, as defined in Eqns. 4.3.
and [4.5], adding when it is needed the delay $(\delta_{ij})$ to allow the lasers to retune. In particular, if $t - t_{rx} < TT_{ij}$, the $\delta_{ij}$ is computed according to Eqn. 5.3.

$$\delta_{ij} = TT_{ij} - (t - t_{rx}) \quad (5.3)$$

When all the possible transmission start times are computed, the Est-VF+$TT_{ij}$ chooses the minimum start time and it assigns this start time and the corresponding wavelength to the ONU. This strategy reduce the average delay introduced by the ONUs in $N1$, but it also indirectly reduce $i)$ the amount of retunings of the tunable lasers, and $ii)$ the average wavelength gap (i.e., the average distance between two wavelengths between which a laser has to retune). This algorithm differs from the EFT-VF+$TT_{ij}$ because the EFT-VF+$TT_{ij}$ first chooses the minimum start time and then it adds, when needed, the delay to allow the lasers to retune.

## 5.5 Results and discussion

In this section, we compare the performance in term of average delay of all the proposed DBWAs. In particular, in section 5.5.1 we present the simulative scenarios in which we tested our algorithms. In section 5.5.2 we present the comparison between the proposed TT-aware DBWAs and their simple version. In section 5.5.3 we evaluate the performance of EFT+$TT$ and EFT-VF+$TT$ when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths. Finally, in section we present the results of our proposed EsT-VF+$TT_{ij}$ compared with EFT-VF+$TT_{ij}$.

### 5.5.1 Simulative Scenario

The proposed algorithms are evaluated and compared through simulation. For this purpose, we implemented the DBWA algorithms in a network simulator based on the Discrete Event Simulation Library (DESL) [36], which we modified to model a WDM/TDM PON (both short-reach and long-reach). This simulation tool only implements the upstream transmissions. To reflect the property of the real Internet traffic, we generate self-similar traffic by aggregating multiple sub-streams, each consisting of alternating Pareto-distributed ON/OFF periods, with a Hurst parameter of 0.8. The simulator generates Ethernet frames with a length distributed between 64 and 1518 bytes.
The buffer size of each ONU is set to 10 Mbytes. The maximum polling cycle time is 2 ms, and accordingly the $B_{\text{max},i}$ is calculated with the same equation that we previously proposed in [47]. We assume the same $B_{\text{max},i}$ for every ONU. The guard band time between two subsequent transmissions is 1 $\mu$s. The TT value for ONUs in $N_1$ is 100 $\mu$s. Our tree topology includes 24 wavelengths (W) and 192 ONUs divided into three groups, depending on their transmission technology. We tested all the DBWAs in three scenarios where we changed the capacity of the wavelengths and the distribution of the transmission technology (i.e., number of ONUs using a particular transmission technology in each group). Then, for each scenario we varied the OLT-ONU distance distribution. The parameters of the 9 scenarios are shown in Table 5.2 where $C_w$ refers to the capacity of the wavelengths. Note that we assume that the network is properly designed to support increased losses due to high splitting ratios and
5.5. Results and discussion

5.5.2 TT-aware DBWAs vs Simple DBWAs

In this section, we compare EFT+TT, EFT-VF+TT, Simple EFT+TT, and Simple EFT-VF+TT algorithms in Scenario 1 where three different transmission technologies operate in the PON, and the OLT-ONUs distances distribution is between 500 m and 100 km. Note that, even if the EFT+TT and the EFT-VF+TT algorithms are designed to manage several different values of TT of the transceivers, in this simulations we test the DBWAs using only two different values of TT: TT = 0 and TT = 100 µs.

Figure 5.2 reports the comparison in the average packet delay of all ONUs in the PON vs. offered load among EFT+TT, EFT-VF+TT, Simple EFT+TT, and Simple EFT-VF+TT algorithms, when the TT of the tunable lasers is 100 µs. We can state that, with our settings,
Chapter 5. Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs

Figure 5.4: Average delay reduction of the Simple EFT-VF+TT and the EFT-VF+TT algorithms compared to EFT+TT scheme, where the ONUs in N1 have a TT = 100 µs, in Scenario 3.

the coexistence of different transmission technologies over the same WDM/TDM PON is possible since all the evaluated DBWAs can still provide a low average delay (less than 1.5 ms), except for very high load (over 80%-90%), as shown in Fig. 5.2. In Fig. 5.2 we notice that Simple EFT+TT has a higher average delay with respect to EFT+TT, and Simple EFT-VF+TT has a higher average delay compared to EFT-VF+TT. This is due to the fact that in the Simple EFT+TT a delay is added to all the ONUs that have to retune while in EFT+TT the retuning delay is added only for the ONUs that really have a TT > 0. We can also notice that the Simple feature has a higher impact in the case of the EFT-VF+TT algorithm than in the case of EFT+TT. Since the EFT-VF+TT has a better usage of the voids compared to Simple EFT-VF+TT, the gap between the average delay of EFT-VF+TT and Simple EFT-VF+TT is higher compared to the gap between EFT+TT and Simple EFT+TT.
5.5. Results and discussion

In the following, we compare the relative gain of EFT-VF+TT and Simple EFT-VF+TT respect to EFT+TT. In Figs. 5.3 and 5.4 we show the average delay reduction of EFT-VF+TT and Simple EFT-VF+TT with respect to EFT+TT in different scenarios where we vary: \(i\) the OLT-ONUs distances (Fig. 5.3), \(ii\) the percentage of tunable transceivers within the network (Fig. 5.3 vs Fig. 5.4). Particularly, in Fig. 5.3 we show the results obtained in Scenario1 and in Fig. 5.4 we report the results obtained in Scenario 3. We can notice that EFT-VF+TT always reduces the average delay compared to Simple EFT-VF+TT, in all the investigated scenarios. Moreover, when the load increases, the difference between Simple EFT-VF+TT and EFT-VF+TT increases, since the average delay reduction of Simple EFT-VF+TT decreases. This is due to the fact that at high loads the number of voids in the network decreases and the behavior of Simple EFT-VF+TT becomes closer to the behavior of EFT+TT. This is true for all the scenarios, but is much more evident in the short-reach cases (i.e., OLT-ONUs distances between 1-20 km), as shown in Figs. 5.3 and 5.4 where the distance between the curves of EFT-VF+TT (1-20 km) and Simple EFT-VF+TT (1-20 km) is much higher than in all the others cases.

Variation of the OLT-ONUs distances. The curves of EFT-VF+TT (80-100 km) and Simple EFT-VF+TT (80-100 km) in Figs. 5.3 and 5.4 refer to a long reach scenario and show that the average delay reduction of both Simple EFT-VF+TT and EFT-VF+TT with respect to EFT+TT is very limited (less than 5%). This is due to the fact that in a long-reach scenario, where there is low distribution of the distances between OLT and ONUs (i.e., between 80-100 km), the number of voids decreases and then the performance of EFT+TT and EFT-VF+TT (both Simple and not Simple versions) are very close. Comparing the short-reach cases with the long-reach cases (i.e., OLT-ONUs distances between 80-100 km and 500 m-100 km) shown in Figs. 5.3 and 5.4 we notice that the difference in the average delay reduction between Simple EFT-VF+TT and EFT-VF+TT is higher in the short-reach cases. Therefore, we can state that in short-reach scenarios is more important to have a DBWA which is able to manage the TT in a smart manner.

Variation of the percentage of transceivers. Comparing Fig. 5.3 with Fig. 5.4 we can notice that in the case where the percentage of tunable transceivers increases (as in Scenario 3) the difference between the performance of Simple EFT-VF+TT and EFT-VF+TT slightly increases with respect to Scenario 1 (i.e., the transmission technologies are
uniformly distributed). In Scenario 3, the average delay reduction of Simple EFT-VF+TT is lower than in Scenario 1. This happens because in Scenario 1 (Fig. 5.3), in Simple EFT-VF+TT, $2/3$ of the transmissions are not delayed (i.e., from ONUs in N2 and N3) while in Scenario 3 (Fig. 5.4) only $1/3$ of the transmissions are not delayed (i.e., from ONUs in N2 and N3).

*Variation of the wavelength capacity.* We also evaluated the differences in terms of average delay reduction between Simple EFT-VF+TT and EFT-VF+TT when the wavelengths have different capacities. The graph for Scenario 2 is not reported here since we noticed that a different capacity does not affect significantly the performance of the evaluated DBWAs. In the case where the capacity is 10 Gbit/s the average delay reduction of both Simple EFT-VF+TT and EFT-VF+TT are slightly higher (around 5% more) than in the case where the capacity of each wavelength is 1 Gbit/s.

### 5.5.3 Average delay evaluation: Array of fixed tuned lasers with limited tunability

In this section, we evaluate the performance of some the previously studied DBWAs (i.e., EFT+TT and EFT-VF+TT) when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths, as explained in section 5.3.2. The simulation scenario used for this evaluation is Scenario 1c defined in Tab. 5.2. In this scenario, each ONU in N2 can transmit over a limited set of contiguous wavelengths. These sets of wavelengths are assigned in a round robin manner to the ONUs in N2, this means that these ONUs do not transmit over the same limited number of wavelengths. For example, if in a network we have 3 ONUs with arrays of fixed tuned lasers and 4 wavelengths, and each ONUs can transmit only over 3 wavelengths. The sets of wavelengths are assigned as follows: ONU1 can transmit over wavelengths 1, 2, and 3. ONU2 can transmit over wavelengths 2, 3, and 4. ONU3 can transmit over wavelengths 3, 4, and 1. In this scenario, all the ONUs in N2 are equipped with the same arrays of fixed tuned lasers. In particular, we test three cases where the array of fixed tuned lasers are composed by 4, 8, or 12 lasers. We compare these three case with the case where the number of lasers in the arrays of fixed tuned lasers is the same of the number of available wavelengths in the network (i.e., 24 wavelengths), we refer to this case as full tunability.
5.5. Results and discussion

Figure 5.5: Average delay of the EFT+TT algorithm when the ONUs in N2 have different limited tunability, compared to the case where the ONUs in N2 have full tunability. The ONUs in N1 have a TT = 100 µs.

Figure 5.5 shows the average packet delay of the EFT+TT algorithm when the ONUs in N2 have different limited tunability (i.e., they can use only to 4, 8, or 12 wavelengths since the array of fixed tuned lasers are composed only of 4, 8, or 12 lasers), compared to the case where the ONUs in N2 have full tunability (i.e., they can tune to every wavelength of the network since these ONUs are equipped with large arrays of lasers). We can notice that the average delay of EFT+TT in the full tunability case and where the arrays of lasers can use only a limited number wavelengths has a very small variation, which is almost negligible. Lowering the number of available wavelengths for each array of fixed tuned lasers slightly increases the average delay, if compared with the case where the arrays of lasers can use all the
Figure 5.6: Average delay of the EFT-VF+TT when the ONUs in $N_2$ have different limited tunability, compared to the case where the ONUs in $N_2$ have full tunability. The ONUs in $N_1$ have a $TT = 100$ µs.

wavelengths in the networks. However, this difference in terms of average packet delay is very limited, i.e., less than 2 µs.

