Design of Energy-Efficient Next-Generation Telecommunications Networks

Sandu Padmapanie Abeywickrama

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

March 2016

Abstract

OVER the past 35 years, optical telecommunications networks have revolutionised almost every aspect of human interactions. The recent popularity of bandwidth-intensive applications and deployment of virtual private networks which span across the globe have resulted in the need for large amounts of data to be transported over the telecommunications networks. Currently, fifth-generation optical communication networks are widely deployed in order to cater for this increasing bandwidth requirement. Nonetheless, the sudden growth in information and communications technology has had a profound impact on the environment due to increasing electricity consumption and its non-negotiable contribution to global green house gas emissions. Therefore, energy-efficiency is becoming a critical factor in the design of next-generation telecommunications networks.

This thesis presents a series of novel solutions to address the increasing energy consumption of telecommunications networks. The proposed solutions focus on three important segments of the telecommunications network, namely the optical fibre access network, wireless access network, and optical core network. First, an energy-efficient approach to enhance video-on-demand services over optical access networks is proposed. To date, video-on-demand has been identified as one of the top traffic contributors to the telecommunications network. Moreover, it is expected to become more popular in the foreseeable future. Therefore, a more scalable and energy-efficient video-on-demand delivery mechanism should be considered in the design of next-generation telecommunications networks. We propose a novel VoD solution that exploits the use of a local storage caching server which stores a collection of the most popular videos closer to the customer. In this work, we discuss the opportunities for quality-of-service improvements

and their respective power consumption and cost trade-offs.

Second, we focus on the wireless access network segment. The wireless access network connects mobile users to the base stations via the use of radio waves in the ultra high frequency spectrum of the electromagnetic spectrum. In today's cellular networks, due to the large number of base stations being deployed, the power consumption of these base stations is identified as the main contributor to the network power requirement, contributing to about 60% of the network power usage. We present an energy-efficient transmit power control mechanism considering a heterogeneous wireless access network and present two approaches to solve the transmit power control problem. Our solutions exploit dynamic transmit power control of the base stations considering temporal variations of the daily traffic demand. In this work, we present the simulation results and discuss opportunities to reduce the power consumption of the wireless network.

Finally, we bring our focus towards the core network segment. The core network aggregates traffic from the access network segment and facilitates communication across large distances. Thus, the survivability of the network to maintain operation against failures in network equipment and infrastructure becomes an important parameter. We present an energy-efficient approach to enhance survivability of the core networks by exploiting the dual-homing architecture in the access network. In particular, we propose a protection scheme that protects core network data transmission against fibre and equipment failures and discuss some energy saving opportunities in the core network.

Overall, the technical contribution presented in this thesis provides a series of energy-efficient solutions towards designing an energy-efficient next-generation telecommunications network. We also discuss some of the future research directions and interesting problems arising from our efforts.

Declaration

This is to certify that

1. The thesis comprises only my original work towards the PhD,
2. Due acknowledgement has been made in the text to all other material used,
3. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.
4. The chapters 1, 3, and 5 of this thesis have been professionally proofread.

Sandu Padmapanie Abeywickrama, March 2016

Acknowledgements

"If I have seen further, it is by standing on the shoulders of giants."

~ Sir Isaac Newton

I would like to convey my sincere gratitude to everyone who helped me in numerous ways to make my PhD candidature an enjoyable one. First and foremost, I would like to thank my supervisory committee: Associate Professor Elaine Wong, Professor Christina Lim, and Associate Professor Brian Krongold. My research would not have been a success if not for your supportive efforts. I specially thank Associate Professor Wong for her patience, technical expertise and for providing continuous opportunities for my professional development. I am truly honoured and privileged to have worked with such a passionate and dedicated researcher.

During my PhD candidature, I was fortunate to collaborate with the Royal Institute of Technology (KTH), Kista, Sweden. I convey my special thanks to Professor Lena Wosinska, Associate Professor Paolo Monti, Associate Professor Jiajia Chen, and Dr. Marija Furdek of KTH for being my mentors during my visit to Kista. I further thank them for their technical expertise, valuable comments and constructive feedback on my research. I also would like to extend my thanks to Dr. Avishek Nag of the Centre for Telecommunication Value-Chain Research, Dublin, Ireland for the helpful discussions and comments.

I also extend my gratitude to Dr. Chien Aun Chan of the Centre for Energy-Efficient Telecommunication in the University of Melbourne, Dr. Tharaka Samarasinghe, Dr. Chamil Jayasundara, Dr. Chathurika Ranaweera, Pubuduni Imali Dias, and Bhagya Amarasekara for the valuable discussions and constructive feedback on my work. My special thanks go to my colleagues, friends, and administrative staff of the Department of Electrical and Electronics Engineering for their support throughout my candidature. I also would like

to thank the University of Melbourne and National Information and Communications Technology Australia (NICTA) for providing the resources to make my research a success. I also express my salute to the Sri Lankan education system for providing me the strong technical foundation that was required for my graduate studies.

Last but not least, I express my heartfelt gratitude to my parents Mr. Sumathipala and Mrs. Sunitha Abeywickrama, for their truly unconditional love. I also thank my sister Mrs. Banie Abeywickrama Siriwardena and my brother Mr. Thuthimal Abeywickrama and their families for the limitless support. Thank you all.

To My Nephew, Zukith Skylle Siriwardena.

Contents

1 Introduction **1**
- 1.1 Evolution of Optical Communication Systems . . . 1
- 1.2 Motivation towards Energy-Efficient Optical Networks . . . 4
- 1.3 Outline of the Thesis . . . 7
- 1.4 Publications . . . 13
  - 1.4.1 Publications arising from Chapter 2 of the Thesis . . . 13
  - 1.4.2 Publications arising from Chapter 3 of the Thesis . . . 14
  - 1.4.3 Publications arising from Chapter 4 of the Thesis . . . 14

2 Video-on-Demand over Passive Optical Networks **15**
- 2.1 Introduction . . . 15
- 2.2 Passive Optical Networks (PONs) . . . 17
- 2.3 Video-on-Demand Architectures . . . 22
- 2.4 Optimising Video-on-Demand Delivery - Related Work . . . 25
- 2.5 VoD Delivery Over Local Storage based PONs . . . 36
  - 2.5.1 Dual-Receiver Dynamic Bandwidth Allocation (DR-DBA) Algorithm 38
  - 2.5.2 Single-Receiver Dynamic Bandwidth Allocation (SR-DBA) Algorithm 42
  - 2.5.3 Single-Receiver Dual-Channel Dynamic Bandwidth Allocation (SRDC-DBA) Algorithm . . . 45
  - 2.5.4 Model Evaluation . . . 48
  - 2.5.5 Results and Discussion . . . 53
- 2.6 Summary . . . 61

3 Green Heterogeneous Cellular Networks **63**
- 3.1 Introduction . . . 63
- 3.2 Wireless Access Networks . . . 65
  - 3.2.1 Wireless Access Network Technologies . . . 66
  - 3.2.2 Heterogeneous Networks and Their Challenges . . . 68
- 3.3 Energy-Efficiency of Heterogeneous Networks . . . 73
- 3.4 Related Work . . . 80
- 3.5 Heuristic-based Transmit Power Estimation for Small Cell Networks . . . 84
  - 3.5.1 Centralised Small Cell Power Control Algorithm . . . 85
  - 3.5.2 Model Evaluation . . . 88
  - 3.5.3 Results and Discussion . . . 90
- 3.6 Transmit Power Estimation for Small Cell Networks: Linear Program . . . 95

3.6.1 Problem Formulation . . . 95
3.6.2 Model Evaluation . . . 99
3.6.3 Results and Discussion . . . 100
3.7 Summary . . . 105

4 Survivable Core Networks **107**
4.1 Introduction . . . 107
4.2 Wavelength Division Multiplexing (WDM) Networks . . . 109
4.2.1 ITU Wavelength Grid . . . 110
4.2.2 The Evolution of WDM Networks . . . 111
4.2.3 Lightpaths . . . 113
4.2.4 WDM Mesh Networks . . . 115
4.2.5 Fault Management in WDM Mesh Networks . . . 119
4.3 Dual-Homing (DH) Architecture . . . 123
4.4 Related Work . . . 126
4.5 Dedicated Path Protection with Dual Homing (DPP-DH) Approach . . . 129
4.5.1 Dedicated Path Protection with Dual Homing Considering Core Link Failures (DPP-DH-*l*) . . . 131
4.5.2 Dedicated Path Protection with Dual Homing Considering Core Node and Link Failures (DPP-DH-*n*) . . . 133
4.5.3 Model Evaluation of DPP-DH . . . 136
4.5.4 Results and Discussion . . . 140
4.6 Summary . . . 146

5 Conclusion **149**
5.1 Summary of Contributions . . . 149
5.2 Future Directions . . . 153
5.2.1 Energy- and Cost-Efficient VoD Systems with Cloud Infrastructure 153
5.2.2 Energy Harvesting Small Cell Networks . . . 154
5.2.3 Considering the Sleep-Mode Features for Energy-Efficient Core Network Survivability . . . 155
5.3 Summary of Thesis . . . 156

List of Tables

2.1 EPON class of service [1]. 20
2.2 Network and Protocol Parameters. 49
2.3 Equipment Specifications. 50
2.4 Costs of considered components. 54
2.5 Percentage of incremental deployment cost considering duct reuse and greenfield deployment of LS based PON. 61

3.1 LTE quality of service (QoS) classes. 67
3.2 Comparison of green cellular network approaches. 79
3.3 Parameters considered in the evaluation of the heuristic-based algorithm. 90
3.4 Parameters considered in the evaluation of the linear program. 99

4.1 Simulation parameters. 138
4.2 Considered power consumption values of the components/elements of the core network. 140

List of Figures

1.1 Recent trends in number of devices connected to the Internet [2]. 6

2.1 Network architecture of a passive optical network (PON). 18
2.2 Centralised server based VoD system architecture. 23
2.3 Distributed server based VoD system architecture. 23
2.4 Peer-to-peer based VoD system architecture. 25
2.5 Cloud based VoD system architecture. 26
2.6 Enhanced PON architecture with LS [3]. 34
2.7 Video request arrivals at the VoD server [4]. 36
2.8 Local Storage based PON architecture for VoD delivery [4]. 38
2.9 Format of the REPORT control frame with a modified field for the Requested Video ID. 39
2.10 Format of the GATE control frame with a modified field for the VoD content information. 40
2.11 Timing diagrams to illustrate communication between the OLT, LS, and ONU when the (a) central office services the request and (b) local storage services the request. 41
2.12 DR-DBA receiver configuration with dual receivers. 42
2.13 Flow chart of the DR-DBA algorithm. 43
2.14 Single receiver configuration for SR-DBA algorithm. 45
2.15 Receiver configurations used for DR-DBA, SR-DBA, and SRDC-DBA. . . . 54
2.16 Delay as a function of number of active ONUs. 55
2.17 Jitter as a function of number of active ONUs. 56
2.18 Maximum achievable aggregated throughput on λ_D and λ_S. 57
2.19 Power consumption per customer and percentage power increment as a function of number of active ONUs. 57
2.20 Power consumption per ONU. Equipment-based contribution break-down for 1 OLT line card servicing 32 ONUs and 32 OLT line cards servicing 1024 ONUs. 59
2.21 Cost per ONU. Equipment-based cost breakdown for 32 OLT line cards servicing 1024 ONUs. 60

3.1 Wireless access network architecture. 65
3.2 (a) Macrocellular network (b) Heterogeneous network. 70
3.3 Breakdown of energy consumption in cellular networks (Source: Vodafone [5]). 75

3.4 Considered single macro cell access network segment. 84
3.5 Hourly capacity demand variation per user during a day [6]. 89
3.6 Total transmit power versus number of users under varying capacity demand throughout the day. 91
3.7 Total transmission power versus number of users and number of deployed small cells during peak traffic hour. 92
3.8 Total transmission power under urban, sub-urban and rural cases during the day. 93
3.9 Percentage transmission power saving. 94
3.10 Average transmit power per small cell as a function of number of cells deployed. 95
3.11 Total transmission power versus number of users under varying capacity demand throughout the day. 101
3.12 Total transmission power versus number of users and number of deployed small cells during peak traffic hour. 102
3.13 Total transmission power under urban, sub-urban, and rural cases during the day. 103
3.14 Percentage transmission power savings. 103
3.15 Average transmit power per small cell as a function of number of cells deployed. 104

4.1 An example of dedicated path protection over a dual-homed network. . . 109
4.2 The ITU wavelength grid with 100 GHz channel spacing [7]. 110
4.3 A four-channel point-to-point WDM transmission system [8]. 111
4.4 A typical WDM optical network covering access to long haul. 113
4.5 (a) SDH/SONET node (b) WDM node (adapted from [8]). 114
4.6 A wavelength routed WDM mesh network. 115
4.7 Different schemes for surviving failures in WDM mesh networks [9]. . . . 123
4.8 Dual-homed architecture at the access segment. 124
4.9 An example protected end-to-end connection showing different possible working/backup path options through the metro/core network. 125
4.10 The Metro/Core node architecture proposed by DISCUS FP7 project [10]. 126
4.11 Core network architecture with dual-homed access network. 130
4.12 Dual-homed LE architecture with link disjoint paths and link/node disjoint paths. 134
4.13 Network topology of Ireland. 137
4.14 Average path lengths for dual-homed and baseline architectures using DPP-DH-*l*-*W* and DPP-DH-*n*-*W* (number of active wavelengths minimised) for normalised total traffic load of 0.5. 141
4.15 Average path lengths for dual-homed and baseline architectures using DPP-DH-*l*-*L* and DPP-DH-*n*-*L* (path length minimised) for normalised traffic load of 0.5. 142
4.16 Number of wavelengths used by algorithm variants on dual-homed and baseline architecture under different network loads. 143

4.17 Percentage of connections in different availability categories from DPP-DH-*l*-*W* and DPP-DH-*n*-*W*. 144
4.18 Percentage of connections in different availability categories from DPP-DH-*l*-*L* and DPP-DH-*n*-*L*. 144
4.19 Percentage of reduction in number of used wavelength segments against the baseline approach of each variant. 145
4.20 Percentage of power savings of the DPP-DH algorithm variants under different network loads. 146

Acronyms

3GPP	Third-Generation Partnership Project
ARP	Allocation and Retention Priority
BE	Best Effort
BS	Base Station
CAGR	Compound Annual Growth Rate
CBD	Central Business District
CO	Central Office
CoS	Class of Service
CPSA	Clustering-based Power Saving Algorithm
CU	Cost Unit
DBA	Dynamic Bandwidth Allocation
DC	Difference of Convex
DH	Dual-Homing / Dual-Homed
DPP	Dedicated Path Protection
ertPS	Extended Real Time Pooling Service
FBS	Femtocell Base Station
FDMA	Frequency Division Multiple Access
FL	FBS Leader
FM	FBS Member
GHG	Greenhouse Gases
GBR	Guaranteed Bit Rate
GPON	Gigabit Passive Optical Network
ICIC	Inter-cell Interference Coordination
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronic Engineers

ITU	International Telecommunications Union
LAN	Local Area Network
LE	Local Exchange
LLID	Logical Link Identity
LOS	Line of Sight
LP	Linear Program
LS	Local Storage
LTE	Long-Term Evolution
M2M	Machine-to-Machine
MAC	Media Access Control
MACPDU	Medium Access Control Packet Data Unit
MIMO	Multiple-Input and Multiple-Output
MPCP	Multi Point Control Protocol
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
nLOS	Near-line of Sight
NLOS	Non-line of Sight
nrtPS	Non-real Time Pooling Service
OBS	Optical Burst Switching
O-E-O	Optical-Electrical-Optical
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLT	Optical Line Terminal
OPS	Optical Packet Switching
OXC	Optical Crossconnect
P2P	Peer-to-Peer
PAE	Power Added Efficiency
PON	Passive Optical Network
PPP	Poisson Point Process
QCI	Quality of Service Class Identifier
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency

RG	Residential Gateway
rtPS	Real Time Pooling Service
RWA	Routing and Wavelength Assignment
SAP	Small cell Access Point
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCN	Small Cell Network
SINR	Signal to Interference and Noise Ratio
SLA	Service Level Agreement
SON	Self-Organising Network
SS	Subscriber Station
STB	Set-top-box
TDMA	Time Division Multiple Access
UGS	Universal Grant Service
UHF	Ultra High Frequency
VoD	Video-on-Demand
WAN	Wide Area Network
WCDMA	Wideband Code Division Multiple Access
WDM	Wavelength Division Multiplexing
WR	Wavelength Routing

Chapter 1
Introduction

A communications system transmits information from one location to another, whether separated by a few kilometres, or transoceanic distances. Information is often transmitted by electromagnetic carrier wave whose frequency can vary from a few megahertz to a several hundred terahertz. Optical communication systems use higher carrier frequencies ($\sim$ 100 THz) in the visible or near-infrared region of the electromagnetic spectrum. Such optical communication systems are referred to as lightwave systems to distinguish them from microwave systems, whose carrier frequency is typically smaller by five orders of magnitude ($\sim$ 1 GHz). Fibre optic communication systems are lightwave systems that employ optical fibres for data transmission. Such optical systems have been deployed worldwide since 1980 and have revolutionised the field of telecommunications.

1.1 Evolution of Optical Communication Systems

The first-generation of lightwave systems operated near 0.8 μm, used GaAs semiconductor lasers and became commercially available in 1980 [11]. These systems considered repeater spacings of up to 10 km and operated at a bit rate of 45 Mbps. The larger 10 km repeater spacing compared with the 1 km spacing of the coaxial systems was an important motivation mainly because it reduced installation costs associated with each repeater. Subsequently, it was clear that the repeater spacing could be further increased by operating the lightwave system near 1.3 μm, where fibre loss is less than 1 dB/km. Moreover, optical fibres exhibit minimum dispersion in this wavelength region. This realisation led

to a worldwide effort in developing InGaAsP semiconductor lasers and detectors operating near 1.3 μm. The second-generation of lightwave systems became available shortly after, in the early 1980s. However, the bit rate of these early systems was limited to below 100 Mbps due to the dispersion of multi-mode fibres [12]. This limitation was later overcome by the introduction of *single-mode* fibre. A laboratory experiment conducted in 1981 demonstrated 44 km of single-mode fibre transmitting at 2 Gbps [13]. The introduction of commercial systems with single-mode fibre soon followed. By 1987, second-generation lightwave systems, operating at bit rates of up to 1.7 Gbps with a repeater spacing of about 50 km, became commercially available.

The repeater spacing of second-generation lightwave systems was limited by fibre losses at the operating wavelength of 1.3 μm (typically at 0.5 dB/km). Silica fibre loss is minimum near 1.55 μm. Indeed, a 0.2 dB/km loss was realised in 1979 in this spectral region [14]. However, the introduction of third-generation lightwave systems operating at 1.55 μm was considerably delayed due to large fibre dispersion near 1.55 μm. Conventional InGaAsP semiconductor lasers could not be used due to the occurrence of pulse spreading as a result of simultaneous oscillation of several longitudinal modes. This dispersion problem was eventually overcome either by using dispersion-shifted fibres designed to have minimum dispersion near 1.55 μm or by limiting the laser spectrum to a single longitudinal mode. By 1985, laboratory experiments demonstrated the possibility of transmitting information at bit rates of up to 4 Gbps over distances exceeding 100 km [15]. Third-generation lightwave systems operating at 2.5 Gbps became commercially available in 1990. These systems were capable of operating at a bit rate of up to 10 Gbps [16]. The best performance was achieved using dispersion-shifted fibres in combination with lasers oscillating in single longitudinal mode. The main drawback of the third-generation 1.55 μm systems was the need to periodically regenerate the signal using electronic repeaters spaced apart typically by 60 - 70 km. This limitation was eventually addressed by the fourth-generation lightwave systems.

The fourth-generation lightwave systems make use of *optical amplification* for increasing repeater spacing and *wavelength division multiplexing (WDM)* for increasing the bit rate. The introduction of WDM around 1992 started a revolution that resulted in doubling

of the system capacity every 6 months or so that led to lightwave systems operating at a bit rate of 10 Tbps by 2001. In most WDM systems, fibre losses are compensated periodically using erbium-doped fibre amplifiers (EDFAs) spaced 60 - 80 km apart. EDFAs were developed in 1985 and became commercially available in 1990. In 1991, using EDFAs in a recirculating-loop configuration, an experiment showed the possibility of data transmission over 21,000 km at 2.5 Gbps, and over 14,300 km at 5 Gbps [17]. This demonstration indicated that an amplifier-based, all-optical, submarine transmission system was feasible for inter-continental communication. By 1996, transmission of over 11,300 km at a bit rate of 5 Gbps had been demonstrated by using actual submarine cables [18]. Since then, a large number of submarine lightwave systems have been deployed worldwide. For example, a 27,000 km fibre optic link around the globe (known as FLAG) became operational in 1998, linking many Asian and European countries [19]. Another major lightwave system, known as *Africa One* started operation in 2000. It circles the African continent and covers a total transmission distance of about 35,000 km [20]. Several more WDM systems were deployed across the Atlantic and Pacific oceans during 1998 - 2001 in response to the increase in Internet data traffic. Indeed, such a rapid deployment of WDM systems led to a worldwide over-capacity that resulted in the so called *telecom bubble* in 2001.

The main emphasis of most WDM lightwave systems is to increase the capacity of data transmission by transmitting more and more data channels through the WDM technique. With increasing signal bandwidth, it is often not possible to amplify all wavelength channels using a single amplifier. As a result, new amplification schemes (for example, distributed Raman amplification) were developed to cover the spectral region extending from 1.45 to 1.69 μm. In 2000, these amplification schemes led to a 3.28 Tbps experiment in which 82 channels, each operating at 40 Gbps, were transmitted over 3000 km. Within a year, the system capacity increased to nearly 11 Tbps (273 WDM channels, each operating at 40 Gbps) but the transmission distance was limited to 117 km [21]. Commercial lightwave systems with a capacity of 3.2 Tbps, transmitting 80 channels (each at 40 Gbps) with the use of Raman amplification, became available by the end of 2003.

The fifth-generation of fibre optic communication systems were motivated by the need to extend the wavelength range of the existing WDM systems. The conventional wavelength window, known as the C band, covers the wavelength range of 1.53 to 1.57 μm, was extended on both the long and short wavelength sides, resulting in the L and S bands, respectively. The Raman amplification technique can be used for signals amplification in all three wavelength bands. Moreover, a new type of fibre, known as *dry fibre*, were eventually developed such that fibre losses are small over the entire wavelength region extending from 1.3 to 1.65 μm [22]. The focus of the modern fifth-generation lightwave systems was also on improving the spectral efficiency of WDM systems. The idea is to employ advanced modulation formats in which information is encoded using both the amplitude, phase, and polarisation of the optical carrier [23]. Using such advanced modulation formats, 64 Tbps transmission was successfully demonstrated over 320 km using 640 WDM channels that spanned both C and L bands with 12.5 GHz channel spacing [24]. In another recent experiment (2012) [25], a capacity of 101.7 Tbps was achieved over 165 km using C and L bands.

Even though fibre optic communication technology is barely 30 - 35 years old, to date, it has progressed rapidly and has reached a certain stage of maturity. However, with this rapid growth, the increased energy consumption due to operating and maintaining the communication infrastructure becomes a crucial challenge, not only to network providers in terms of financial burdens, but also to the sustainability of the environment. Therefore, energy-efficiency becomes an important parameter to be carefully considered and optimised during both network design and operational stages.

1.2 Motivation towards Energy-Efficient Optical Networks

Today, traditional non-renewable energy sources such as hydrocarbon energy provide for most of the energy demand globally, for example, 70% of USA's electricity in 2015 [26]. The combustion of hydrocarbon materials releases large amounts of green house gases (GHGs) and the greenhouse effect caused by GHG emission has been recognised as the main cause of global warming [27].