Figure 5.6 shows the average packet delay of the EFT-VF+TT algorithm when the ONUs in $N_2$ have different limited tunability, compared to the case where the ONUs in $N_2$ have full tunability. Similar considerations drawn for the EFT+TT algorithm apply in the case of EFT-VF+TT. With EFT-VF+TT the difference in terms of average delay is a little larger than in the EFT+TT case, but however it is very limited.
5.5. Results and discussion

Figure 5.7: Average delay of the EsT-VF+TT\textsubscript{ij} algorithm, compared to the EFT-VF+TT algorithm applied in the Scenario 1 where tunable lasers have different TTs according to the number of wavelengths they have to retune. The ONU\'s in N\textsubscript{1} have a TT\textsubscript{base} = 100 µs.

5.5.4 TT-aware DBWAs with Different Wavelength TTs

Figure 5.7 shows the results in the scenario where tunable lasers have different TTs according to the number of wavelengths they have to retune. To test this two algorithms we used Scenario 1 defined in Tab. 5.2 where the OLT-ONUs distances distribution is between 500 m and 100 km. We use a value of TT\textsubscript{base} of 100 1µs. In Fig. 5.7 we compare the EFT-VF+TT\textsubscript{ij} and the EsT-VF+TT\textsubscript{ij} in terms of average delay. First, we can notice that the EsT-VF+TT\textsubscript{ij} algorithm has the lowest average delay. Moreover, the average delay provided is lower than the average delay usually required for an access network, i.e., 1.5ms. Indeed, the EsT-VF+TT\textsubscript{ij} is specifically designed for the case when tunable lasers have different TTs according to the number of wavelengths they have to retune, and we can prove that this allocation technique, minimizing the transmission start time of the ONU\'s in N1,
Chapter 5. Dynamic Bandwidth and Wavelength Allocation with Coexisting Transceiver Technology in WDM/TDM PONs

is able to reduce the overall average packet delay of the network. In Fig. 5.7, we can notice that the average delay of the EFT-VF+TT$_{ij}$ is much higher than the corresponding version of the algorithm (i.e., EFT-VF+TT) shown in Fig. 5.2, tested in the same scenario.

5.6 Conclusion

In this work we proposed DBWA algorithms that solve the problem of managing the traffic originated by transceivers with different characteristics (e.g., different tuning time). Starting from two existing algorithms (i.e., EFT and EFT-VF) we proposed two new DBWAs that take into account the laser TT considering that different transceivers at the ONU can have different values of TT. These DBWAs are able to manage at the same time transmissions of transceivers which cannot retune (e.g., fixed tuned lasers), transceivers with a $TT = 0$ (e.g., array of fixed tuned lasers), and transceivers with $TT > 0$ (e.g., tunable lasers). We refer to these solutions as EFT+TT and EFT-VF+TT. We compared the performance of EFT+TT, EFT-VF+TT, and two simplified versions that are not aware that ONUs can have different TT values. Our results show that the EFT-VF+TT is the solution, among the three considered, that provides the lower average packet delay. Results show also that in the case of short-reach scenarios (about 20 km) it is more important to use a DBWA that is aware of the different values of $TT$ in the network, while in a long reach-scenario (up to 100 km) Simple EFT-VF+TT can be used without having a significant loss in terms of average packet delay.

Then we proposed a new DBWA that we called EsT-VF+TT$_{ij}$ that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength that they have to retune to. Our results show that the proposed TT-aware DBWA is able to reduce the overall average packet delay of the network, compared to the case where a DBWA that is not specifically designed for this scenario is applied.

Finally, we considered the case where the arrays of fixed tuned lasers are equipped with a number of lasers which is lower than the total number of wavelengths available in the network. According to our results, we can conclude that lowering the number of lasers in the arrays of lasers of ONUs negligibly increases the average packet delay, this difference in terms of average packet delay is very limited (i.e., less than 2 µs).
Table 5.1: Definition of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>Time when the Report from $ONU_i$ arrives at the OLT.</td>
</tr>
<tr>
<td>$t_{guard}$</td>
<td>Guard band time between two subsequent transmissions of different ONUs.</td>
</tr>
<tr>
<td>$t_{rx}$</td>
<td>Time when the Gate message is received at the ONU.</td>
</tr>
<tr>
<td>$R_i$</td>
<td>Round-trip propagation time between OLT and $ONU_i$.</td>
</tr>
<tr>
<td>$t_c$</td>
<td>Time needed for the transmission of Report or Gate frame.</td>
</tr>
<tr>
<td>$W_i$</td>
<td>Set of wavelength supported by $ONU_i$.</td>
</tr>
<tr>
<td>$S_{hj}$</td>
<td>Arrival time at the OLT of the first bit of the $j$-th scheduled transmission on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$F_{hj}$</td>
<td>Finish time of the $j$-th transmission on the $h$-th wavelength corresponding to the reception of the last bit of the control frames which is piggy-backed to the data packets.</td>
</tr>
<tr>
<td>$L_h$</td>
<td>Finish time of the last transmission scheduled on the $h$-th wavelength.</td>
</tr>
<tr>
<td>$B_{max,i}$</td>
<td>Maximum length for a transmission of $ONU_i$, in each cycle time, in order to ensure fairness among the ONUs and grant the maximum capacity achievable from each of them.</td>
</tr>
<tr>
<td>$V_i$</td>
<td>Set of eligible voids for $ONU_i$ calculated according to its $R_i$ and its requested length of data, which must be granted.</td>
</tr>
<tr>
<td>$t_{g,i}$</td>
<td>Time needed to transmit the length of data which $ONU_i$ asks to transmit, can be partitioned.</td>
</tr>
<tr>
<td>$w_{eft}$</td>
<td>Upstream wavelength which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$w_{vf}$</td>
<td>Upstream wavelength which could be chosen by the Void Filling algorithm.</td>
</tr>
<tr>
<td>$t_{eft}$</td>
<td>Transmission start time within the $w_{eft}$ which could be chosen by the EFT algorithm.</td>
</tr>
<tr>
<td>$t_{vf}$</td>
<td>Transmission start time within the $w_{vf}$ wavelengths, which could be chosen by the Void Filling algorithms.</td>
</tr>
<tr>
<td>$w$</td>
<td>Chosen upstream wavelength.</td>
</tr>
<tr>
<td>$t$</td>
<td>Transmission start time within the $w$ wavelength.</td>
</tr>
<tr>
<td>$N_1$</td>
<td>Set of ONUs that transmit using tunable lasers.</td>
</tr>
<tr>
<td>$N_2$</td>
<td>Set of ONUs that transmit using array of fixed-tuned lasers.</td>
</tr>
<tr>
<td>$N_3$</td>
<td>Set of ONUs that transmit using fixed-tuned lasers.</td>
</tr>
<tr>
<td>$TT$</td>
<td>Tuning Time of a tunable transmitter, which is the time needed to retune its laser to a different wavelength.</td>
</tr>
</tbody>
</table>
### Table 5.2: Network parameters for different network scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>C_w</th>
<th>Transmission technologies distribution</th>
<th>OLT-ONUs distances distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 Gbit/s</td>
<td>N1=1/3 N2=1/3 N3=1/3</td>
<td>a) 1-20km b) 80-100km c) 500m-100km</td>
</tr>
<tr>
<td>2</td>
<td>10 Gbit/s</td>
<td>N1=1/3 N2=1/3 N3=1/3</td>
<td>a) 1-20km b) 80-100km c) 500m-100km</td>
</tr>
<tr>
<td>3</td>
<td>1 Gbit/s</td>
<td>N1=2/3 N2=1/6 N3=1/6</td>
<td>a) 1-20km b) 80-100km c) 500m-100km</td>
</tr>
</tbody>
</table>
Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

TWDM PON is a promising candidate to become the backhauling infrastructure of Fixed Mobile Converged (FMC) access-aggregation networks. One of the main features of a FMC network is that different types of traffic (that support, e.g., fixed and mobile network services) are transported over the same infrastructure, and the coexistence of different traffic types poses complex challenges in terms of capacity assignment as well as traffic management. If TWDM is the technology of choice, one question naturally raises: is it better to isolate different types of traffic over different wavelengths (i.e., Layer-1 convergence) or to share wavelength capacities among the different traffic types (i.e., Layer-2 convergence)? In the first case, to isolate the different types of traffic, we can assign logically-separated Virtual PONs (VPONs) to different traffic types, that will operate on disjoint sets of wavelengths. In the second case, different traffic types can share the same wavelengths in the network (i.e., a single convergent TWDM PON is used).

In this chapter we consider the problem of assigning separately
Chapter 6. Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

a wavelength or a subset of wavelengths to different types of traffic, in order to keep them separated. In such a way, several logically-separated Virtual PONs (VPONs) can be created over the same physical network infrastructure in such a way that each VPON transports a different type of traffic. To perform such investigation, we model the capacity assignment problem considering some basic variations of the traditional bin packing problem [58] and we explore the tradeoff between the case of isolated traffic types over separated wavelengths (referred as VPON case) and the case where the traffics are mixed over all the available wavelengths in the network (referred as TWDM case) under different scenarios in terms of traffic loads, user distribution, and network capacity. In particular, in this chapter, we want to address the problems that might arise when different types of traffic are multiplexed over the same wavelengths of a TWDM PON, such as: i) a single, converged, layer-2 protocol must be used to transport different types of traffic (e.g., fixed traffic from residential users, fixed traffic from business users, mobile 3G traffic, mobile 4G traffic, ...) which might have very heterogeneous requirements (e.g., in terms of bandwidth, QoS, latency or synchronization); ii) the OLT will be required to look up each received frame in order to forward it towards the appropriate destination (as different types of traffic, e.g., 3G, 4G, fixed traffic, etc. might have to be forwarded towards different core nodes); iii) tunable lasers with a large tunability range might be needed, which are still very expensive [21].

The rest of the chapter is organized as follows. In Section 6.1, we provide a quick review of previous works on the concept of VPONs. In Section 6.2, we introduce the VPON assignment in FMC networks problem and present the optimization model that we use to solve it. In Section 6.3, we introduce the network-optimization scenarios and we discuss our results on them. Section 6.4 concludes our work.

The work presented in this chapter was published in:


6.1 Related Works

In this section, we provide a quick overview of previous works that have investigated the VPON concept.
6.2. VPON assignment in FMC networks: Problem Definition and Modeling

In [56], authors propose a novel hybrid WDM/TDM-PON architecture that utilizes VPONs to increase the bandwidth of individual ONUs in a dynamic fashion by providing a varying number of wavelength channels to each ONU. When some ONUs in the network request a higher bandwidth, these ONUs can dynamically join a new TDM-PON with more available bandwidth resources. These ONUs form a new VPON, and they may not even be in the same physical PON (i.e., they may not be physically connected to the same splitter).