Information and communication technology (ICT) takes a dual role in this scenario. On the one hand, broad and intensive use of advanced ICT applications and services promises substantial improvements in many areas such as in industry, logistics, trade, healthcare, and education. Mainly, thanks to virtualisation, information transmission, and teleconferencing ICT avoids the need to transport documents and people and, thus, decreases GHG emissions caused by transportation [28–30]. However, the ever increasing number of ICT equipment and intensive usage of ICT services is leading to a continuous increase of ICT-related energy consumption. Moreover, the short lifetime of ICT-related devices results in increased production rates, thus contributing towards producing hazardous e-waste which is harmful to the environment.

Moreover, as we have discussed earlier in this chapter, the continuous growth of Internet traffic has become a major driver to increase the capacity of network infrastructures. Reports have shown that the main contribution to Internet traffic growth is the increased traffic generated from end customers [31]. This is due to the introduction of bandwidth-hungry applications for users and the fast growth of the number of residential and mobile broadband subscribers. This trend is evident from Fig. 1.1 which compares the forecasts made in 2010 and 2014 on the number of connected devices to the Internet. Due to the concurrent growth of Internet traffic and the number of subscribers, both the number of network elements and their capacities are also expected to increase. The recent introduction and wide penetration of smart phones and tablets confirm this trend. Furthermore, according to the vision of the *Internet of Things*, it is expected that in the foreseeable future, a huge number of smart autonomous devices will communicate via the Internet. This concept is referred to as machine-to-machine (M2M) communication. As indicated in Fig. 1.1, it is projected that 50 billion of smart autonomous devices will be connected to the Internet in 2020 thus confirming the exponential growth of ICT in the near future [31].

In 2007, the carbon footprint of ICT was estimated to account for about 2% of the total carbon emissions worldwide, which is comparable to the carbon footprint of the aviation sector [32]. At that time, the estimated ICT-related electricity consumption was above 17 TWh, which corresponds to the production volumes of 15 medium-scale nuclear power

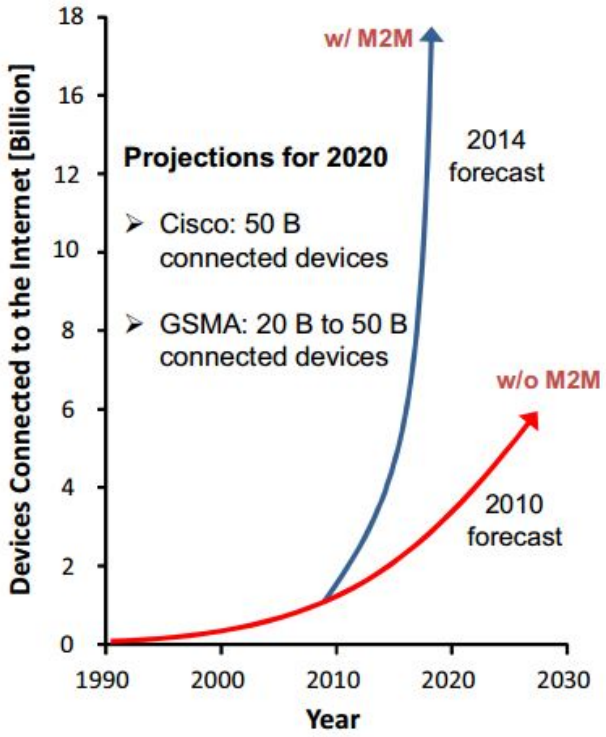


Figure 1.1: Recent trends in number of devices connected to the Internet [2].

plants today [30]. In 2020, the contribution of the ICT sector to the global carbon emissions is expected to increase to above 4% [28]. Nearly 51% of the ICT-related energy consumption is from telecommunications infrastructure and data centres. However, in [33], it has been estimated that ICT-enabled solutions can potentially reduce overall annual emissions by 16.5% of the projected total in 2020.

The thesis presents a series of novel solutions to address the increasing energy consumption of telecommunications networks. The proposed solutions focus on three important segments of the telecommunications network, namely the optical fibre access network, wireless access network, and optical core network. The next three chapters of the thesis consider each of these network segments separately. The outline of the thesis and the original contributions arising from each chapter are discussed in the next section.

1.3 Outline of the Thesis

This thesis comprises five chapters including the current Chapter 1, Introduction. In this chapter, we discuss the evolution of the optical communication systems, research objective, outline, and the original contributions of this thesis. In Chapter 2, our focus is on the optical fibre access networks, where an energy-efficient approach to enhance video-on-demand (VoD) services over passive access networks (PONs) is proposed. In Chapter 3, we focus on the wireless access network systems, where a power control mechanism for heterogeneous mobile access network is proposed. In Chapter 4, we focus on the core network segment, where an energy-efficient approach to enhance survivability of the core network is presented. Finally, we conclude the thesis in Chapter 5 giving a summary and potential future directions of our research work.

Note that each chapter concentrates on a different segment of the telecommunication network, is therefore arranged in a self-contained fashion. Each chapter contains an introduction and a detailed literature review that discuss the background, motivation, and related work relevant to the considered network segment and the conducted research work. A brief outline of each of the remaining chapters, highlighting the original research contributions, is presented next.

Chapter 2: Video-on-Demand (VoD) over Passive Optical Networks (PONs)

In Chapter 2, we optimise the VoD delivery over PONs. We believe delivering VoD service over the Internet becomes a key challenge in the design of next-generation telecommunications networks as it is already identified as the main traffic contributor in today's optical networks and is expected to grow further at an alarming rate [34]. The chapter starts with an introduction which discusses the optical wired access networks, commonly deployed VoD system architectures and a comprehensive review on the most notable related work in literature that optimises VoD delivery. We exploit the use of a local storage (LS) server to store a subset of highly demanded videos. The LS is strategically placed within the boundaries of the PON, storing the video content much closer to the customer thus improving the quality of service (QoS) attributes of VoD delivery. As the Internet is an IP network, it is by nature a best effort network[1] thus making QoS attributes such as delay, jitter and bandwidth decisive network parameters in determining the success of VoD delivery. In our work, we present three dynamic bandwidth allocation (DBA) algorithm variants considering the presence of the LS server to improve the user-perceived QoS of VoD delivery over a PON. The algorithms are simulated using packet-level simulations to accurately calculate the QoS improvements achieved by distributing the VoD content from a server located much closer to the customer. Next, we formulate a power consumption model to estimate the power consumption of the access network segment that is attributed to the VoD system. The power consumption of the three presented VoD schemes are evaluated. Moreover, we formulate a cost model to estimate the deployment cost of the LS based PON. We then present the observed results along with a discussion of the trade-off between the QoS improvements, energy-efficiency, and deployment cost of LS based VoD delivery.

[1] In a best effort IP network, all IP packets are treated in the same fashion. The network undertakes its best effort to deliver every packet as quickly as it can, but makes no undertaking to treat any class of packets preferentially to any other.

Original contributions of Chapter 2

- Proposed a dual-receiver based dynamic bandwidth allocation (termed as DR-DBA) algorithm leveraging the availability of video content at the local storage server placed closer to the customer.
- Introduced a receiver configuration which exploits the use of a single dynamically-tunable receiver module at the optical network unit (ONU).
- Proposed two single-receiver based dynamic bandwidth allocation (SR-DBA and SRDC-DBA) algorithms that uses the proposed single receiver configuration.
- Developed a packet-level simulator to demonstrate the three variants of the algorithms considering PON architectures with, and without the presence of the LS.
- Formulated mathematical models to estimate the power consumption of the video delivery scheme in order to compare the power consumption of the proposed algorithm variants and their operation.
- Developed a mathematical model to estimate the deployment cost of a local storage based passive optical network considering different deployment scenarios.
- Proved that, storing the video content in the local storage help improve the QoS performance. We then show that our SRDC-DBA algorithm shows the best QoS improvement to power consumption trade-off. Further, we show that the deployment cost of the LS based PON is dominated by the fibre-related costs and the cost contribution from the LS equipment is very marginal.

Chapter 3: Green Heterogeneous Cellular Networks

In Chapter 3, our focus is on the wireless mobile access network as the wireless mobile access network being identified as the main contributor to the network power consumption. We exploit the concept of cell zooming[2] to improve the energy-efficiency of the cellular network and propose a transmit power control mechanism for the heterogeneous

[2]Cell zooming is a network layer technique enabling dynamic adjustment of the cell size according to traffic conditions by adjusting antenna tilt angles, height, or transmit power.

wireless access network systems. The introduction discusses the most commonly recognised state-of-the-art wireless broadband access network architectures and technologies. We then provide a detailed overview on the heterogeneous mobile access network architecture. We also discuss the key aspects of heterogeneous networks, such as the coexistence of small cell BSs with existing macro cell BSs in a heterogeneous network, self-organising capability, and backhauling of the small cells in this section. Next, we discuss the related work found in literature on the energy-efficiency of heterogeneous networks with a classification of commonly used approaches to improve the energy-efficiency of heterogeneous access networks. We propose a heuristic-based transmit power control algorithm to solve the power control problem to derive the transmit power of small cell BSs of the heterogeneous network in an sub-optimal fashion. Next, we take a different approach to solve the same problem by formulating a linear program (LP) which produces a global optimal solution. However, the heuristic-based approach is computationally straightforward compared to the LP. As a result, the heuristic approach is able to model larger network segments than the LP. We show that the total transmission power of the network can be reduced by 12% and 14% by solving the transmit power control problem with the heuristic-based approach and with the LP, respectively. We compare the two approaches in terms of their scalability versus optimality and discuss the opportunities to save energy by efficient control of BS transmit power.

Original contributions of Chapter 3

- Proposed a heuristic-based algorithm to solve the optimal transmit power control problem for heterogeneous mobile access networks. The algorithm is then simulated in urban, sub-urban, and rural test environments to evaluate its performance and the optimality of the results.

- Proposed an optimisation framework based on a LP to find the global optimal solution to the same problem considered above in the heuristic-based approach. The LP is then used to find the optimal transmit powers for urban, sub-urban and rural test environment similar to the heuristic-based algorithm evaluation to maintain

comparability.

- Compared the sub-optimal results from the heuristic algorithm versus the global optimal results gathered from the LP.

- Showed that our proposed transmit power control approach provides opportunity for power savings in the heterogeneous mobile access network. The total transmission power of the network can be reduced by 12% on average as for the heuristic-based algorithm estimations whereas the LP showed a 14% reduction of total transmission power on average.

Chapter 4: Dual-Homing-based Survivability for Core Networks

In Chapter 4, we bring our focus towards the core network segment of the telecommunications network. As the core networks serve large bandwidth demands across long distances (aggregated traffic from access networks), survivability of the core network to maintain service in the event of failure of network equipment and fibre cables becomes important. For this reason, in this chapter, we exploit the dual-homing (DH)[3] capability of the access network to improve core network survivability. The chapter starts with an introduction which discusses the WDM mesh networks, their operation and fault management. Further, we discuss the DH architecture of the access network. Then, we propose a series of routing and wavelength assignment (RWA) algorithms to provide dedicated path protection for data communication through the core networks while utilising core network resources effectively. These algorithms are designed to exploit the dual-homing capability of the access networks to provide protection with different optimisation objectives. The RWA algorithms are then simulated in a country-wide optical core network to analyse the improvements to the network survivability. We also conduct an availability analysis to evaluate the proposed algorithm variants. Next, we formulate a mathematical model to estimate the power consumption of the core network. The power consumption model is used to evaluate the energy-efficiency of the proposed algorithms.

[3]Dual-homing is a concept initially introduced to ensure survivable connectivity between the access network and the metro/core network. In a dual-homed network, a local exchange, which aggregates end-user traffic from the access network is connected to two nodes from the core network.

This chapter presents a detailed analysis and a comparison of the proposed algorithms considering the energy trade-off of the proposed protection schemes.