In [57], the same authors enhanced their previously proposed scheme to provide also highly flexible wavelength/bandwidth assignment capability. In [58], VPONs are used to promote energy savings. Authors argue that, in off-peak hours, it would be more efficient to group ONUs into a smaller number of wavelengths, and to switch off some linecards at the OLT. The trade-off between minimizing the migration of traffic from one VPON to another, and minimizing the overall energy consumption is investigated.

Also, regarding the optimization model used in our study, similar models used for PON capacity assignment already exist, but they are applied to different networking problems [60–62].

In our study we use VPONs to separate different types of traffic in FMC network, while still trying to minimize the number of wavelengths in the network, which is assumed to be related to the network deployment cost (e.g., number of linecards in the OLT), and its power consumption. Moreover, our work aims at evaluating the trade-off between the deployment of a pure TWDM PON and the deployment of VPON-based network (i.e., different types of traffic are separated over different sets of wavelengths).

In the following section, we introduce the problem of VPON assignment in FMC networks.

6.2 VPON assignment in FMC networks: Problem Definition and Modeling

A VPON is composed by a wavelength or a set of wavelengths over which only a subset of ONUs of the network transmit. In our study, VPONs aggregate the traffic of ONUs belonging to the same technology (e.g., ONU supporting fixed vs. mobile traffic). So, in the same physical access-aggregation infrastructure several VPONs can coexist. In our work, we want to investigate the trade-off between i) using different VPONs to separate the different types of traffic, and
Figure 6.1 shows an example of the proposed VPON-based architecture in the case of ring-and-spur access-aggregation network topology. A ring-and-spur topology consists of a ring of fibers connecting several nodes, called Remote Nodes (RN) to a main central office where the OLT is located. The RNs (e.g., OADMs, ROADMs) are also connected to a set of ONUs and they are responsible for dropping the wavelengths that circulate on the ring towards these ONUs. Each RN together with its associated ONUs forms a, typically tree-based, sub-network called spur. In our scenario, an ONU can support either 2G/3G users, or 4G users, or fixed users (note that each ONU serves only one traffic type). For each spur, there are several ONUs connected that might serve all the three types of traffic, i.e., 2G/3G traffic, 4G traffic, and fixed traffic. In Fig. 6.1, each type of traffic is transmitted over a separate wavelength, and at the OLT the corresponding wavelengths are directly routed towards the core nodes designated for each specific type of traffic.

The VPON assignment problem in FMC networks consists in assigning capacity (i.e., one or more wavelengths) to each ONU such that ONUs attaining to different types of traffic do not share the same wavelength capacity. While the problem of capacity assignment in a convergent TWDM-PON (i.e., where all the types of traffic are mixed together) has been already modeled using the well-known bin-packing problem [58], capacity assignment for our proposed VPON case requires some changes to the original optimization model. The models used in our work are presented in the following section.

6.2.1 VPON Assignment in FMC networks: the ILP Model

In this subsection, we introduce the Integer Linear Programming (ILP) formulation for the VPON assignment problem in FMC networks. The considered network scenario is similar to the one showed in Fig. 6.1.

In Table 7.1, we define the notation used hereafter.

The following Eqns. 7.1-6.5 describe the MILP model that we use for VPON assignment in an FMC network. More precisely, the following model has to be applied separately to assign the VPON capacity for each type of traffic (e.g., two separate optimization instances shall be used for fixed and mobile traffic). Each VPON is separated and independent from the others, therefore we can
6.2. VPON assignment in FMC networks: Problem Definition and Modeling

Figure 6.1: Example of the VPON scenario in a ring-and-spur metro-access network.

optimize each VPON separately. It follows that the objective function of our problem will be given the sum of the objective functions of the separated sub-problems that we use to assign each VPON.

\[
\sum_{i \in I_q} x_{ij} \cdot c_i \leq y_j \cdot w_j \quad \forall j \in W \quad (6.1)
\]

\[
\sum_{j \in W} x_{ij} = 1 \quad \forall i \in N \quad (6.2)
\]

\[
\sum_{i \in G_s} x_{ij} \leq |G_s| h_{js} \quad \forall j \in W \quad (6.3)
\]

\[
\sum_{j \in W} h_{js} \leq z_s \quad \forall s \in S \quad (6.4)
\]

\[
\sum_{j \in W} y_j \leq z \quad (6.5)
\]

Equation 6.1 assigns to each wavelength the capacity required by a set of ONUs belonging to the same technology up to the total capacity of the wavelength. Equation 6.2 forces each ONU to transmit over a single wavelength (so traffic of an ONU cannot be split over multiple wavelengths).

Equations 6.3 and 6.4 are used to evaluate the number

\footnote{Note that in this model the ONU transmitters are assumed to be fixed-tuned, but with simple modification also the case of tunable transmitters can be covered}
Chapter 6. Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

Table 6.1: Definition of parameters and variables

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>number of ONUs in the network, ( i \in N ).</td>
</tr>
<tr>
<td>( S )</td>
<td>number of spurs connected to the ring, ( s \in S ).</td>
</tr>
<tr>
<td>( W )</td>
<td>number of available wavelengths, ( j \in W ).</td>
</tr>
<tr>
<td>( Q )</td>
<td>number of traffic types managed in the network, ( q \in Q ), ( Q \leq W ).</td>
</tr>
<tr>
<td>( G_s )</td>
<td>ONUs of the spur ( s \in S ).</td>
</tr>
<tr>
<td>( L_q )</td>
<td>ONUs belonging to the same technology ( q \in Q ).</td>
</tr>
<tr>
<td>( L_{q,s} )</td>
<td>ONUs belonging to the same technology ( q \in Q ) and to the same spur ( s \in S ).</td>
</tr>
<tr>
<td>( c_i )</td>
<td>capacity required by ONU ( i \in N ).</td>
</tr>
<tr>
<td>( w_j )</td>
<td>capacity of wavelength ( j \in W ).</td>
</tr>
</tbody>
</table>

Variables of the MILP

\[
\begin{align*}
  x_{ij} &= 1 \text{ if ONU } i \in N \text{ transmits over wavelength } j \in W. \\
  y_j &= 1 \text{ if the wavelength } j \in W \text{ is used.} \\
  z &\in Z \text{ total number (integer) of wavelengths used in the network.} \\
  z_s &\in Z \text{ number of wavelengths used in each spur: } \sum z_s \neq z. \\
  h_{js} &= 1 \text{ if wavelength } j \in W \text{ is used in spur } s \in S. 
\end{align*}
\]

of wavelengths dropped in each spur, while \( 6.5 \) sums up the total number of wavelengths used in the network.

We consider the two following objective functions in our study.

\[
\begin{align*}
  &\min(z) \quad (6.6) \\
  &\min(\sum_{s \in S} z_s) \quad (6.7)
\end{align*}
\]

Equation \( 6.6 \) minimizes the total number of wavelengths in the network (as calculated in Eqn. \( 6.5 \)), as we assume that the number of active wavelengths is related to the cost of the network (e.g., number of linecards in the OLT), and its power consumption. Equation \( 6.7 \) minimizes the sum of number of wavelengths used in each spur. We assume here that the sum of the wavelengths dropped at each spur represents an indirect metric capturing the complexity of the RNs (e.g., number of elements needed in the node to drop each wavelength).

In order to compare the VPON case with the case where the traffic is mixed over a single TWDM-PON we can use exactly the same ILP model, but considering all the ONUs of the network together, and not only those attaining to a specific traffic type. More precisely, to assign the capacity in the VPON case we solve multiple bin packing
problem over set $I (BP_I)$, where the set $I$ comprises all the ONUs of a particular traffic type, i.e. $I = L_q, \forall q \in Q$. Instead, for case where the traffic is mixed over a TWDM-PON we solve a single bin packing problem $BP_I$, where $I$ comprises all the ONUs of the network, i.e., $I = N$.

In the following section we present the scenarios in which we tested our model and the results provided by our optimizations.

6.3 Illustrative Numerical Results

6.3.1 Network Scenarios

In this work, we compared the two solutions $i)$ VPON case, and $ii)$ converged TWDM-PON in different network scenarios in order to investigate under which conditions (e.g., topology, network loads, users distribution) it is more convenient to use VPONs or a single TWDM PONs. The metrics used for the comparison are the total number of wavelengths needed to support the entire traffic and the sum of wavelengths dropped at each spur, as for Eqns. 6.6 and 6.7, respectively.

We consider that 3 types of traffic are backhauled over the same network: $i)$ Fixed traffic, $ii)$ Mobile 3G traffic, and $iii)$ Mobile 4G traffic. In particular, in our work, the ONUs that transmit 3G and 4G traffic are collocated with nodeBs and eNodeBs, respectively. While, the ONUs that transmit the fixed traffic are located in an aggregation point (e.g., a Cabinet). Practical representative values for the capacity required by the different kinds of ONUs are taken from [63]. Wavelength capacity is 1Gbit/s unless stated differently.

Using the models discussed above, we evaluate several network scenarios: $i)$ **VNI traffic distribution**, where the total traffic is split among the various service types according to the values provided by the CISCO Visual Networking Index (VNI) [64], i.e., Fixed traffic 76%, 3G traffic 12%, and 4G traffic 12%. Each spur has 40 ONUs, and each ONU belonging to a particular traffic type has a capacity which is uniformly distributed between a minimum and a maximum capacity that vary for each traffic type. $ii)$ **Uniform traffic distribution** scenario, where the aggregated traffic for each service type is the same (i.e., one third of the total traffic). Also in this case, each ONU belonging to a particular traffic type has a capacity which is randomly and uniformly distributed between a minimum and a maximum capacity that varies...
for each traffic type. iii) VNI traffic distribution with higher wavelength capacity, where the traffic in the network is distributed according to the figures of the CISCO VNI, but the wavelength capacity is 10Gbit/s instead of 1Gbit/s. iv) VNI traffic distribution with different number of ONUs per spur, where the traffic is distributed according to the figures of the CISCO VNI, but in each spur the number of connected ONUs varies. In all the investigated scenarios we consider peak hour traffic.

The exact values that we used in the different scenarios are reported in Table 6.2.