Original contributions of Chapter 4

- Proposed the dedicated path protection with dual homing (DPP-DH) concept.
- Proposed four variants of the DPP-DH algorithm.
    1. DPP-DH-*l*-*W* algorithm which considers core link failures only and minimises the number of used wavelengths,
    2. DPP-DH-*l*-*L* algorithm which considers core link failures only and minimises the total path length,
    3. DPP-DH-*n*-*W* algorithm which considers either core node or link failures and minimises the number of used wavelengths, and
    4. DPP-DH-*n*-*L* algorithm which considers either core node or link failures and minimises the total path length.
- Developed a simulator to demonstrate the proposed algorithm variants in a country-wide optical core network to evaluate the improved survivability and the improved resource utilisation of the core network.
- Formulated a mathematical power consumption model to estimate the core network power consumption considering the DH architecture in order to evaluate the energy-efficiency of the resulting core network.
- Showed that DH capability of the access network could be exploited to provide improved survivability to the core network to protect end-to-end connections against fibre and network hardware failures. Compared to baseline network architectures without DH, the proposed DPP-DH approaches can reduce the backup path length by up to 35% on average, used up to 25% fewer wavelengths, and increased the portion of connections with 99.999% and 99.995% availability by 77% and 13%, respectively. Further, the estimated power consumption of the network showed that DPP-DH can provide up to 31% power savings.

Chapter 5 : Conclusions and Future Directions

In Chapter 5, the thesis concludes with a summary of the research work reported. Furthermore, in this chapter, we discuss potential future directions of our research.

1.4 Publications

1.4.1 Publications arising from Chapter 2 of the Thesis

Journals

- S. Abeywickrama and E. Wong, Delivery of video-on-demand services using local storages within passive optical networks, *OpticsExpress*, Vol. 21, 2083-2096, (2013).
- S. Abeywickrama and E. Wong, Dynamic bandwidth allocation algorithms for local storage based VoD delivery: Comparison between dual and single receiver configurations, *Optics Communications*, Vol. 336, Feb, (2015).

Conferences

- S. Abeywickrama and E.Wong, Delivery of VoD services on PONs using Local Storages: Analysis of QoS, *Proc. Of Asia Communications and Photonics conference (ACP)*, (2012).
- S. Abeywickrama and E. Wong, Impact on Local Storages on performance of PONs, *Proc. Of Photonics Global Conference (PGC)* , (2012).
- S. Abeywickrama and E. Wong, Single-receiver Dynamic Bandwidth Allocation (SR-DBA) algorithm for Local Storage based VoD delivery, *Proc. Of Asia Communications and Photonics conference (ACP)* , (2013).
- S. Abeywickrama and E. Wong, Single-receiver Dual-channel dynamic bandwidth algorithm for local storage VoD delivery, *Proc. Of Global Communications conference (GLOBECOM)* , (2013).

1.4.2 Publications arising from Chapter 3 of the Thesis

Conferences

- S. Abeywickrama and E. Wong, Estimation of Transmit Power for Small Cell Networks, *Proc. Of Opto-Electronics and Communications Conference*, (2014).
- S. Abeywickrama and E. Wong, Transmit Power Control for Small Cell Networks in Urban, Suburban, and Rural Environments, *Proc. Of Conf. on Optical Internet* , (2014).
- E. Wong, I. Akhtar, S. Abeywickrama, C. Ranaweera, C. Lim, and A. Nirmalathas, Towards a Framework for Small-cell Network Planning, *Progress in Electromagnetics Research Symposium (PIERS)* , (2014).

1.4.3 Publications arising from Chapter 4 of the Thesis

Journals

- S. Abeywickrama, M. Furdek, P. Monti, L. Wosinska, and E. Wong, Protecting Core Networks with Dual-Homing: A Study on Enhanced Network Availability, Resource Efficiency, and Energy-Savings, *Optics Communications*, Vol. 381, 327-335, Dec,(2016).

Conferences

- S. Abeywickrama, M. Furdek, P. Monti, L. Wosinska, A. Nag, and E. Wong, Dual-Homing Based Protection for Enhanced Network Availability and Resource Efficiency, *Proc. of Asia Communications and Photonics conference (ACP)* , (2014).
- S. Abeywickrama, E. Wong, M. Furdek, P. Monti, and L. Wosinska, Energy-Efficient Survivability for Core Networks using Dual-Homing, *Proc. of Asia Communications and Photonics conference (ACP)* , (2015).

Chapter 2

Video-on-Demand over Passive Optical Networks

This chapter focuses on the optical access network segment. We discuss the use of a local storage caching server which stores video content closer to the customer to improve the quality-of-service of video-on-demand delivery over the fibre-based passive optical access networks. The chapter also presents a power consumption analysis of the proposed video-on-demand delivery schemes and discusses the power consumption trade-off of using the local storage caching server.

2.1 Introduction

In recent years, the deployment of passive optical networks (PONs) has resulted in significant growth of on-demand services over the access segment. Data communications forecasts have highlighted that the global Internet video related traffic from all services, such as video-on-demand (VoD), peer-to-peer (P2P), etc., will be as high as 80% of all consumer Internet traffic by 2019 [35]. Providing VoD streams while maintaining the required quality-of-service (QoS) levels will be especially challenging during peak hours with potentially hundreds of customers of a single PON, watching the same video. Furthermore, research is currently being carried out to enhance the quality-of-experience (QoE) for the customer with the aid of higher resolution screens, 3D effects, and higher definition audio features. Forecasts in [35] further predict that by 2019, high definition (HD) internet video will comprise 69.9% of the global VoD, an increase from 55.8% in 2014. Therefore, it is to be expected that video file sizes will become larger, thus raising the bandwidth requirements to suit their functions. QoS requirements concerning network parameters for different classes of services provided over IP networks includ-

ing VoD have been specified in the ITU-T and ETSI recommendations [36, 37]. Further, it has been shown that customers are demanding even higher QoS attributes in studies conducted using techniques such as user opinion scoring systems [38]. Therefore, the minimum QoS requirement levels can be expected to be more stringent in years to come.

As a solution to address the increasing VoD-related Internet traffic, distributed storage systems have been proposed to achieve higher capacities and lower network transmission costs in VoD systems [39]. Additionally, studies on VoD and IPTV architectures highlight that delivering video content from strategic locations in the network in a distributed manner [40–42] can help improve the energy-efficiency of the network. Considering these benefits, the placement of a local storage (LS) server within the PON has been proposed as a viable distributed storage solution for VoD delivery over PONs [4]. In [4], the LS arrays carry a set of highly popular videos which is dynamically updated depending on customer demand using a novel last-k algorithm to measure the popularity of the considered videos. The proposed LS architecture was shown to achieve a 40% bandwidth saving in the downstream direction from the central office (CO) of the PON [4].

In this chapter, we critically study and compare the QoS attributes, power consumption, and deployment cost of PON with and without the use of the above-mentioned LS. We carry out packet level simulations for two architectures whereby VoD services are delivered with and without the presence of LS within the PON. We introduce dynamic bandwidth allocation (DBA) algorithms to optimise packet delivery in the LS PON with the aim of enhancing QoS levels. Further, we formulate power consumption models for the two architectures to analyse the additional power requirement that may have been introduced to the PON by the LS equipment. The power consumption of a LS is attributed to the processing and video storage of LS server and storage arrays respectively. The QoS and power consumption values are then critically analysed to study the trade-off between the QoS performance and the network power consumption. Furthermore, we formulate a cost model to estimate the deployment cost of local storage based PONs. We critically study the trade-off between QoS performance and the deployment costs of local storage based VoD delivery.

The rest of the chapter is organised as follows. First, in Section 2.2, we discuss the

conventional PON architecture in detail. We present a brief review of the proposed DBA algorithms for PON operation. Second, in Section 2.3, we discuss the video-on-demand (VoD) system architectures in general. Then, we review some of the most notable contributions made in literature to optimise VoD delivery in Section 2.4 . In Section 2.5, we focus on the LS-based PON architecture and propose three dynamic bandwidth allocation algorithms to optimise VoD delivery considering different receiver configurations at the ONU. We critically compare the three algorithms with respect to their end-user QoS performance and network power consumption to study the possible improvements of LS-based VoD delivery. We measure the QoS attributes by conducting packet level simulations, and the power consumption estimations by formulating mathematical models to represent the network architectures. Moreover, in this section, we present a cost model to estimate the deployment cost of the local storage based PON in an attempt to study the financial trade-off of using the local storage server. The results obtained from the packet level simulations, power consumption model, and cost model are presented and discussed in detail in this section. Finally, a summary of the chapter is presented in Section 2.6.

2.2 Passive Optical Networks (PONs)

As the name implies, a passive optical network (PON) consists of a collection of passive elements in the optical path between the central office (CO) and its end user terminals. A typical architecture of a PON is presented in Fig. 2.1. A central controller which is termed optical line terminal (OLT) is placed at the CO. As shown in Fig. 2.1, fibre emanating from the OLT is split into multiple (N) paths using a 1:N passive optical splitter. Here, the parameter N represents the number of connected end user terminals. Each of the split paths are terminated at an optical network unit (ONU) placed near the customer premises. In general, such fibre based passive access network architectures are refereed to as fibre-to-the-X (FTTX) architectures. Where, X signifies the end terminal of the optical path. For example, if the optical fibre is terminated at the curb, at a building, or at a house premises, the architecture is termed as FTTC, FTTB, and FTTH, respectively.

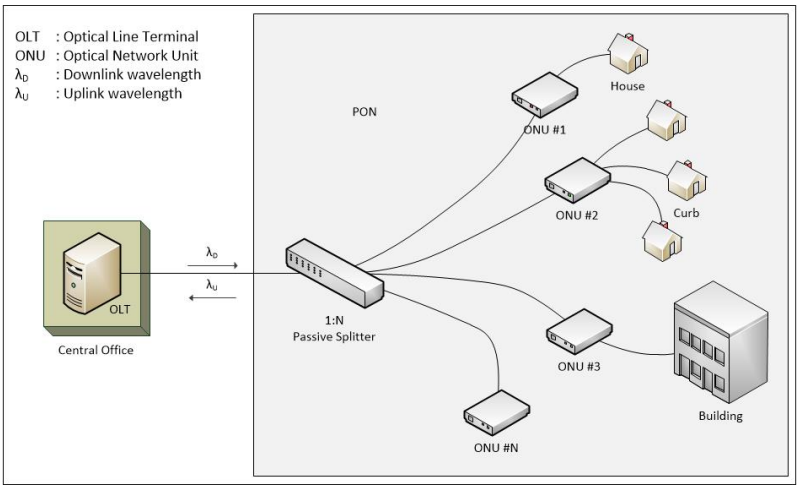


Figure 2.1: Network architecture of a passive optical network (PON).

To date, many variations of PON technologies have been proposed in literature [43], such as time division multiplexing (TDM) PON, wavelength division multiplexing (WDM) PON, orthogonal frequency division multiplexing (OFDM) PON, sub-carrier multiplexing (SCM) PON, and code division multiplexing (CDM) PON. Amongst these different PON technologies, TDM-PON and WDM-PON have been considered as the most suitable candidates for commercial deployments due to their practicality for widespread deployment [44]. Specifically, TDM-PON has become very popular amongst network operators due to its cost effectiveness, ability to support a large number of users and simplistic nature in deployment [45]. Therefore, our work in this chapter is predominantly focused on the TDM-PON technology.

In TDM-PON, two distinct wavelengths are used, one for uplink and one for downlink transmission. The number of ONUs connected to the OLT depends on the power budget of the network due to the downlink transmission being equally split amongst the connected ONUs by the passive splitter. Typical split ratios in passive splitters are 1:8, 1:16, 1:32, and 1:64. Two TDM-PON technologies that have been well studied and widely deployed are Ethernet PON (EPON) and Gigabit capable PON (GPON). EPON technology is based on the Institute of Electrical and Electronics Engineers (IEEE) 802.3ah standard [46]. In a EPON, all data packets are encapsulated in Ethernet frames and are carried natively in the PON with no changes or modification. On the other hand, GPON

is standardised by International Telecommunications Union (ITU) [47]. Unlike the EPON, data are encapsulated using GPON encapsulation method (GEM) which avoids the direct transmission of the native Ethernet frames. Moreover, the latest release of EPON and GPON, which are referred to as 10GEPON [48], and XGPON [49] respectively, support data rates of up to 10 Gbps. Both of these TDM-PON standards use specifically designed resource allocation mechanisms for their uplink transmission in the shared channel, in order to avoid upstream packet collisions between the end users connected to the same PON. However, in our work, we consider the EPON technology due to the simplicity of handling Ethernet frames in the packet level simulation avoiding the extra layer of encapsulation suggested by the GPON standard.