**Table 6.2: Network parameters for different network scenarios**

<table>
<thead>
<tr>
<th></th>
<th>VNI traffic distribution</th>
<th>Uniform traffic distribution</th>
<th>Higher wavelength capacity</th>
<th>Different ONUs per spur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of ONU s</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Number of RNs (spurs)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total Fixed ONU s</td>
<td>64</td>
<td>11</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Total 3G ONU s</td>
<td>312</td>
<td>365</td>
<td>312</td>
<td>312</td>
</tr>
<tr>
<td>Total 4G ONU s</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Wave-length capacity</td>
<td>1 Gbit/s</td>
<td>1 Gbit/s</td>
<td>10 Gbit/s</td>
<td>1 Gbit/s</td>
</tr>
<tr>
<td>Fixed ONU s rate</td>
<td>[0.5 - 1 Gbit/s]</td>
<td>[0.5 - 1 Gbit/s]</td>
<td>[0.5 - 1 Gbit/s]</td>
<td>[0.5 - 1 Gbit/s]</td>
</tr>
<tr>
<td>ONUs per spur</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>[20-60]</td>
</tr>
</tbody>
</table>
6.3.2 Numerical Results and Discussion

In Fig. 6.2, we compare the total number of wavelengths used in the network in the two cases, VPON and converged TWDM PON for all the 4 scenarios. As expected, in all the tested scenarios, the total number of wavelengths in the converged TWDM case is lower than in the VPON case. This is due to the fact that in the converged TWDM case the different types of traffic are mixed over all the available wavelengths, therefore higher traffic multiplexing is enabled and the capacity of the wavelengths is used more efficiently, i.e., less bandwidth is left unused. However, it shall be noted that, for the practical values of ONUs’ traffic used in this analysis, the VPON case pays only a minor increase in the number of wavelengths (i.e., at most 12% more wavelengths than in the converged TWDM case). While paying a slight increase in terms of required capacity, the VPON case yields several benefits: i) At the OLT-side, there is a simplification of the traffic management method. In the VPON case the wavelength which transports all the traffic of a particular VPON is routed towards the same core node and it is not necessary for the OLT to check each packet to decide which types of traffic is transporting. ii) At the data-link layer, using VPONs to segregate the different types of traffic it is not necessary to have a convergent protocol at Layer 2 in the access-aggregation networks. iii) At the ONU-side, by using...
Chapter 6. Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

Figure 6.3: Comparison in the number of wavelengths per spur between the VPON case and TWDM-PON case, in the network scenarios investigated.

several VPONs in the same network it is possible to install cheaper transceivers (e.g., tunable lasers with smaller tuning range) since ONUs have to transmit over a limited set of wavelengths, and not over all the wavelengths of the network.

In Fig. 6.3, we compare the number of wavelengths dropped at each spur (for all the 10 spurs) of the network in the two cases, VPON and converged TWDM, and in all the 4 studied scenarios. We can notice that in the converged TWDM case the number of wavelengths dropped in each spur is lower than in the VPON case in all the studied scenarios. As for the total number of wavelength, also in this case the fact that different traffic types shall be served on separate wavelengths increases the number of wavelengths to be dropped at each spur, and therefore the complexity of the node.

Varying the Traffic Distribution. In Fig. 6.2, if we compare the results for VNI traffic distribution and for the uniform traffic distribution scenario, we can evaluate the effect of varying the traffic distribution in the comparison between the VPON case and the converged TWDM case. First, we can notice that also in the uniform traffic distribution scenario the total number of wavelengths used in the VPON case is about 12% higher than in the converged TWDM case (in other words, a variation of the traffic distribution does not affect the relative difference in terms of number of wavelengths between VPON and converged TWDM cases). Note also that in the uniform traffic distribution scenario the total number of wavelengths is lower than
in the VNI traffic distribution scenario, since in the uniform traffic distribution scenario the total traffic is lower than in the VNI traffic distribution case.\footnote{The reason for this behavior is that we decided to keep a constant number, namely 400, of ONUs in the network, so, if we increase the percentage of 3G ONUs, that require a relatively low capacity, for 12\% to 33\%, the overall traffic decreases.} The same considerations apply in Fig. 6.3, where it is confirmed that the number of wavelengths per spur in the uniform traffic distribution scenario is lower than in the VNI traffic distribution scenario. Within each single scenario of each single case (i.e., VPON and converged TWDM cases), we notice that the number of wavelength dropped in each spur vary even if the total number of ONUs per spur is the same in all the cases. This is due to the fact that the capacity of each ONU varies between a maximum and a minimum, therefore among the different spurs the aggregated traffic provided to the network is subject to small changes.

**Varying Wavelength Capacity.** Let us now consider the comparison between the VPON case and the converged TWDM case for the VNI traffic distribution using higher wavelength capacity (10Gbit/s). In Fig. 6.2, we notice that using 10Gbit/s wavelength capacity the total number of wavelengths is much lower as, having wavelengths with higher capacity, a smaller number of wavelengths is needed to accommodate the same amount of traffic. Moreover, for higher wavelength capacity the difference in terms of number of wavelengths is smaller between the VPON case and the converged TWDM case. This happens because, when the wavelength capacity is much higher than the transmission capacity required by each ONU, the capacity constraint of the bin-packing problem becomes much less strict, and it is much easier to fit ONUs’ required capacity in the wavelengths without wasting bandwidth. This effect emerges also from Fig. 6.3 where, when considering results for 10Gbit/s, in both VPON and converged TWDM cases, the number of wavelengths dropped in each spur is the same for all the spurs of the network. That is, in this case, the variation of the aggregated capacity of each spur does not translate in a variation of number of dropped wavelengths in the different spurs, because of the higher wavelength capacity.

**Number of ONUs per spur variation** To evaluate the effects of the variation of the number of ONUs per spur we now consider the case of VNI traffic distribution with different number of ONUs per spur (fourth case in Fig. 6.2). The total number of wavelengths in both the VPON case and converged TWDM case are same as in the case of
fixed number of users per spur, i.e., unbalanced ONUs’ distribution among the various spurs does not affect the overall result in terms of number of wavelengths. This is also related to the underlying bin-packing problem structure, for which the spatial distribution of users is not relevant in terms of the efficiency of capacity filling. Contrariwise, Fig. 6.3 shows that the number of wavelengths per spur is different compared to the case of fixed number of ONUs per spur as the distribution of the ONUs in the spurs is now more unbalanced. Therefore, assuming that the number of wavelengths dropped in each spur is proportional to the complexity of the RNs used for the drop, in the different number of ONUs per spur scenario it is necessary to install more diverse, and more complex RNs.

Figure 6.4: Comparison in the total number of wavelengths used between the VPON case and TWDM-PON case, in a penalizing network scenario.

Considerations on a network scenario penalizing VPON approach

In this subsection we consider a specific network scenario where the difference between the VPON case and the converged TWDM case gets much larger than in all the previous practical scenarios. VPON is penalized with respect to a converged TWDM-PON case when the required capacities of the ONUs associated to the different traffic
6.3. Illustrative Numerical Results

Figure 6.5: Comparison in the number of wavelengths per spur between the VPON case and TWDM-PON case, in a penalizing network scenario.

Types assume specific values (typically slightly larger than half of the wavelength capacity) that do not allow an efficient utilization of wavelength capacity. In this way, we aim at identifying the worst conditions for VPON deployment in an FMC network.

A possible set of network parameters of a penalizing network scenario is reported in Tab. 6.3.

In this network scenario, the total number of wavelengths used in the VPON case is more than 25% higher than in the converged TWDM case. This is due to the scarce fit of the required capacities of fixed and 4G ONUs in the wavelength capacity in the VPON case (e.g., in the fixed traffic case, a 600Mbit/s ONU will occupy an entire 1Gbit/s wavelength, without any possibility to use the remaining 400Mbit/s by some other types of traffic). Instead, in the converged TWDM case, the total number of wavelengths can be strongly reduced with respect to the VPON case as different types of traffic can be mapped together in the same wavelength. As future work, we will extend this preliminary insight in the scenarios penalizing VPON adoption by extending existing methodologies to derive worst case performance.
Chapter 6. Virtual PON Assignment for Fixed-Mobile Convergent Access-Aggregation Networks

Table 6.3: Network parameters for the particular case evaluated

<table>
<thead>
<tr>
<th>Particular scenario</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of ONUs</td>
<td>400</td>
</tr>
<tr>
<td>Number of RNs (spurs)</td>
<td>10</td>
</tr>
<tr>
<td>Total Number of Fixed ONUs</td>
<td>64</td>
</tr>
<tr>
<td>Total Number of 3G ONUs</td>
<td>312</td>
</tr>
<tr>
<td>Total Number of 4G ONUs</td>
<td>24</td>
</tr>
<tr>
<td>Wavelength capacity</td>
<td>1 Gbit/s</td>
</tr>
<tr>
<td>Fixed ONUs transmission rate</td>
<td>600 Mbit/s</td>
</tr>
<tr>
<td>3G ONUs transmission rate</td>
<td>30 Mbit/s</td>
</tr>
<tr>
<td>4G ONUs transmission rate</td>
<td>400 Mbit/s</td>
</tr>
<tr>
<td>Number of ONUs per spur</td>
<td>40</td>
</tr>
</tbody>
</table>

analysis for the bin-packing problem (see, e.g., [65]).

6.4 Conclusion

In this work, our goal was to evaluate if, in a FMC access-aggregation network, it is more convenient to isolate the different types of traffic over different sets of wavelengths (i.e., VPON case) or to mix the traffic types over all the available wavelengths (i.e., converged TWDM case). To model such comparison we developed ILP formulations for both cases, based on the classical bin-packing problem. Through numerical evaluation we investigate the trade-off between the VPON case and the converged TWDM case, under different network scenarios. We can conclude that, in terms of total number of wavelengths assigned the converged TWDM case is always slightly better (i.e., about 12%) than the VPON case. However, to use VPONs provides several benefits: i) At the OLT-side, there is a simplification of the traffic management method. ii) At the data-link layer, using VPONs to segregate the different types of traffic it is not necessary to have a convergent protocol in the access-aggregation segment of the network. iii) At the ONU-side, by using several VPONs in the same network it is possible to install cheaper transceivers (e.g., tunable lasers with smaller tuning range). As a future works we plan to quantitatively evaluate these benefits mentioned above to perform a more accurate comparison between VPON and converged TWDM case.

Finally, in our work we also discussed a specific and penalizing network scenario, where the utilization of VPONs is not suitable due to scarce fit of required ONU capacities in the wavelength capacity.
We plan to generalize such analysis using worst-case bin packing problem analysis.
CHAPTER 7

Multi-Operator Network Sharing

The massive uptake of smartphones and tablets, of the mobile Internet, and digitization technologies such as cloud computing will increase the data traffic. Between now and 2016, it is estimated that mobile data traffic will multiply tenfold with video acting as the biggest driver. By 2020, the OTTs video offerings will account for more than half of total data volume.

Exponential growth of data volume can be considered as one of the main drivers for investments in telecom infrastructure, and in particular in FMC. Moreover, the evolution of customer demand to be online everywhere every time with different devices at the same time is the main driver behind the exponential growth of data traffic and therefore for FMC infrastructures.

The need for more capacity in Internet access, in relation to connectivity speed as well as network dimensioning, will lead to the necessity of changes in the network types due to the limited capacity actual networks can offer in relation to the exponential growth of data traffic and customer demand. Therefore, in the next years, the number of both fixed and mobile broadband subscriptions will continue to rise, and many people will have access to broadband through both of
them.

The FMC architecture will allow various fixed and mobile applications to share the same infrastructure, while at the same time aiming at allowing strict separation (unbundling) and by thus, ensuring that general regulation principles are satisfied.