In EPON, all ONUs share the same wavelength channel for the uplink transmission. When multiple ONUs simultaneously transmit uplink data, transmission collisions may occur. The multi point control protocol (MPCP) was introduced to eliminate these collisions [46]. MPCP introduces two standardised control frames to govern a collision-free uplink transmission. A REPORT control frame is used by the ONUs to request for uplink bandwidth from the OLT and a GATE control frame is used by the OLT to assign bandwidth to a specific ONU. On the other hand, in the downlink transmission, the OLT broadcasts the data and control messages to all connected ONUs. All active ONUs listen to the downlink channel and receive all the frames sent out from the OLT. Then, each ONU identifies the data destined to it based on the unique identifier named as the logical link identity (LLID). The LLID is assigned to each ONU by the OLT upon registration. The EPON uses differentiated class of services (CoS) to support miscellaneous quality of service (QoS) for different services. This is accomplished by using IEEE 802.1p and 802.1q (virtual bridged local area networks) standard extensions [1]. To support queue oriented QoS mechanism, an ONU maintains up to eight different priority queues. Each of these queue priorities and their usage are listed in Table 2.1 [1].

Dynamic Bandwidth Allocation for PONs

The data transmission of a TDM-PON is conducted by using two wavelength channels, one for downlink transmission, and one for uplink transmission. Uplink transmission is

Priority	Class of Service	Specification
1	Network Control	Maintain and support network infrastructure
2	Voice	Characterised by less than 10 ms delay
3	Video	Characterised by less than 100 ms delay
4	Controlled load	Admission control
5	Excellent load	BE type service, for important customers
6	—Spare—	
7	Best effort	Local area network (LAN) traffic
8	Background	Bulk transfers

Table 2.1: EPON class of service [1].

when the ONU transmits data towards the central office. Conversely, downlink transmission is when the central office transmits data towards the ONUs. Downlink transmission is broadcasted to all ONUs. Each ONU receives all the transmitted data on the downlink transmission and filters the relevant data while discarding the irrelevant data. This makes managing the downlink transmission relatively straightforward. However, in the uplink direction, due to the fact that all ONUs sharing the same wavelength channel, only a single ONU can be transmitting upstream data at a given time to avoid data collisions. For this reason, multi point control protocol (MPCP) has been introduced to manage the ONUs in taking turns using the uplink transmission. According to the MPCP, each ONU uses a REPORT message to notify its bandwidth requirement to the OLT. Depending on the requested bandwidths and the implemented bandwidth allocation mechanism, the OLT calculates the bandwidth allocation and the transmission start time for each ONU to perform the uplink transmission in the next cycle. Then, the OLT includes these two information in a GATE message and broadcasts it in the downstream. Each ONU receives a unique GATE message that includes information about its next transmission. During the allocated transmission window, each ONU sends its queued data and a REPORT message to the OLT. Note that the MPCP does not specify any particular dynamic bandwidth allocation algorithm. Instead, it is intended to facilitate the implementation of a bandwidth allocation algorithm. Many research groups [50–53] have proposed various dynamic bandwidth allocation mechanisms to improve the bandwidth utilisation of PONs. In this section, we briefly discuss some of these bandwidth allocation mechanisms.

Interleaved polling with adaptive cycle time (IPACT) [54], is one of the first band-

width allocation algorithms proposed for the EPON. IPACT consists of an interleaved polling mechanism which has the capability to adjust the PON cycle time. In IPACT, while one ONU is transmitting the uplink data, another ONU is polled to schedule the next transmission. The amount of bandwidth allocated to an ONU is dependent on the algorithm implemented in the OLT. Different algorithms that can be used in the OLT to dynamically allocate the available bandwidth for its ONUs are investigated in [55]. These algorithms are limited service, constant credit, linear credit, and elastic service. The performance of each of these algorithms is compared with each other and also with the fixed service in which all the ONUs are granted a fixed amount of bandwidth. The results showed that IPACT with limited service allocation schemes performs better than the other schemes implemented.

Another bandwidth allocation mechanism that considered differentiated services is proposed in [56]. The proposed mechanism redistributes the surplus bandwidth available from ONUs that have requested a small amount of bandwidth to those that have requested large amount of bandwidth. Moreover, the bandwidth allocation decisions for the lightly loaded ONUs are made instantaneously, whereas for the heavily loaded ONUs, these decisions are made once the OLT receives all the REPORT messages from all of its connected ONUs. Considering the PON's differentiated CoS, the average packet delays of three types of priority classes are investigated in this study.

To further improve the efficiency of EPON's QoS performance, traffic predicting mechanisms have been proposed [57–59]. For example, in [58], data delays and losses are improved by effectively predicting the traffic arrival during the waiting period and also by limiting the maximum resources allocation for an ONU. In [57], an analytical model is used to estimate the bandwidth allocations to reduce the ONU buffer size and hence the queuing delays.

Another type of bandwidth allocation mechanism that allows users to share uplink bandwidth depending on the service level agreement (SLA) is proposed in [60]. The proposed mechanism assures the bandwidth requirements of the users who purchased bandwidth guaranteed services. This is achieved by categorising ONUs according to their SLA parameters and dividing the PON cycle into several sub-cycles.

A bandwidth allocation mechanism to efficiently handle the multi-service provisioning in EPON uplink traffic is proposed in [61]. In this algorithm, bandwidth requirements of multiple services are embedded into a REPORT message sent from each ONU. Hence, the OLT uses a class based bandwidth allocation mechanism to grant bandwidth for each CoS separately.

In this chapter, we propose a series of dynamic bandwidth allocation algorithms to improve video-on-demand delivery over PONs. In the next section, we discuss the most commonly used VoD system architectures.

2.3 Video-on-Demand Architectures

Data communications forecasts have highlighted that global Internet video related traffic from all services such as VoD, peer-to-peer (P2P), etc., will be as high as 80% of all consumer Internet traffic by 2019 [35]. The same report also estimates that in 2019, it would take an individual over 5 million years to watch the amount of video that will cross global IP networks each month. That is, every second, nearly a million minutes of video content will cross the network. Due to the combination of demand growth for VoD, and size of videos getting larger considering the new developments such as high definition (HD) audio/video and 3D features, the current network capabilities will clearly be insufficient in the future. As a solution, several network architectures and algorithms have been proposed to meet this impending growth in bandwidth.

In this section, we discuss architectural concepts that have been considered for VoD systems. A single or combination of these architectures can be used in practice. Moreover, while these concepts are not completely independent and may have some overlap amongst each other, we discuss the concepts separately for the sake of clarity.

A. Centralised Server based VoD Architecture

Figure 2.2 shows a centralised server based VoD system architecture. This is a simple architecture in which a central server is responsible for serving all its users. The videos are delivered from an origin server to the requesters using unicast. This architecture is

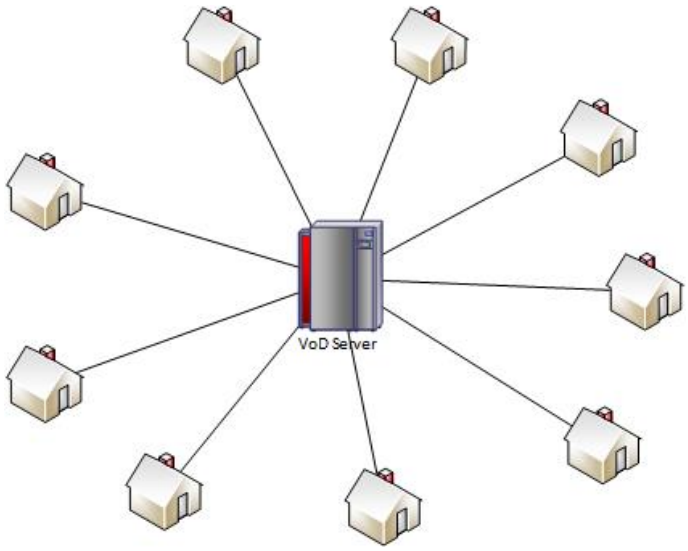


Figure 2.2: Centralised server based VoD system architecture.

straightforward and simple to implement. However, it is not scalable. VoD systems based on this architecture are unable to sustain against the expected future growth without significant equipment upgrades. Moreover, when this architecture is used for a VoD system, the load on the VoD server as well as on the network will be significantly high. Due to these shortcomings, the centralised server based architecture is not suitable for large scale VoD deployments.

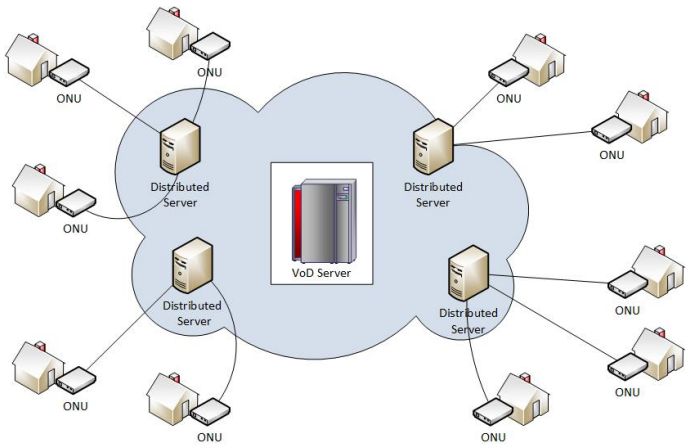


Figure 2.3: Distributed server based VoD system architecture.

B. Distribute Server based VoD Architecture

The distributed server based VoD system architecture is shown in Fig. 2.3. Currently, this architecture is widely used for large scale VoD deployments. In this architecture, distributed servers are installed in strategic locations across the network to optimise VoD delivery to the users. These storages can be either proxy servers installed at locations close to the users [62,63], or a hierarchical network of distributed content storages [64–68]. Moreover, these storages may either store a subset of popular videos in full [64, 66], or only portions of videos [69–71]. When a user requests a video, the request is served by the nearest server if it has the requested content, or else, the request is served from a different server, generally, located at a higher level in the distribution tree. Consequently, the use of distributed server architecture reduces the load on the origin server and improves the bandwidth efficiency of the VoD system.

C. Peer-to-peer (P2P) based VoD Architecture

The peer-to-peer paradigm has recently drawn a lot of attention as one of the strategies that can be used to reduce the load from distributed VoD servers and to increase the capacity of VoD systems. Many efforts have been made to exploit P2P for VoD services [72–74]. Fig. 2.4 shows a simplified high level architecture of a P2P based VoD system. In such a system, users participate in the video content delivery by re-distributing video content that they have previously downloaded.

D. Cloud based VoD Architecture

The cloud computing paradigm is increasingly becoming popular amongst VoD service providers as a promising alternative to actually deploying physical servers and building privately owned data centres. Cloud computing allows multiple applications to run on the same physical server so that those applications can share resources such as processing, storage and bandwidth. Commercial cloud platforms such as Amazon AWS [75] and Microsoft Azure [76], are established cloud platforms today. Moreover, VoD service providers such as Netflix, are leveraging on resources from the cloud service providers

[77]. As a result, cloud based VoD systems have drawn significant attention from both academia and industry [78–81]. Fig. 2.5 shows a simplified high-level architecture of a cloud based VoD system. In such a system, VoD service providers usually have their own data centres and servers in addition to the cloud based infrastructure. These self-owned infrastructure store video content and serve a fraction of the total video requests. Additionally, these self-owned servers upload video to the cloud storages. In turn, these videos are pushed towards the servers located at the edge of the network via the cloud. In studies reported in [80, 81], the authors show that multiplexing can be used to optimise the resource utilisation of VoD since multiple virtual machines can be run on a single server using virtualisation.