For these reasons, it will be needed to provide to the end-users an increased capacity in the access network. Several methods can be adopted in order to increase the access network capacity:

1. Increase of network capacity e.g., number of fibers, capacity per fiber;

2. Traffic offloading, e.g., Wi-Fi offloading, Multi-operator network sharing.

In a FMC network several types of handover can occur: i) Horizontal handover, ii) Vertical handover, and iii) Multi-operator handover. Horizontal handover occurs when a user changes its connection point within the same technology, e.g., a user previously connected to a LTE eNodeB connects to a different LTE eNodeB. Vertical handover happens when a user roams from one technology to another, e.g., a user connected to a LTE eNodeB connects to a Wi-Fi Access Point (AP). A Multi-operator handover is performed when a user is offloaded from the network of one operator to the network of another operator. Also combinations of these three types of handover may arise. For example a Vertical handover can occur between the networks of two different operators, e.g., a user connected to a Wi-Fi AP of operator A connects to a LTE eNodeB of operator B.

Recently, network sharing has become a popular solution among operators to solve different issues: i) to speed the rollout of 3G/4G services, ii) to reduce substantial investments in network deployment and iii) 3G licenses purchase. Today, operators begin to pay attention to sharing network resources between each other in order to lower capital expenditure (CAPEX) in infrastructure establishment and lower operation expenditure (OPEX) in the long run.

There are some existing schemes to resolve this problem.

In our work, we propose a multi-operator network sharing approach based on non-cooperative game theory (GT) that allow operators to increase the traffic served during peak-hours. At the same time, our proposed approach aims to avoid that one operator fills all the unused bandwidth of the second operator causing a consistent traffic
7.1. Related Works

In this section, we review already proposed method to perform network sharing. In [66], a Fixed Spectrum Allocation scheme (FSA) among Mobile Virtual Network Operators (MVNOs) is proposed. This approach is simple to be realized and managed, but it has a low level of sharing and a low spectral efficiency, which will cause serious waste of spectrum resource. In [67]- [68], a Partial Spectrum Sharing scheme (PSS) can promote spectral efficiency, but spectral utilization rate is still low. Also in [70]- [71] is proposed a dynamic spectrum sharing scheme (DSS) based dynamic prioritization of MVNOs, however in this scheme MVNOs short term fairness can not be guarantee due to insensitive priority factor. Instead, our proposed collaboration strategy aims at maintaining fairness for the users of each network, and consequently for the operators. Fairness for the users of each network means that our proposed collaboration procedure wants to preserve some capacity of the network of operator in order to avoid to refuse the connection to users of its own network, while serving as much traffic as possible of the other operator. All the three traditional schemes (i.e., FSA, PSS, and DSS) can not guarantee full usage of radio spectrum and the basic interest of MVNOs at the same time. In order to protect MVNOs interest, promote MVNOs throughput and enhance spectral efficiency as much as possible, a Based Multi operator Shared Network Opportunistic Spectrum Sharing (BMSN OSS) scheme is presented in [72].

The benefits, drivers, drawbacks and risk with shared networks, with focus on macro cell networks, have been investigated in many papers, e.g. [73] - [75], and is quite well understood.

In order to improve indoor coverage two types of solutions are widely used; Distributed Antenna Systems (DAS) [76]. Here com-
Chapter 7. Multi-Operator Network Sharing

peting operators cooperate with each other and also with the facility owner and/or with companies using the indoor infrastructure. In this multi-operator settings the physical infrastructure i.e. the DAS network and the repeater equipment, is shared. However, the radio capacity (the base stations), the spectrum and the access control are managed by each operator. In our work, the collaboration consists in offloading the excess traffic load of one operator to the network of another operator. This type of collaboration occurs i) when two operators have two different network infrastructures in the same area (e.g., a cell with two base stations), or ii) when two operators use the same network infrastructure sharing also the spectrum (i.e., the capacity) when needed.

In [77] are investigate strategies for active sharing of radio access among multiple operators, and analyses the benefits depending on the sharing degree. It is assumed that each operator owns certain percentage of the total resources and enters a sharing agreement. The agreement defines the portion of resources that are shared among all operators and can be accessed on a first come first served basis. Conversely, in our proposed collaboration strategies there are not any agreement among operators and the decision about the collaboration (i.e., to accept the traffic of the other operator or not) depends on the traffic load conditions of the operators involved in the decision proces.

In the following section we present in details the scenario that we consider in this work.

7.2 Scenario

In our work we consider an offloading scenario. We call offloading scenario the situation where several operators have already deployed their networks and they perform traffic offloading for several reasons such as: i) Energy efficiency, ii) Throughput improvement. The aim of our proposed multi-operator network sharing scenario is to increase the total amount of traffic served by the collaborating operator while maintaining the fairness for the users of each network.

In our basic scenario, for each site (e.g., 3G/4G cells, Wi-Fi access points) we have multiple coverage, i.e., each operator has its own access infrastructure. Moreover, we consider a scenario where only two operators coexist. We will refer to the operator that wants to offload the traffic as the Offloader, while we call the operator that can
accept the traffic the Receiver. This scenario is presented in Fig. 7.1.

Figure 7.1: FMC network infrastructure considered for the offloading scenario.

In particular, Fig. 7.1 shows a scenario where two operators have two separated FMC infrastructures in the metro/access segment and several technologies (e.g., 3G antennas, 4G antennas, Wi-Fi APs) are backhauled over this infrastructure. For each cell (3G, 4G, or Wi-Fi) both operators have antennas. The Decision Engine is the core node in which the decision about the collaboration is taken. The two operators have to send the information needed to decide if the operators will collaborate or not. The information that have to be sent to the Decision Engine depends on the collaboration strategy used. In this case, when the collaboration occurs the users of one operator are served by the network infrastructure of the second operator. In Fig. 7.1 we show only two users: one (in blue) belonging to operator 1 and one (in red) belonging to operator 2. Another interpretation of our scenario is that operators share the same FMC network infrastructure using two different portion of the spectrum. In this case, when the collaboration occurs the users of one operator are served using a part of the portion of spectrum of the second operator.

In our scenario, the traffic is offloaded from one operator to another when the amount of traffic that have to be served it is higher than the network capacity. Therefore, the traffic exceeding the network capacity of the operator will be offloaded to the network of another operator. The Receiver can decide whether accept or not this amount of traffic using the game theoretic approach proposed in our work. If the Receiver decides to not accept, then the traffic of the Offloader
can not be served and therefore is lost.

Within this scenario the collaboration between the two operators is evaluated when one operator needs to serve more traffic with respect to the capacity of its own network and the second operators has some available capacity, and viceversa. Instead, when both operators have to serve more traffic with respect the capacity of the network no collaboration occurs. The same applies to the case when both operators are in a light loaded traffic condition. In our model, we are considering the networks of the operators as mono- and same-technology networks: we are not assuming any particular technology for the different networks. The fact that offloading the traffic from one network type (e.g., Wi-Fi) to another type of network (e.g., LTE) is listed in the future works.

In the following section we will explain in details the game theoretic model proposed to solve the multi-operator collaboration decision.

### 7.3 Game Theoretic model for Multi-operator Network Sharing

Game theory is the study of strategic decision making. It is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers. The games studied in game theory are well-defined mathematical objects. To be fully defined, a game must specify: i) the players of the game, ii) the information and actions available to each player at each decision point, and iii) the payoffs for each outcome. These elements are used together with a solution concept to deduce a set of equilibrium strategies for each player. When the equilibrium strategies are employed no player can profit by unilaterally deviating from their strategy. These equilibrium strategies determine an equilibrium to the game, i.e., a stable state in which one outcome or a set of outcomes occur with known probability.

The normal form game, also called strategic form, is usually represented by a matrix which shows the players, the strategies, and the payoffs. More generally it can be represented by any function that associates a payoff for each player with every possible combination of actions. A game can be cooperative or non cooperative. A game is cooperative if the players are able to form binding commitments. In non-cooperative games players make decisions independently. Moreover, a game can be a symmetric game or a non-symmetric game. A symmetric game is a game where the payoffs for playing a particular strategy
depend only on the other strategies employed, not on who is playing them. If the identities of the players can be changed without changing the payoff to the strategies, then a game is symmetric. Asymmetric games are games where there are not identical strategy sets for both players. Also, games can be zero-sum games or non-zero-sum games. Zero-sum games choices of the players can neither increase nor decrease the available resources. In zero-sum games the total benefit to all players in the game, for every combination of strategies, always adds to zero, e.g., if one player gain 1, then the other player lose 1. In non-zero-sum games outcomes have net results greater or less than zero.

The game that we model in this work is a non cooperative, asymmetric, and non-zero-sum game. Indeed, in the proposed game each player chooses its strategy independently to improve its own performance (i.e., utility) or reducing its losses (i.e., costs). The players of this game are the network operators. The only information shared between players is the amount of traffic exceeding the capacity of the network of an operator, i.e., the amount of bandwidth that an operator wants to offload. In this game each player has two strategies that depend on the role of player: the Offloader can decide whether offload or not, while the Receiver can decide to accept or not to accept the traffic of the Offloader. This game is played every time an operator has some traffic exceeding the capacity of its network. Therefore, the role of the player can change according to the load condition of each network. The solution of the proposed game is represented by the Nash Equilibrium. The Nash equilibrium is a solution concept of a non-cooperative game in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only their own strategy.

In Table 7.1 we define the notation used hereafter.
Chapter 7. Multi-Operator Network Sharing

Table 7.1: Definition of parameters and variables

<table>
<thead>
<tr>
<th>Definitions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>number of operators, i.e., players.</td>
</tr>
<tr>
<td>$C_n$</td>
<td>total capacity of the network of player $n$, $n \in N$.</td>
</tr>
<tr>
<td>$L_n$</td>
<td>network load (number of users served) of the player $n$, $n \in N$.</td>
</tr>
<tr>
<td>$R_n$</td>
<td>available capacity of player $n$, $n \in N$.</td>
</tr>
<tr>
<td>$Ex_n$</td>
<td>excess load of player $n$, $n \in N$.</td>
</tr>
<tr>
<td>$T_n$</td>
<td>amount of excess load accepted by player $n$, $n \in N$.</td>
</tr>
<tr>
<td>$Ex_{m,n}$</td>
<td>maximum amount of excess load of player $n$ that can be accepted by player $m$, $m,n \in N$.</td>
</tr>
<tr>
<td>$\rho_s$</td>
<td>total amount of traffic served.</td>
</tr>
</tbody>
</table>
7.3. Game Theoretic model for Multi-operator Network Sharing

In Tab. 7.2 we report the table that describes our proposed game. In particular, the game modeled in 7.2 represents a game with two players: player 1 is the Offloader while player 2 is the Receiver. For each strategy of the two players we report the equations used to compute the payoffs.