2.4 Optimising Video-on-Demand Delivery - Related Work

Currently, on-demand video already accounts for a bulk of Internet traffic [35]. Its traffic growth is expected to continue in the coming years, as the number of users, streaming rates and library sizes grow. As a result, VoD traffic volume is estimated to increase by a few orders of magnitude in the next few years [35]. Therefore, implementing bandwidth efficient VoD delivery techniques has become an important requirement for future VoD systems. Many studies have been published, aiming at optimising the delivery of VoD.

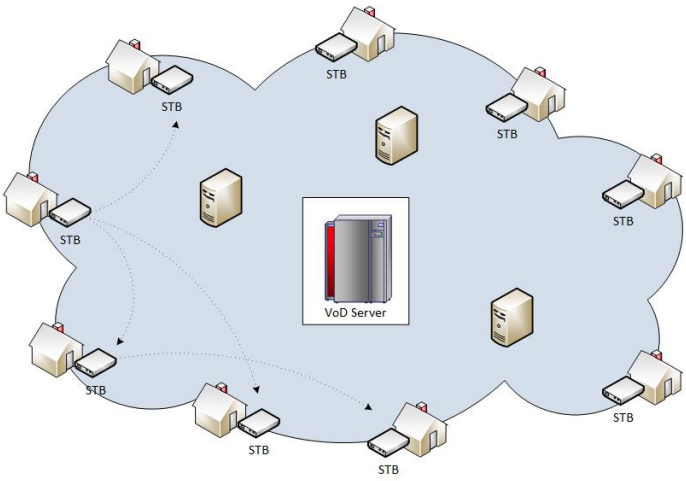


Figure 2.4: Peer-to-peer based VoD system architecture.

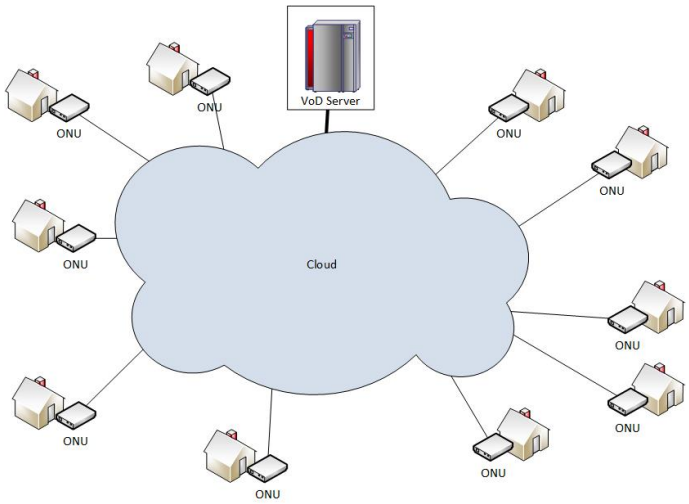


Figure 2.5: Cloud based VoD system architecture.

In this section, we review some of the most notable work related to our work.

A. Exploitation of Multicast

Typical VoD systems deliver unicast video streams from server to users. This approach is very expensive since a separate unicast stream is required to serve each request. Moreover, using unicast for VoD delivery is less scalable since bandwidth consumption increases linearly with the number of users watching videos simultaneously. The main bottlenecks here are the server streaming capacity[1], and the bandwidth capacity of the delivery network.

On the other hand, if multicast is used instead of unicast, then a group of users can share a single stream to receive video content from the server. Use of multicast has many benefits including but not limited that listed below:

- Use of multicast decreases the load on servers. This is due to the fact that a single stream transmitted from the server can serve multiple users simultaneously.
- Multicast reduces network bandwidth consumption significantly. This is because when multicast is used, only one stream is delivered to a location close to the user i.e. central office whereby the stream is replicated to multiple receiving users.

[1]Streaming capacity is the number of simultaneous streams that the VoD server can support

- Since the constrained system resources such as server capacity and network bandwidth are efficiently utilised when multicast is used, overall operational cost is reduced. This increases profitability for the VoD service provider.

- Multicast provides excellent scalability for VoD systems enabling them to sustain well against expected future growth. This also reduces the overall capital costs as the VoD systems can support increased numbers of users without requiring equipment upgrades.

While multicast is enabled by default in most modern IPTV and transport networks, some challenges arise when it comes to VoD delivery due to the underlying characteristics of VoD. This is because, multicast is applicable only in scenarios where multiple users watch the same part of the same video at the same time, which is highly unlikely in VoD. However, there are many studies attempting to enable multicast in VoD systems at the expense of a slight reduction of user interactivity and/or user quality of experience [82, 83].

However, these multicast approaches are considered out of scope for the work in this thesis as our aim is to optimise VoD delivery by strategic content placement close to end users. The related work on strategic content placement approaches for optimising VoD are discussed next.

B. Strategic Content Placement

Strategic placement of video content in content storages servers deployed at different locations within the network is one of the most promising and most popular approach that is used to improve the efficiency of VoD systems. The most intuitive and common approach is to replicate the popular items in multiple content storages located closer to users while keeping the less popular items in centralised data centres. Despite being simple and straightforward, such approach has the potential to not only reduce the server and network bandwidth consumption, but also to improve the QoS, QoE, and energy-efficiency of the VoD system.

Recent advances in storage technologies have made disk space plentiful and afford-

able. Moreover, these storage technologies together with advances in network technologies have facilitated the placement of content very close to users, for example, in optical line terminals (OLTs) in passive optical networks (PONs), base stations in mobile networks, or even in the end user's equipment itself. Thus, placement of content in a wide range of possible storage locations [63] from customer set-top-boxes (STBs) or residential gateways (RGs) [84–86], to external proxy servers located closer to users [64] to centralised pool of servers [87], have been studied in the literature. In general, the content placement is carried out by actively pushing the content, e.g. pre-population, or dynamically storing the content on the fly, e.g. caching. In this subsection, we review some of the most notable work aimed at optimising VoD delivery through strategic placement of video content.

Proxy-based Content Placement

Strategic placement of video content in proxy servers installed at locations close to users, is one of the most popular approaches widely used to improve the efficiency of VoD delivery. Typically, proxy servers are installed in the same local network within the same access network segment as the clients since communications within the local network is more reliable and cheap. Unlike web services for which the proxies are initially used, the video file sizes and the data rates required by VoD systems are tremendously high. Consequently, implementing proxies in VoD systems have many distinct challenges.

In [69], the authors proposed a proxy *prefix caching* technique. In this scheme, the proxy stores only the prefixes of popular videos. When a request for a video is received, the proxy server immediately starts sending out the prefix of that video to the user while requesting the rest of the video (i.e. the suffix of the video) from the VoD server. This suffix is relayed to the user through the proxy, as it streams from the server. Since proxy servers absorb the delay, throughput, and loss variations in the paths between the VoD server and the proxy servers, service providers can guarantee good QoS even if the paths between the proxy servers and the VoD server has a weak service level agreement (SLA). In addition, proxy servers also perform *work-ahead smoothing* into client's playback buffer for variable bit rate traffic. Work-ahead smoothing reduces the network resource require-

ments between the proxies and the clients by transmitting large frames in advance of each burst.

A more generalised class of schemes that exploits the idea of storing only portions of videos in the proxy servers, is *segment caching* [70, 71, 88, 89]. Unlike in prefix caching, the portions stored at the proxy servers under segment caching are not necessarily the initial portions. These segments of videos stored in caches are dynamically determined by a cache admission policy and a cache replacement policy. In [88], authors consider the caching of uniform size segments in layer encoded videos. On the other hand, an exponential segmentation method is used in [70]. In this method, the size of a segment is determined based on the distance from the beginning of the video, i.e. the closer it is to the beginning, the smaller the size of the segment. This strategy is based on the assumption that the probabilities of accessing later segments of videos are less than that of the initial segments. Such a strategy allows the cache manager to quickly discard a big chunk of cached items. An adaptive and lazy segmentation based caching strategy is proposed in [71]. The proposed segmentation strategy dynamically adapts to real time access patterns and subsequently segments the videos as late as possible. The lazy segmentation policy is further improved in [89], such that it effectively addresses the conflicting interest between reducing the start-up latencies and improving the byte hit ratio.

In [62, 90], a proxy based delivery technique called *video staging* is proposed for end-to-end VoD delivery over wide area networks (WANs). The objective of the proposed scheme is to reduce the bandwidth consumption in the backbone WAN. In the VoD architecture considered for this study, videos are streamed from a centralised VoD server to multiple requesting users through a WAN. Moreover, a proxy server is installed in each local area network (LAN). In the video staging approach, each video is partitioned along the rate axis in contrast to the segment and the prefix caching, where the partitioning is carried out along the time axis. More specifically, each video is divided into two parts: an *upper part* with more variability in the bandwidth requirement, and a *lower part* with less variability in the bandwidth requirement. The upper part is pre-fetched and delivered from the proxy servers whereas the lower part is retrieved from the central server in which the entire video is stored. Consequently, only a part of the video data

is delivered from the VoD server through the WAN whereas the rest of the data is retrieved from the proxy server located inside the LAN. While the video staging approach efficiently reduces bandwidth requirements, there are difficulties from the practical implementation point of view. For example, two portions of the video stored at the proxy and central server, need to be synchronised during the playback of a video. This adds extra processing and buffering requirements. Moreover, signalling messages from the user need to be delivered to both the proxy server and central VoD server, which also further increases the complexity of the system. Another framework that uses the concept of staging portions of videos in proxy servers, is proposed in [91]. This scheme reduces the bandwidth requirement in WAN by using proxy servers to facilitate a constant bit rate for the video streaming over WAN. The authors propose algorithms that select video frames for staging at the proxy servers, such that a given constant bit rate transmission is achievable.

A *frame-wise* proxy caching technique called *selective caching* is proposed in [92]. In this method, the proxy servers store only certain parts of videos. Unlike video staging approaches, the selective caching schemes are frame-wise in that a given frame is either not available in the cache or stored in its entirety. The authors propose two selective caching algorithms, selective caching for QoS networks (SCQ) and selective caching for best-effort networks (SCB). These algorithms are designed for caching a single video with a preallocated cache space. Moreover, these algorithms may select a group of non-consecutive frames to store at the proxies. The proposed SCQ algorithm can reduce the network cost of bandwidth reservation using a small client buffer. On the other hand, the SCB algorithm can increase the playback robustness without violating the client buffer budget.

One of the main challenges in caching only portions of videos in proxy servers is the in-time delivery of different portions. That is, the portions stored in the proxies and the VoD server need to be delivered to the users without delays in order to ensure a smooth continuous playback. Since proxy servers are located within the same local network as the users, timely delivery of cached portions is not a significant concern. However, there can be delays in fetching the segments that are not cached, which can result in discontin-

uous playback. Such delays are also known as *proxy jitters*. In [89], an active pre-fetching strategy that reduces the proxy jitters is proposed. To this end, authors present an optimisation model that minimises the proxy jitter subject to the requirements on start-up latencies and byte hit ratios.

Another class of much researched proxy based content placement schemes is *collaborative proxy caching*. In collaborative proxy caching schemes, a group of proxies typically in the same Intranet, collaborate with each other so that aggregated resources such as network bandwidth and storage space can be used more efficiently. In the *Middleman* collaborative proxy caching architecture proposed in [93], the proxy servers are arranged as a peer group, and a centralised *coordinator* coordinates the caching decisions and redirects the requests.

In [94], a set of light-weight cooperative cache management algorithms are proposed. These algorithms are aimed at maximising the traffic volumes served from the caches and minimising the bandwidth costs. The authors consider a cluster of distributed caches connected directly or through a parent node and use linear programming to determine the content placement solution that achieves the optimal performance. In [95], a novel collaborative caching mechanism based on real-world network topologies is proposed. The authors consider a heterogeneous system in which the request patterns and available storage capacities are different across different locations, and design a content placement strategy and a request routing rule that take the request patterns, cache sizes, link capacities and the topology of the system into account. They use an optimisation based approach for this purpose. The objective of their optimisation problem is to maximise the amount of requests that can be supported using the existing system infrastructures. They solve this problem by dividing it into sub-problems that focus on cache cooperation in different levels.

Other Content Placement Schemes

There are many studies in literature where VoD is optimised through content placement, but not necessarily using proxy caches. In this subsection, we discuss some of these works.

Video content placement in hierarchical VoD distribution networks is a widely studied area. In [96], a three-level hierarchical network storage architecture for VoD is presented. In their proposed scheme, a few videos with highest request probabilities are stored in the first level. An increased number of videos with relatively higher request probabilities are then stored at the second level, with all remaining videos stored at the third level. If a video is stored near the root of a hierarchical network, storage cost can be reduced since fewer copies will be stored. However, this will adversely affect the bandwidth cost since video data must travel further through the network to reach its destination. On the other hand, if video content is stored in a lower level of the hierarchy, then bandwidth cost is low but storage costs is high. Hence, there is a trade-off between bandwidth costs and storage costs when placing videos in a hierarchical network. This trade-off is studied in [64]. Moreover, models for video placements in hierarchical VoD distribution networks are developed in [97]. In [98], a video placement strategy that determines the optimal number of video copies that should be stored at each server of a hierarchical VoD system in studied. The objective of the proposed strategy is to minimise costs arising from storage and transport, while a lower bound for the probability of rejecting a request is guaranteed. The authors also study the implications of different system parameters such as video request probability and arrival patterns, on the performance of their placement strategy. In a subsequent publication [99], the same authors propose fast heuristic algorithms for video placement in hierarchical VoD networks, which can achieve near optimal operating costs.

Considering the advances in network technologies and rapidly decreasing storage costs, placement of video content much closer to customers has been highlighted as a viable approach for the future. There have been many studies that attempted to store videos in locations very close to the users, such as in residential gateways (RGs) [85, 100] or in set-top-boxes (STBs) [72]. In [85], the authors propose a VoD system architecture based on *nano data centres (NaDa)*. The objective of this NaDa based architecture is to reduce the energy consumption arising from VoD systems. In this method, RGs located at the customer premises are used to store some of the video content. More specifically, each RG is pre-loaded with a subset of video content, which will be subsequently re-

distributed to the other gateways. The content pre-loaded to the gateways can be full or partial replicas of videos, which are split into smaller segments to increase manageability. The authors formulate an optimisation problem that determines the number of replicas for each of the videos, which in turn yields the number of segments that will be stored in the system. Using a large set of empirical data, the authors show that the NaDa based architecture can save 20% - 30% of the energy, in comparison to conventional data centre based architecture. However, NaDa based architectures have its own limitations. Managing a very large number of nano data centres at each customer location requires complicated protocols considering the privacy of the content, which makes the implementation of such systems difficult. Further, due to the addition of a NaDa server at each customer location results in high initial capital cost.

An interesting solution for VoD delivery exploiting both multicast and strategic placement of videos closer to the customers is proposed in [3]. In [3], the authors propose a PON architecture where the PON architecture is modified to fit a local video storage server within the local access network. This local storage (LS) consists of an additional light source, and hence a wavelength channel which is used to deliver video data to the users simultaneously without consuming the upstream or downstream bandwidth of the standard PON. Moreover, the LS also consists of storage arrays that is used to store the video content. Considering the enabling high bandwidth efficiency and simplicity of management when compared to having a data centre at each customer location as discussed above, we believe introducing a LS into the access network to be a viable solution for future VoD delivery and is considered in our work in this chapter. In the next subsection, the local storage (LS) based PON architecture is discussed in detail.

Local Storage based PON Architecture for VoD

The typical operation of a passive optical network (PON) was discussed previously in Section 2.2. In this section, the modified PON architecture with the local storage (LS) is discussed. We consider this architecture for the VoD delivery approaches proposed in this thesis.

In [3], the authors proposed an enhanced architecture of a conventional PON with

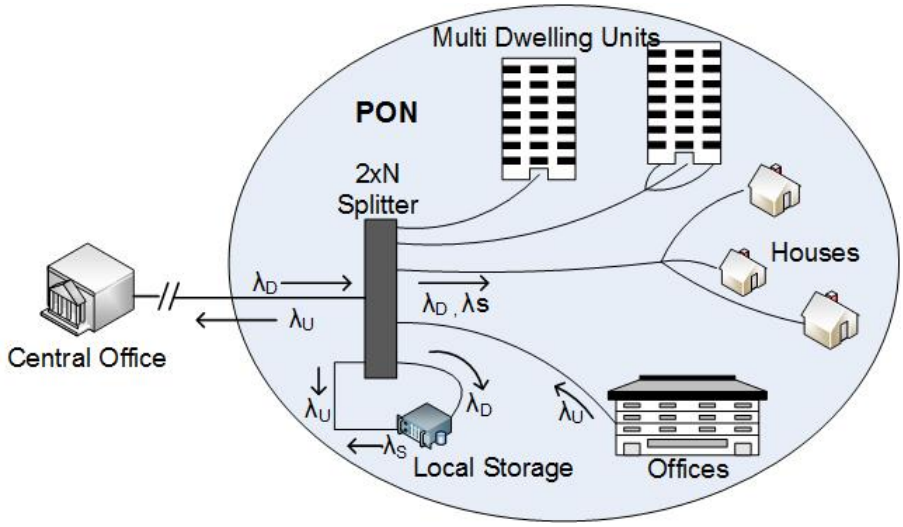


Figure 2.6: Enhanced PON architecture with LS [3].

efficient VoD delivery. As illustrated in Fig. 2.6, the authors introduce a local storage (LS) server with an additional light source which allows an additional wavelength channel transmission on a separate wavelength (λ_S). In Fig. 2.6, λ_D and λ_U denote the downstream and upstream wavelength channels, respectively. The objective of LS is to store a subset of videos within the PON and transmit the video content locally to the customers at the event of a receiving a video request. As shown in the figure, this LS is connected to a 2xN passive splitter using two fibres.

Popular videos are stored locally in the LS once they are transmitted from the central office (CO) on λ_D, where as subsequent requests for stored videos are served by a second data stream to the customer using a separate wavelength channel on λ_S from the LS. VoD requests sent by the optical network units (ONUs) on λ_U are received by the LS and if the requested video is available in its local library, the LS responds to the request. The authors also suggest that for best results, the LS can be housed closer to the nearest possible power source to the passive splitter.

Caching algorithm for the local storage (LS)

Various cache replacement algorithms have been proposed to determine which content should be stored and likewise evicted from the cache servers [101]. Due to the large size

of cached video content, the decision of which content to be retained or removed from the server is crucial for the performance. Least recently used (LRU)[2] and least frequently used (LFU)[3] algorithms are considered as the most straight forward and widely used caching algorithms. However, for the purpose of VoD delivery, both are reported to have algorithm specific drawbacks adversely effecting VoD delivery performance. As a result, a popularity-aware Last-*k* algorithm has been proposed [4]. The authors show that Last-*k* algorithm can reduce the number of active video streams by 40% during peek hours by caching less than 5% of the videos stored at the CO. In the context of this thesis, we consider the Last-*k* algorithm in order to determine the subset of most popular videos. The Last-*k* algorithm is discussed next.

Last-*k* Algorithm

The Last-*k* algorithm tracks the popularity of video contents using the most recent statistics and facilitates intelligent cache decisions that store the most popular content precisely at the time of update. Here, the relationship between the movie popularity and the inter-arrival times, which is the time between two consecutive video request arrivals for a specific video, is taken into account to estimate the popularity of a video. Generally, the inter-arrival time of a popular video should be lower than the inter-arrival time of a less popular video. In Fig. 2.7, backward distance in time $t_{i,j}$ is defined as the time since last k backward distances of the movie i so that,

$$T_{i,j} = \sum_{j=1}^{k} t_{i,j} \tag{2.1}$$

$T_{i,j}$ is used as the estimation parameter, which increases with decreasing popularity. $T_{i,j}$ is calculated for all videos in the library and the LS cache is updated with videos with least $T_{i,j}$ values. Here, k denotes the number of past arrivals considered when estimating the popularity.

[2]Least recently used (LRU) caching algorithm discards the least recently used items first. This algorithm requires keeping historical access statistics for each video, which is expensive if one wants to ensure the algorithm always discards the least recently used item.

[3]Least frequently used (LFU) caching algorithm calculates how often an item is needed. Items that are used least often are discarded first.

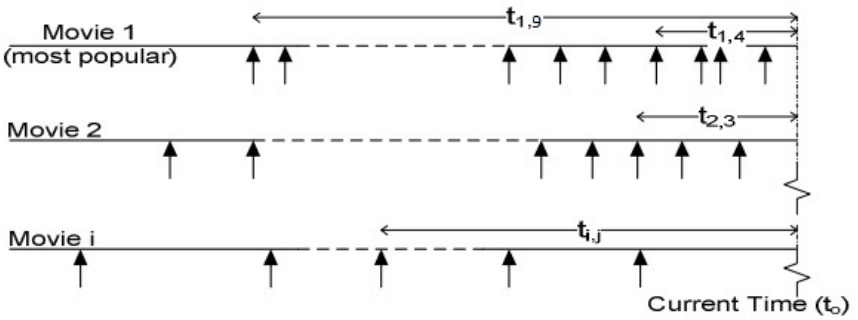


Figure 2.7: Video request arrivals at the VoD server [4].

The above approach is simulated using discrete time simulations using MATLAB in [3], the authors show that 40% of downstream bandwidth could be saved by storing only less than 5 % of videos at the LS. In our work, we exploit the ability of the LS to deliver high downstream bandwidth using an additional light source at the LS as discussed previously. We consider several receiver configurations at the ONU in order to maintain continuous reception of data/videos from both λ_D (from the CO) and λ_S (from the LS) simultaneously. Unlike in [3], we present several novel DBA algorithms to utilise the proposed receiver configurations to perform a packet level simulation to further analyse VoD delivery over PONs with high accuracy in the packet-specific level. In the next section, we will discuss the considered VoD system over the local storage based PON architecture, and the proposed dynamic bandwidth allocation algorithms and their receiver configurations.

2.5 VoD Delivery Over Local Storage based PONs

Video content is generally distributed using unicast video streams from central IPTV servers or intermediate cache servers upon customer request. Due to the downstream broadcast nature of the PON, copies of the content are transmitted to every customer connected to the network, thus resulting in very inefficient downstream bandwidth utilisation. Multiple customers requesting a popular video at different times will cause the same content to be distributed repeatedly, maximising downstream bandwidth wastage in an exponential manner. Each customer connects to the PON using an ONU which

is located at the customer premises. In the illustrative example used in our analysis in this thesis, we consider a PON to service 32 ONUs or equivalently 32 customers. This number can be increased to 64 or 128 depending on the power budget of the network. Fig. 2.8 shows the local storage based PON architecture that is considered in this work. The LS is placed at the remote node location of the optical distribution network. At the remote node, the LS is strategically connected to a 2 x 32 splitter enabling the LS to receive a copy of all content distributed from the OLT which is located at the CO. The LS also receives a copy of all video requests and content uploaded by the ONUs towards the OLT. As such, the LS has the advantage of knowing the exact downstream and upstream statuses of the PON at all times. Most importantly, the LS can broadcast content within the PON in a localised manner. It is equipped with a transmitter allowing it to broadcast video on a separate wavelength, denoted λ_S. Correspondingly, the ONUs are equipped with an additional photodetector to receive content on λ_S. Downstream communication is a broadcast amongst all connected ONUs on λ_D. Time division multiplexing is used to efficiently and equally share downstream bandwidth. Upstream communication is governed by the IEEE 802.3ah standard Multi-Point Control Protocol that uses *REPORT* and *GATE* control frames for bandwidth request and allocation [102]. During operation, the ONU sends a request to the OLT on λ_U requesting for a specific video, Video *h*. The LS receives a copy of this request on λ_U . If Video *h* is available in its library, the LS will deliver Video h on λ_S. During operation, the OLT will keep track of the videos that are stored in the LS and will ignore a request on the knowledge that the LS will service that particular request. If the requested video, e.g. Video *j*, is not available in the LS, the LS will ignore the request and will allow the OLT to deliver Video j on λ_D. In this case, if Video j is recognised as a new popular movie by the Last-*k* caching algorithm at the LS, the LS will store the contents of Video j it receives on λ_D in its storage array. Subsequently, when another ONU requests for the same video, the LS can service this request locally without utilising additional bandwidth on λ_D. As a result of this continuous process, bandwidth usage for video delivery of popular moves is limited to λ_S, thus allowing bandwidth on λ_D to be made available for other downstream services.