Table 7.2: Table of the proposed game

<table>
<thead>
<tr>
<th>Player 1</th>
<th></th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Offload</td>
<td>Not Offload</td>
</tr>
<tr>
<td>Accept</td>
<td>$$(E_{X_1} - E_{X_1} + 1; L_1 - C_1)$$</td>
<td>$$(E_{X_1} - E_{X_1} + 1; C_1 - L_1 + 1)$$</td>
</tr>
<tr>
<td>Not Accept</td>
<td>$$(E_{X_1} - E_{X_1}; L_1 - C_1)$$</td>
<td>$$(E_{X_1} - E_{X_1}; C_1 - L_1 + 1)$$</td>
</tr>
</tbody>
</table>

The choice of the strategy to be adopted for each player depends on the traffic conditions of the network of each operator. In particular, the Offloader (i.e., Player 1) decides to offload the traffic if $$L_1 - C_1 > 0$$, i.e., if the traffic offered to the network is higher than the total capacity of the network. The Receiver (Player 2) decides to Accept the excess load of player 1 ($$E_{X_1}$$) if $$E_{X_1} \leq \bar{E}_{X_1}$$. $$\bar{E}_{X_1}$$ is the maximum amount of excess bandwidth of Player 1 that the Receiver can serve in order to increase the total amount of traffic served while avoiding to increase the loss probability. $$\bar{E}_{X_1}$$ is computed through the function in Eqn. 7.1 where $$E_{X_1}$$ is the value of $$E_{X_1}$$ that maximize the function $$U_2(E_{X_1})$$ reported in Eqn. 7.1. We call the function in 7.1 the utility function of the Receiver. The solution of the game in Tab. 7.2 is given by the Nash Equilibrium of the game.

$$U_2(E_{X_1}) = \rho_s - E_b(E_{X_1}, C_2 - L_2)$$ \hspace{1cm} (7.1)

$$E_b(E_{X_1}, C_2 - L_2)$$ is the blocking probability of the Receiver where $$C_2 - L_2$$ is the number of servers and $$E_{X_1}$$ is the traffic that has to be served. The total traffic served by player 2, $$\rho_s$$ is computed through Eqn. 7.2

$$\rho_s = \frac{(E_{X_1} + L_2)}{C_2}$$ \hspace{1cm} (7.2)

During this procedure, the only information that is shared between the two operators is the value of $$E_{X_1}$$. In the following Tables 7.3, 7.4, and 7.5 we describe three possible scenarios with different traffic conditions and we compute the related payoffs to provide some example on how the game is solved (i.e., the decision about the
collaboration is taken) in these different situations. In Tab. 7.3 we represent a situation where the network of operator 1 (the Offloader) is lightly loaded, i.e., the load $L_1$ is lower than the network capacity $C_1$. In this case the solution of the game is given by the Nash Equilibrium that corresponds with the strategy where player 1 not offload and player 2 can indifferently accept or not accept the traffic.

**Table 7.3:** Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 60, E\bar{x}_1 = -40, E\bar{x}_1 = 0$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td>Accept</td>
<td>(41; -40)</td>
</tr>
<tr>
<td></td>
<td>Not Accept</td>
<td>(-40; -40)</td>
</tr>
</tbody>
</table>

In Tab. 7.4 we represent a situation where the network of operator 1 (the Offloader) is overloaded, i.e., the load $L_1$ is higher than the network capacity $C_1$. In this case the Nash Equilibrium corresponds with the strategy where player 1 offload and player 2 accepts the traffic.

**Table 7.4:** Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 120, E\bar{x}_1 = 20, E\bar{x}_1 = 30$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td>Accept</td>
<td>(11; 20)</td>
</tr>
<tr>
<td></td>
<td>Not Accept</td>
<td>(-40; -40)</td>
</tr>
</tbody>
</table>

In Tab. 7.5 we represent again a situation where the network of operator 1 (the Offloader) is overloaded but the maximum amount of traffic accepted by the Receiver ($E\bar{x}_1$) is lower than the excess traffic that the Offloader wants to offload. In this case the Nash Equilibrium corresponds with the strategy where player 1 offload and player 2 not accepts the traffic.

**Table 7.5:** Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 140, E\bar{x}_1 = 40, E\bar{x}_1 = 30$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td>Accept</td>
<td>(-9; 40)</td>
</tr>
<tr>
<td></td>
<td>Not Accept</td>
<td>(10; 40)</td>
</tr>
</tbody>
</table>
7.4 Game theoretic Approach with Partial Accept

In this section we will present the game in which players (i.e., network operators) can accept also a portion of the excess load $Ex_n$ of the other player. In this version of the game the players accept to serve a portion of traffic $\overline{Ex}_n$ if $\overline{Ex}_n > 0$. If $Ex_n > \overline{Ex}_n$ then the rest of the excess load of player $n$ is lost. The Offloader (i.e., Player 1) decides to offload the traffic if $L_1 - C_1 > 0$, i.e., if the traffic offered to the network is higher than the total capacity of the network. Again, the choice of the strategy to be adopted for each player depends on the traffic conditions of the network of each operator. $\overline{Ex}_1$ is computed through the function in Eqn. 7.1 where $\overline{Ex}_1$ is the value of $Ex_1$ that maximize the function $U_2(Ex_1)$ reported in Eqn. 7.1. We call the function in 7.1 the utility function of the Receiver. In Tab. 7.6 we present the game used to decide whether collaborate or not. The solution of the game in Tab. 7.6 is the Nash Equilibrium of the game.

### Table 7.6: Table of the proposed game with Partial Accept

<table>
<thead>
<tr>
<th>Player 2</th>
<th>Player 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Offload</td>
</tr>
<tr>
<td>Accept</td>
<td>$(\overline{Ex}_1; L_1 - C_1)$</td>
</tr>
<tr>
<td>Not Accept</td>
<td>$(\frac{L_1}{\overline{Ex}_1 + 1}; L_1 - C_1)$</td>
</tr>
</tbody>
</table>

Also in this version of the game, the only information that is shared between the two operators is the value of $Ex_1$. However, in this case the Players could take a decision about the strategy to be adopted without using the information about the amount of excess load. In the following Tables 7.7, 7.8, and 7.9 we describe three possible scenarios with different traffic conditions and we compute the related payoffs to provide some example on how the game is solved (i.e., the decision about the collaboration is taken) in these different situations. In Tab. 7.7 we represent a situation where the network of operator 1 (the Offloader) is lightly loaded, i.e., the load $L_1$ is lower than the network capacity $C_1$. In this case the solution of the game is given by the Nash Equilibrium that corresponds with the strategy where player 1 not offload and player 2 can indifferently accept or not accept the traffic. In this case, the solution of the game with Partial Accept is the same compared to the game presented in the previous section.

In Tab. 7.8 we represent a situation where the network of operator
Chapter 7. Multi-Operator Network Sharing

Table 7.7: Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100$, $L_1 = 60$, $E_{X_1} = -40$, $\bar{E}_{X_1} = 0$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>(0; -40)</td>
<td>(0; 41)</td>
</tr>
<tr>
<td>Not Accept</td>
<td>(1; -40)</td>
<td>(1; 41)</td>
</tr>
</tbody>
</table>

1 (the Offloader) is overloaded, i.e., the load $L_1$ is higher than the network capacity $C_1$. In this case the Nash Equilibrium corresponds with the strategy where player 1 offloads and player 2 accepts the traffic. Again, the solutions of the game with Partial Accept and the solution presented in the previous section correspond.

Table 7.8: Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100$, $L_1 = 120$, $E_{X_1} = 20$, $\bar{E}_{X_1} = 30$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>(30; 20)</td>
<td>(30; 19)</td>
</tr>
<tr>
<td>Not Accept</td>
<td>(1.96; 20)</td>
<td>(1.96; -19)</td>
</tr>
</tbody>
</table>

In Tab. 7.9 is studied another situation where the network of operator 1 (the Offloader) is overloaded but the maximum amount of traffic accepted by the Receiver ($\bar{E}_{X_1}$) is lower than the excess traffic that the Offloader wants to offload. In this case the Nash Equilibrium corresponds with the strategy where player 1 offloads and player 2 accepts to serve the traffic up to $\bar{E}_{X_1}$. The solution provided by the game with Partial Accept in this situation differs from the solution provided by the game presented in the previous section in the same scenario. Therefore, we expect that using the strategy with Partial Accept there will be a higher collaboration between the operators.

Table 7.9: Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100$, $L_1 = 140$, $E_{X_1} = 40$, $\bar{E}_{X_1} = 30$

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Offload</th>
<th>Not Offload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>(30; 40)</td>
<td>(30; -39)</td>
</tr>
<tr>
<td>Not Accept</td>
<td>(1.96; 40)</td>
<td>(1.96; -39)</td>
</tr>
</tbody>
</table>

In order to validate our proposed game theoretic approaches we compare them with some benchmarking procedures. Particularly,
we compare our procedures with the cases where: i) collaboration between players never occur, i.e., No collaboration case; ii) the Receiver always accepts to serve the traffic of the Offloader if the amount of traffic that have to be offloaded is smaller or equal to the available capacity of the Receiver, i.e., Full collaboration case; iii) the Receiver always accepts to serve the traffic of the Offloader if its network has some available capacity Full collaboration case Partial Accept. In this case, if the amount of traffic that has to be offloaded is higher than the available capacity of the Receiver, only a portion of the excess load is served.

7.5 Illustrative Numerical Results

In this section we introduce the simulative scenario that we use to validate our method and we explain the results obtained.

The network scenario that we consider is the one presented in Fig. 7.1 where for each cell (Wi-Fi AP or mobile NodeB) there is double coverage. In particular, we consider one single cell where users of two different network operators want to access the network. Therefore, in our game we are considering only two players. In the following, we will compare the presented methods under different traffic loads of the two networks. The load \( L_n \) depends on the arrival rate \( \mu \) of the connection and on the holding time \( \tau \) of the connections served. The traffic provided to the network in our simulations is distributed as a Poisson process with average \( \mu \).

Moreover, we will also vary the holding time \( \tau \) of the connections served in order to study the impact of the value of \( \tau \) on the results.

Finally, we will compare the presented strategies under different traffic profiles that change during the day.

7.5.1 Network scenario

We compared our proposed GT approach with the benchmarks, i.e., Full collaboration and No collaboration approaches in several scenarios with different traffic loads.

We focus here on one particular traffic conditions of operator 1 and operator 2 since similar considerations can be noticed under several traffic loads. We also omit here the scenarios where the collaboration never occurs: i) the cases where both operators have a higher offered traffic with respect to their network capacity; ii) the cases where both
operators have a offered traffic which is lower than their network capacity.

The traffic load that we consider are normalized: we consider a network capacity of 100 and the traffic load number represents the percentage of network capacity used. For example, a traffic load of 80 is an amount of traffic that uses the 80% of the network capacity. A traffic load of 180 represents a traffic that completely uses the network capacity and requires another 80% of that capacity to be served.

7.5.2 Numerical Results and Discussion

In Fig. 7.2 we compare the average traffic served by player 1 in the cases of GT collaboration with Partial Acceptance, Full collaboration with Partial Acceptance, GT collaboration, Full collaboration, and No collaboration. In this scenario, the $\tau$ is 3, the traffic load of player 1 is 180, and the traffic load of player 2 is 90. The average amount of traffic served by a network is composed by both traffic of the operator 1 and traffic offloaded by operator 2, i.e., collaboration traffic. In Fig. 7.2 we can notice that all the strategies serve about a traffic of 100, this because the average traffic provided to the network is 180. The average traffic served in this case is not exactly 100 because of the Poisson nature of the traffic.

Figure 7.3 we compare the average traffic served by player 2 in the cases of GT collaboration with Partial Acceptance, Full collaboration with Partial Acceptance, GT collaboration, Full collaboration, and No collaboration. In Fig. 7.3 we can notice that the Full collaboration with Partial Acceptance approach is the strategy that is able to serve the highest amount of traffic. This is due to the fact that with this strategy player 2 always accepts the traffic offloaded by player 1 if there is some capacity available. The GT collaboration with Partial Acceptance and the Full collaboration approaches serves an amount of traffic which is only slightly lower than the Full collaboration with Partial Acceptance case. The GT collaboration approach the traffic served is about 95. Finally, the No collaboration approach serves the smallest amount of traffic among the collaboration strategies presented. Clearly, this happens because the exceeding traffic of player 1 is never served by player 2.

Figure 7.4 shows the average amount of traffic of player 2 served by player 1, in all the studied approaches. In Fig. 7.4 we notice that the average amount of traffic of player 2 served by player 1 is
very small (almost zero). This is due to the fact that player 1 has a high amount of offered traffic, therefore it almost never has available capacity to serve other traffic. Also in this case, the fact that player 2 has some traffic that needs to be offloaded is due to the Poisson nature of the traffic: the offered traffic has a average of 90, this means that sometimes the instantaneous traffic is higher than 100. Figure 7.5 shows the average amount of traffic of player 1 served by player 2, in all the studied approaches. In Fig. 7.5 we notice that the approach that allows to serve the highest amount of traffic of the other operator is the Full collaboration with Partial Acceptance. This, again, is due to the fact that with this strategy player 2 always accepts the traffic offloaded by player 1 if there is some capacity available.

Then in Fig. 7.6 we show the average amount of traffic lost by operator 1 (i.e., player 1). Using No collaboration approach is lost the highest amount of traffic, shows Fig. 7.6 because player 2 never accepts to serve traffic of player 1. Conversely, with Full collaboration with Partial Acceptance the lowest amount of traffic is lost because with this approach player 2 always accepts to serve the traffic of player 1, when some capacity is available on its network. Figure 7.7 shows the average amount of traffic lost by operator 2. In this case,

![Average traffic served by Player 1 using the different collaboration strategies.](image)
we can notice the impact of our proposed GT approach (both without and with Partial Acceptance). In fact, Fig. 7.7 shows that using the Full collaboration approach with and without Full Acceptance we lose more traffic than in the GT approach with and without Full Acceptance. Moreover, the GT approach lose only a small amount of traffic more than the No collaboration case. This means that part of this traffic is lost because of the offered traffic conditions. Therefore, we can state that our proposed GT approach is able to preserve network capacity for the users of the network, i.e., maintains fairness for the users of the operator.

Then, with our proposed GT approach we are able to increase the average amount of traffic served with respect to the case where No collaboration occurs, while preserving network capacity to serves the traffic of its own users.

**7.6 Conclusion**

In this work, our goal was to propose and evaluate a game theoretic (GT) collaboration strategy based on the traffic load of the networks of the operators involved in the decision process. The objective of our
7.6. Conclusion

The proposed collaboration procedure is to increase the total amount of traffic served (i.e., the traffic served by all the operators involved in the network sharing procedure). At the same time, the GT collaboration model proposed aims at maintaining the fairness for the users of the operators: the decision procedure wants to preserve some capacity of the network of operator in order to avoid to refuse the connection to users of its own network, while serving as much traffic as possible of the other operator. Moreover, with our collaboration strategy, the amount of traffic of other operators that one operator accepts to serve change dynamically according to the traffic load of all the players involved in the game (i.e., decision procedure). Results show that our GT collaboration model is able to improve the amount of traffic served by the operators with respect to the case where operators not collaborate. However, the GT collaboration have the lowest amount of traffic lost due to the collaboration (i.e., traffic of an operator which can not be served because the capacity of the operator is filled by the traffic of other operators) with respect to all the other collaboration methods evaluated in this work. At the same time, with the GT collaboration the amount of traffic lost by the operator serving the excess traffic of another operator is only slightly higher than the case.
where no collaboration occurs. This means that the highest portion of the traffic lost by the Receiver is lost due to the condition of the offered traffic, and not to the collaboration.

Considering the GT model, a future work is to improve the model in order to consider more than 2 operators (i.e., players) in the game. Moreover, a second future work is to extend the GT model in order to consider also the strategies of offloading the traffic over different technologies, e.g., handover between Wi-Fi and 3G.

Figure 7.5: Average traffic of Player 2 served by Player 1 using the different collaboration strategies.
Figure 7.6: Average traffic lost of Player 1 using the different collaboration strategies.

Figure 7.7: Average traffic lost of Player 1 using the different collaboration strategies.
Conclusion

In this thesis the aim was to propose, define and technically assess methods to efficiently manage the convergence of Fixed and Mobile traffics over the same infrastructure.

For this purpose, we first presented, analyzed and evaluated different strategies for LR WDM/TDM PON to perform upstream scheduling and wavelength assignment.

Regarding the implementation of a pure Multi-thread polling strategy in LR WDM/TDM PONs, our results show that, while at low/medium loads it achieves large improvement with respect to EFT algorithms, for medium/high loads it suffers from a sudden increase of delay. So, in a LR WDM/TDM PON, the pure application of Multi-thread algorithm replicates the same gain achieved in LR TDM PONs with respect to single thread polling, at least at low and medium loads. Therefore, in this work we found an effective solution to combine the positive features of the EFT-VF algorithm with the Multi-thread algorithm, in order to provide a new and efficient allocation scheme for LR WDM/TDM PON. In fact, simulation results show that the EFT-partial-VF Multi-threaded algorithm provides lower average packet delay, for a wide range of traffic load, compared to the other
Chapter 8. Conclusion

schemes. Simulation results underline that the performance of the algorithms depends on the average number of ONU transmitting on each wavelength in a cycle time ($N/W$). We also investigated the effect of the overhead due to guard band time on the performance of the algorithms. We can conclude that, for a fixed value of $T_{\text{cycle}}$, when the value of $N/W$ increases, the overhead increases, and in turn the average packet delay becomes higher. Future works in the field of DBWA algorithm for hybrid WDM/TDM PON include the investigation of the impact of the lasers tuning time on the performance of such allocation schemes.

Then, we proposed DBWA algorithms that solve the problem of managing the traffic originated by transceivers with different characteristics (e.g., different tuning time). Starting from two existing algorithms (i.e., EFT and EFT-VF) we proposed two new DBWAs that take into account the laser TT considering that different transceivers at the ONU can have different values of TT. These DBWAs are able to manage at the same time transmissions of transceivers which cannot retune (e.g., fixed tuned lasers), transceivers with a $TT = 0$ (e.g., array of fixed tuned lasers), and transceivers with $TT > 0$ (e.g., tunable lasers). We refer to these solutions as EFT+TT and EFT-VF+TT. We compared the performance of EFT+TT, EFT-VF+TT, and two simplified version that are not aware that ONUs can have different TT values. Our results show that the EFT-VF+TT is the solution, among the three considered, that provides the lower average packet delay. Results show also that in the case of short-reach scenarios (about 20 km) it is more important to use a DBWA that is aware of the different values of TT in the network, while in a long reach-scenario (up to 100 km) Simple EFT-VF+TT can be used without having a significant loss in terms of average packet delay.

Then we proposed a new DBWA that we called EsT-VF+$TT_{ij}$ that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength that they have to retune to. Our results show that the proposed TT-aware DBWA is able to reduce the overall average packet delay of the network, compared to the case where a DBWA that is not specifically designed for this scenario is applied.

Also, we considered the case where the arrays of fixed tuned lasers are equipped with a number of lasers which is lower than the total number of wavelengths available in the network. According to our results, we can conclude that lowering the number of lasers in the
arrays of lasers of ONU’s negligibly increases the average packet delay, this difference in terms of average packet delay is very limited (i.e., less than 2 µs).

We also evaluated if, in a FMC access-aggregation network, it is more convenient to isolate the different types of traffic over different sets of wavelengths (i.e., VPON case) or to mix the traffic types over all the available wavelengths (i.e., converged TWDM case). To model such comparison we developed ILP formulations for both cases, based on the classical bin-packing problem. Through numerical evaluation we investigate the trade-off between the VPON case and the converged TWDM case, under different network scenarios. We can conclude that, in terms of total number of wavelengths assigned the converged TWDM case is always slightly better (i.e., about 12%) than the VPON case. However, to use VPONs provides several benefits: i) At the OLT-side, there is a simplification of the traffic management method. ii) At the data-link layer, using VPONs to segregate the different types of traffic it is not necessary to have a convergent protocol in the access-aggregation segment of the network. iii) At the ONU-side, by using several VPONs in the same network it is possible to install cheaper transceivers (e.g., tunable lasers with smaller tuning range). As a future works we plan to quantitatively evaluate these benefits mentioned above to perform a more accurate comparison between VPON and converged TWDM case.

In this part of the work we also discussed a specific and penalizing network scenario, where the utilization of VPONs is not suitable due to scarce fit of required ONU capacities in the wavelength capacity. We plan to generalize such analysis using worst-case bin packing problem analysis.

Finally, we also studied multi-operator network sharing techniques in FMC networks. The main goal was to propose a collaboration procedure able to increase the total amount of traffic served (i.e., the traffic served by all the operators involved in the network sharing procedure). At the same time, we wanted to maintain the fairness for the users of the operators: the decision procedure wants to preserve some capacity of the network of operator in order to avoid to refuse the connection to users of its own network, while serving as much traffic as possible of the other operator. Moreover, with our collaboration strategy, the amount of traffic of other operators that one operator accepts to serve change dynamically according to the traffic load of all the players involved in the game (i.e., decision procedure). For
these reasons, we proposed a game theoretic model used by two players (i.e., two operators) to decide whether collaborate or not depending on the traffic load condition. Results show that our GT collaboration model is able to improve the amount of traffic served by the operators with respect to the case where operators not collaborate. However, the GT collaboration have the lowest amount of traffic lost due to the collaboration (i.e., traffic of an operator which can not be served because the capacity of the operator is filled by the traffic of other operators) with respect to all the other collaboration methods evaluated in this work. At the same time, with the GT collaboration the amount of traffic lost by the operator serving the excess traffic of another operator is only slightly higher than the case where no collaboration occurs. This means that the highest portion of the traffic lost by the operator serving the traffic of the competitor is lost due to the condition of the offered traffic, and not to the collaboration.

8.1 Future Works

In this work we have studied and proposed methods to efficiently manage the capacity of the network in an FMC scenario where both fixed and mobile traffic have to be managed over the same network infrastructure. Efficient methods to manage the capacity are needed to provide to the end users a good Quality of Experience (QoE), i.e., low delay, high bit rates, seamless connectivity to the network. In order to improve more and more the QoE of the users of the FMC network several future works have to be investigated. Among these future works, one is to evaluate methods to efficiently manage the different types of traffic in a unified manner, considering also new technologies that will be backhauled over the same LR TWDM PON infrastructure. Among these new types of traffic, we have the traffic generated by the Internet of Things (IoT). Internet of Things (IoT) is the network of physical objects embedded with electronics, software, sensors, and connectivity to enable objects to exchange data with the manufacturer, operator, and other connected devices. The Internet of Things allows objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more direct integration between the physical world and computer-based systems, and resulting in improved efficiency, accuracy and economic benefit. Since the applications of the IoT are very extensive, this application is expected to provide a huge amount of traffic to the FMC
8.1. Future Works

Moreover, since these applications are also very diverse (i.e., environmental monitoring, energy management, transportation system management, military applications, etc.) the traffic provided by IoT to the network also have the diverse requirements in terms of QoS (e.g., different types of delay). Another type of traffic that has to be considered is the next generation mobile access technology: the 5G. The main requirements of the 5G are:

1. A super-efficient mobile network that delivers a better performing network for lower investment cost. It addresses the mobile network operators pressing need to see the unit cost of data transport falling at roughly the same rate as the volume of data demand is rising.

2. A super-fast mobile network comprising the next generation of small cells densely clustered together to give a contiguous coverage over at least urban areas and gets the world to the final frontier for true "wide area mobility". It would require access to spectrum under 4 GHz perhaps via the world’s first global implementation of Dynamic Spectrum Access.

3. A converged fibre-wireless network that uses, for the first time for wireless Internet access, the millimeter wave bands (20 - 60 GHz) so as to allow very wide bandwidth radio channels able to support data access speeds of up to 10 Gbit/s. The connection essentially comprises "short" wireless links on the end of local fiber optic cable. It would be more a nomadic service (like WiFi) rather than a wide area mobile service.

These requirements can be addressed with the development of new methodologies to manage the bandwidth in the FMC network. The main questions that need to be answered in this sense are:

- It will be possible to mix all the existing types of traffic over all the wavelengths in order to meet the requirements of all these new types of traffic?

- It will be better to mix or to segregate over different wavelengths the types of traffic according to some particular parameter (e.g., requirements of delay, capacity requirements, etc.)?

- Which dynamic allocation methods will be the more recommended in order to allow the implementation of Dynamic Spectrum Access?
Bibliography


Bibliography


[40] N. Antunes, C. Fricker, P. Robert, and J. Roberts, “GATE-Driven Dynamic Wavelength and Bandwidth Allocation for WDM EPONs”, in proc. of GLOBECOM 2010, 6-10 Decembre 2010, Miami (FL), USA.


List of Figures

1.1 Reference topology of the FMC network architecture. ........................................ 3
2.1 TDM PON configuration. ................................................................. 10
2.2 WDM PON configuration. ................................................................. 12
2.3 TDM/WDM PON: first AWG/SPLITTER configuration. ......................... 13
2.4 TDM/WDM PON: second AWG/SPLITTER configuration. ....................... 14
2.5 PON generations. ................................................................................. 15
2.6 Coexistence of PON and XG-GPON1. ................................................... 18
2.7 Reach extension of LR PON. ................................................................. 20
2.8 Branch and Tree architecture. .............................................................. 21
2.9 Ring and Spur architecture. ................................................................. 22
3.1 Basic idea of the IPACT algorithm. ....................................................... 33
3.2 Basic idea of the Single-Thread algorithm. ............................................. 34
3.3 Basic idea of the Multithread algorithm. ................................................ 34
3.4 Basic idea of the EFT algorithm. .......................................................... 35
3.5 Basic idea of the EFT-VF algorithm. ..................................................... 37
3.6 Connections of the elements for the original IPACT project [36]. ............... 39
3.7 Connections of the elements for the modified simulator used in this work. ........ 40
4.1 Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded. .... 47
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded when the Void Filling solution is used to schedule both first and second transmission.</td>
<td>48</td>
</tr>
<tr>
<td>4.3</td>
<td>Operation of the EFT-VF Algorithm with Partial Void Filling Multi-threaded when the first part of the transmission is scheduled through the Void Filling algorithm, while for the second the EFT strategy is used.</td>
<td>49</td>
</tr>
<tr>
<td>4.4</td>
<td>Average packet delay comparison between EFT, EFT-VF, Multi-thread over LR WDM/TDM PON, and EFT-Partial-VF MT algorithms.</td>
<td>55</td>
</tr>
<tr>
<td>4.5</td>
<td>Average delay reduction of EFT-VF, Multi-thread over LR WDM/TDM PON, and EFT-Partial-VF MT algorithms compared to EFT scheme.</td>
<td>56</td>
</tr>
<tr>
<td>4.6</td>
<td>Average packet delay comparison of EFT-Partial-VF MT algorithm varying the value of ( N ).</td>
<td>58</td>
</tr>
<tr>
<td>4.7</td>
<td>Average packet delay comparison of EFT-Partial-VF MT algorithm varying the value of ( W ).</td>
<td>59</td>
</tr>
<tr>
<td>4.8</td>
<td>Average packet delay of EFT-Partial-VF MT algorithm in a scenario with ( W = 8, N = 128 ) ONU, and where 32 ONUs are transmitting at 250 Mbit/s and 96 ONUs are transmitting at 62.5 Mbit/s. This scenario is compared to a scenario where all 128 ONUs are transmitting at 62.5 Mbit/s (Balanced scenario).</td>
<td>60</td>
</tr>
<tr>
<td>4.9</td>
<td>Average packet delay of EFT algorithm in network scenarios with different wavelengths.</td>
<td>61</td>
</tr>
<tr>
<td>5.1</td>
<td>Example of the transmission allocation of an ONU in ( N_1 ) where a retuning happens and there is a need to add delay to retune the laser.</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Average packet delay comparison among EFT+TT, EFT-VF+TT, and Simple EFT-VF+TT in Scenario 1 where the ONUs in ( N_1 ) have a TT = 100 ( \mu )s. The ONUs-OLT distance distribution is between 500 m and 100 km.</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Average delay reduction of the Simple EFT-VF+TT and the EFT-VF+TT algorithms compared to EFT+TT scheme, where the ONUs in ( N_1 ) have a TT = 100 ( \mu )s, in Scenario 1.</td>
<td>77</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>5.4</td>
<td>Average delay reduction of the Simple EFT-VF+TT and the EFT-VF+TT algorithms compared to EFT+TT scheme, where the ONUs in N1 have a TT = 100 $\mu$s, in Scenario 3.</td>
<td>78</td>
</tr>
<tr>
<td>5.5</td>
<td>Average delay of the EFT+TT when the ONUs in N2 have different limited tunability, compared to the case where the ONUs in N2 have full tunability. The ONUs in N1 have a TT = 100 $\mu$s.</td>
<td>81</td>
</tr>
<tr>
<td>5.6</td>
<td>Average delay of the EFT-VF+TT when the ONUs in N2 have different limited tunability, compared to the case where the ONUs in N2 have full tunability. The ONUs in N1 have a TT = 100 $\mu$s.</td>
<td>82</td>
</tr>
<tr>
<td>5.7</td>
<td>Average delay of the Est-VF+TT\textsubscript{ij} algorithm, compared to the EFT-VF+TT algorithm applied in the Scenario 1 where tunable lasers have different TTs according to the number of wavelengths they have to retune. The ONUs in N1 have a TT\textsubscript{base} = 100 $\mu$s.</td>
<td>83</td>
</tr>
<tr>
<td>6.1</td>
<td>Example of the VPON scenario in a ring-and-spur metro-access network.</td>
<td>91</td>
</tr>
<tr>
<td>6.2</td>
<td>Comparison in the total number of wavelengths used between the VPON case and TWDM-PON case, in the network scenarios investigated.</td>
<td>95</td>
</tr>
<tr>
<td>6.3</td>
<td>Comparison in the number of wavelengths per spur between the VPON case and TWDM-PON case, in the network scenarios investigated.</td>
<td>96</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparison in the total number of wavelengths used between the VPON case and TWDM-PON case, in a penalizing network scenario.</td>
<td>98</td>
</tr>
<tr>
<td>6.5</td>
<td>Comparison in the number of wavelengths per spur between the VPON case and TWDM-PON case, in a penalizing network scenario.</td>
<td>99</td>
</tr>
<tr>
<td>7.1</td>
<td>FMC network infrastructure considered for the offloading scenario.</td>
<td>107</td>
</tr>
<tr>
<td>7.2</td>
<td>Average traffic served by Player 1 using the different collaboration strategies.</td>
<td>117</td>
</tr>
<tr>
<td>7.3</td>
<td>Average traffic served by Player 2 using the different collaboration strategies.</td>
<td>118</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>7.4</td>
<td>Average traffic of Player 1 served by Player 2 using the different collaboration strategies</td>
<td>119</td>
</tr>
<tr>
<td>7.5</td>
<td>Average traffic of Player 2 served by Player 1 using the different collaboration strategies</td>
<td>120</td>
</tr>
<tr>
<td>7.6</td>
<td>Average traffic lost of Player 1 using the different collaboration strategies</td>
<td>121</td>
</tr>
<tr>
<td>7.7</td>
<td>Average traffic lost of Player 1 using the different collaboration strategies</td>
<td>121</td>
</tr>
</tbody>
</table>
2.1 Basic features of GPON and EPON standards . . . . . . 16
3.1 Definition of variables . . . . . . . . . . . . . . . . . . 42
4.1 Definition of variables . . . . . . . . . . . . . . . . . . 65
4.2 Network parameters for different network scenarios . . 66
4.3 Network parameters of different network scenarios for the evaluation of the EFT algorithm . . . . . . . . . . . . . . 66
5.1 Definition of variables . . . . . . . . . . . . . . . . . . 85
5.2 Network parameters for different network scenarios . . 86
6.1 Definition of parameters and variables . . . . . . . . . . 92
6.2 Network parameters for different network scenarios . . 94
6.3 Network parameters for the particular case evaluated . 100
7.1 Definition of parameters and variables . . . . . . . . . . 110
7.2 Table of the proposed game . . . . . . . . . . . . . . . . 111
7.3 Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 60, E_{x_1} = -40,$ $E_{x_1} = 0$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
7.4 Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 120, E_{x_1} = 20,$ $E_{x_1} = 30$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
7.5 Table of the proposed game where the payoffs of the players are computed with $C_1 = 100, L_1 = 140, Ex_1 = 40$.

7.6 Table of the proposed game with Partial Accept.

7.7 Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100, L_1 = 60, Ex_1 = -40, Ex_1 = 0$.

7.8 Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100, L_1 = 120, Ex_1 = 20, Ex_1 = 30$.

7.9 Table of the proposed game with Partial Accept where the payoffs of the players are computed with $C_1 = 100, L_1 = 140, Ex_1 = 40, Ex_1 = 30$. 