In the next section, we propose a novel DBA algorithm termed as the *dual-receiver*

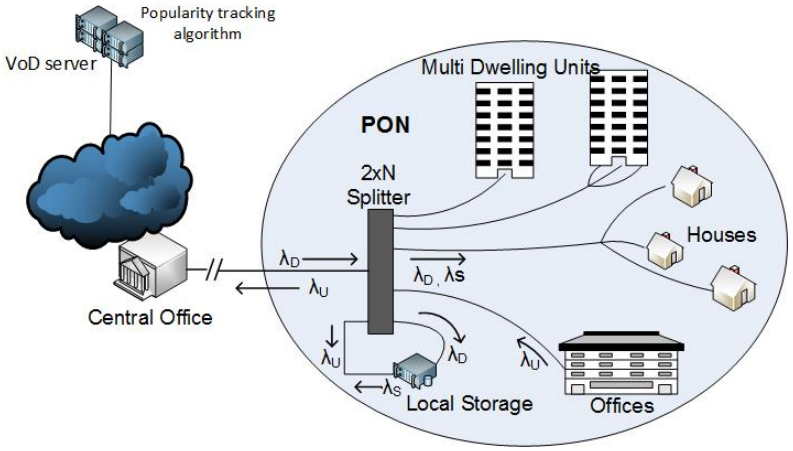


Figure 2.8: Local Storage based PON architecture for VoD delivery [4].

dynamic bandwidth allocation (DR-DBA) algorithm. This algorithm optimises the LS based VoD delivery scheme in the PON.

2.5.1 Dual-Receiver Dynamic Bandwidth Allocation (DR-DBA) Algorithm

In this section, we present the dual-receiver dynamic bandwidth allocation (DR-DBA) algorithm that optimises packet delivery in a PON with LS. The time division multiple access (TDMA) scheme is used in downstream direction on λ_D to effectively share the bandwidth amongst the ONUs. Upstream communication is managed by the centralised OLT granting time windows for each ONU to upload its content based on the reported queue sizes. The proposed architecture in Fig. 2.8 uses REPORT and GATE control frames to manage upstream communication within the PON as suggested by IEEE 802.3ah standard [102]. REPORT control frames are sent upstream on λ_U by each ONU to communicate its bandwidth requirement to the OLT. The OLT collects all REPORT frames and allocates bandwidth as required for the next cycle according to the DBA. These allocations are then communicated downstream to the ONUs using GATE control frames on λ_D. According to the IEEE 802.3ah standard, the DBA is left open for vendor implementation. The formats of control frames are adjustable to suit the services delivered by the PON. Here, we modify the frame format of both GATE and REPORT multi-point control protocol (MPCP) control frames. In the REPORT control frame message, one octet is allo-

Field	No. of Octets
Destination Address	6
Source Address	6
Length/Type = $88\text{-}08_{16}$	2
Opcode = $00\text{-}03_{16}$	2
Timestamp	4
Number of queue sets	1
Report bitmap	1
Queue #1 - #8 reports	16
Requested Video ID	**1**
Pad = 0	0- 21
Frame check sequence	4

Figure 2.9: Format of the REPORT control frame with a modified field for the Requested Video ID.

cated to carry the requested video identity number from the ONU to the OLT. The format of the REPORT control frame is adjusted to accommodate the VoD service parameter, the Requested Video ID, as highlighted in Fig. 2.9. The opcode specific fields, which can carry a multiple number of report bitmaps and queue sizes, depending on the number of queue sets, are generally allocated 40 octets by the standard. One octet is taken out from this set to be used as the Requested Video ID carrier to the OLT. This adjustment will eliminate one cycle from the initial request communication between the ONU and OLT, which in return will directly help improve the delay performance of the network.

Similarly, in the GATE control frame, one octet is allocated to carry VoD content information from the OLT to the LS, as highlighted in Fig. 2.10. This information is used to communicate the popularity information derived by the Last-k algorithm at the OLT, to the LS. If the transmitted video is recognised by the caching algorithm to be a popular video, LS will store a copy of the video in its storage arrays during the first downstream delivery on λ_D. Maintenance of the most popular set of videos in the LS is thus optimised since it does not require additional bandwidth on λ_D other than what will be already utilised for the first downstream delivery of that video. This improvement is only possible due to the above mentioned format adjustment.

Field	No. of Octets
Destination Address	6
Source Address	6
Length/Type = $88\text{-}08_{16}$	2
Opcode = $00\text{-}02_{16}$	2
Timestamp	4
Number of Grants	1
Grant start time + Grant length	6
VoD content information	**1**
Pad = 0	0 - 32
Frame check sequence	4

Figure 2.10: Format of the GATE control frame with a modified field for the VoD content information.

The algorithm by which the VoD service is delivered is illustrated further using the timing diagrams shown in Fig. 2.11. Figure 2.11 (a) illustrates how a video request is handled in the PON when the requested video is not available in the LS video library. Figure 2.11 (b) shows how the same request is handled when the LS library contains the requested video. When a video is requested, the ONU will add the Requested Video ID to the REPORT control frame to be sent next. This frame is then uploaded to the OLT in the next transmission window. Therefore, the initial waiting time for a request at the ONU is limited to at most one cycle. This enhancement is only possible because of the proposed format for the REPORT frame which consists of a field dedicated to accommodate the video ID. Without the format adjustment highlighted earlier in Fig. 2.9, sending the Requested Video ID to the OLT will extend over two cycles, i.e. one cycle for requesting for a timeslot to upload the data portion carrying the Requested Video ID and another cycle for the actual transmission of the Video ID. Once the OLT receives the REPORT frame with the Requested Video ID, it will search for the requested video in the LS content index which the OLT manages for reference. If the requested video is unavailable in the LS library, the OLT will begin to broadcast the video packets in the next cycle as shown in Fig. 2.11 (a). In the case where the LS library contains the requested video, the OLT ignores the request allowing the LS to service the request on λ_S. The

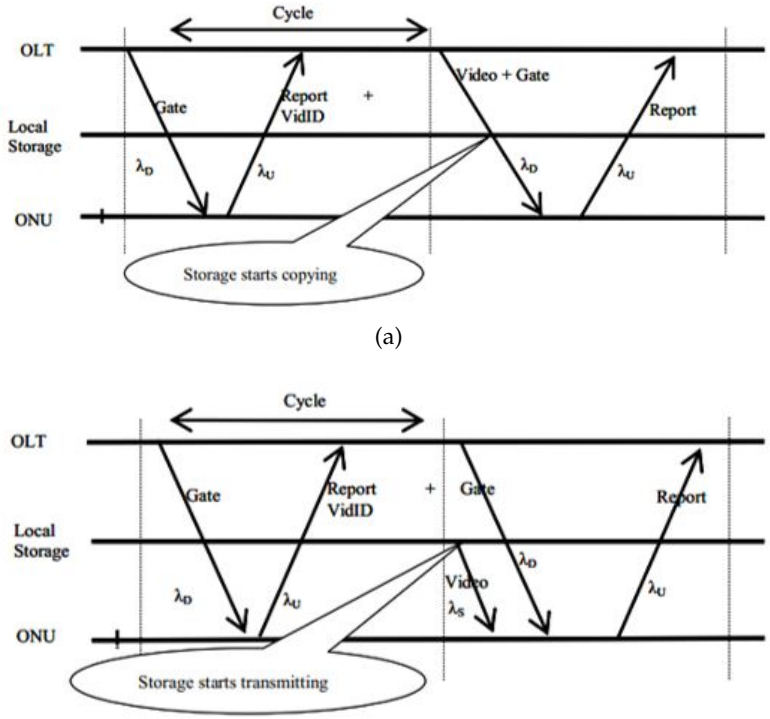


Figure 2.11: Timing diagrams to illustrate communication between the OLT, LS, and ONU when the (a) central office services the request and (b) local storage services the request.

OLT continues its transmissions from the next cycle as shown in Fig. 2.11 (b). The LS transmission differs from that of the OLT mainly because the time division multiplexed bandwidth on λ_S is shared amongst only those ONUs which are receiving video from the LS. In comparison, the OLT will allocate downstream timeslots for each and every ONU irrespective of their active or idle status. The generic steps of the above mentioned DBA algorithm is represented in the flow chart shown in Fig. 2.13.

Algorithm 1 illustrates the basic operational steps of the LS based VoD delivery using the DR-DBA algorithm. The DR-DBA algorithm exploits the capability to utilise two separate wavelength streams (λ_D and λ_S) simultaneously. However, this requires a dual receiver configuration at each ONU. Fig. 2.12 illustrates the dual receiver configuration considered for the DR-DBA algorithm. As shown in Fig. 2.12, each ONU is equipped with two photodetectors that are connected to two fixed-tuned filters that are then coupled by a 3dB coupler. The two filters are tuned to λ_D and λ_S wavelength channels to

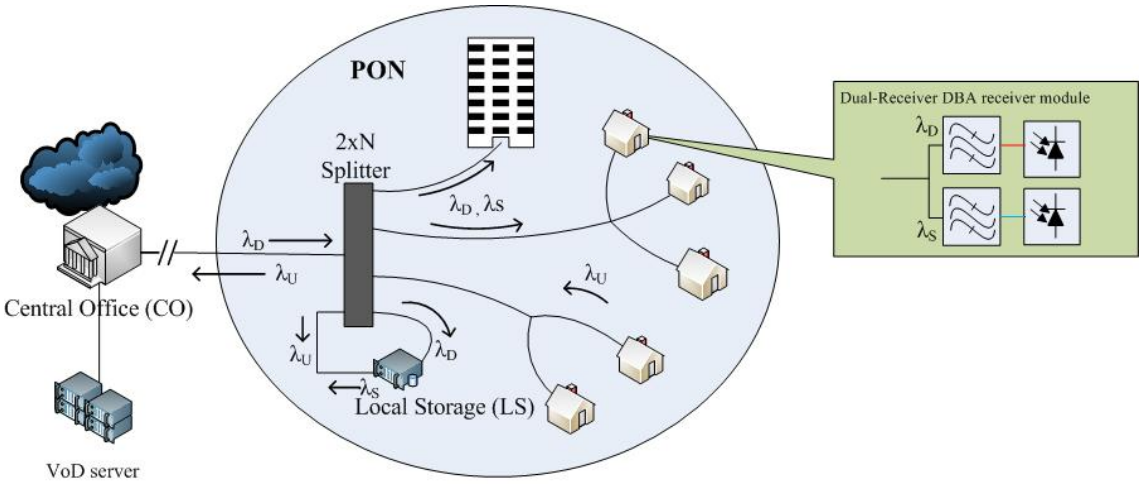


Figure 2.12: DR-DBA receiver configuration with dual receivers.

maintain simultaneous data and video reception.

The main drawback of DR-DBA algorithm is the increased power consumption at the ONU due to the use of two receivers to facilitate the continuous and simultaneous receiving of data on λ_S and λ_D. Therefore, in an attempt to counter this increased power consumption, in the next section, we further extended DR-DBA algorithm to operate with a single, but dynamically tunable receiver configuration at the ONU end. The single-receiver dynamic bandwidth allocation (SR-DBA) algorithm is discussed in the next section.

2.5.2 Single-Receiver Dynamic Bandwidth Allocation (SR-DBA) Algorithm

The SR-DBA algorithm exploits the use of a tunable single receiver module at the ONU to receive data on both λ_D and λ_S wavelength channels. This receiver configuration is shown in Fig. 2.14. The receiver module includes a wide-band receiver along with a tunable wavelength filter that tunes between λ_D and λ_S as required. We consider the use of a high speed liquid crystal filter for the wavelength filtering due to its high-speed switching capabilities. Currently, liquid crystal wavelength filters record a switching speed of approximately 10 μs and consumes low power on the mW scale [103]. Such a tunable receiver module at the ONU eliminates the need for a second receiver at the ONU, thus reducing network power consumption.

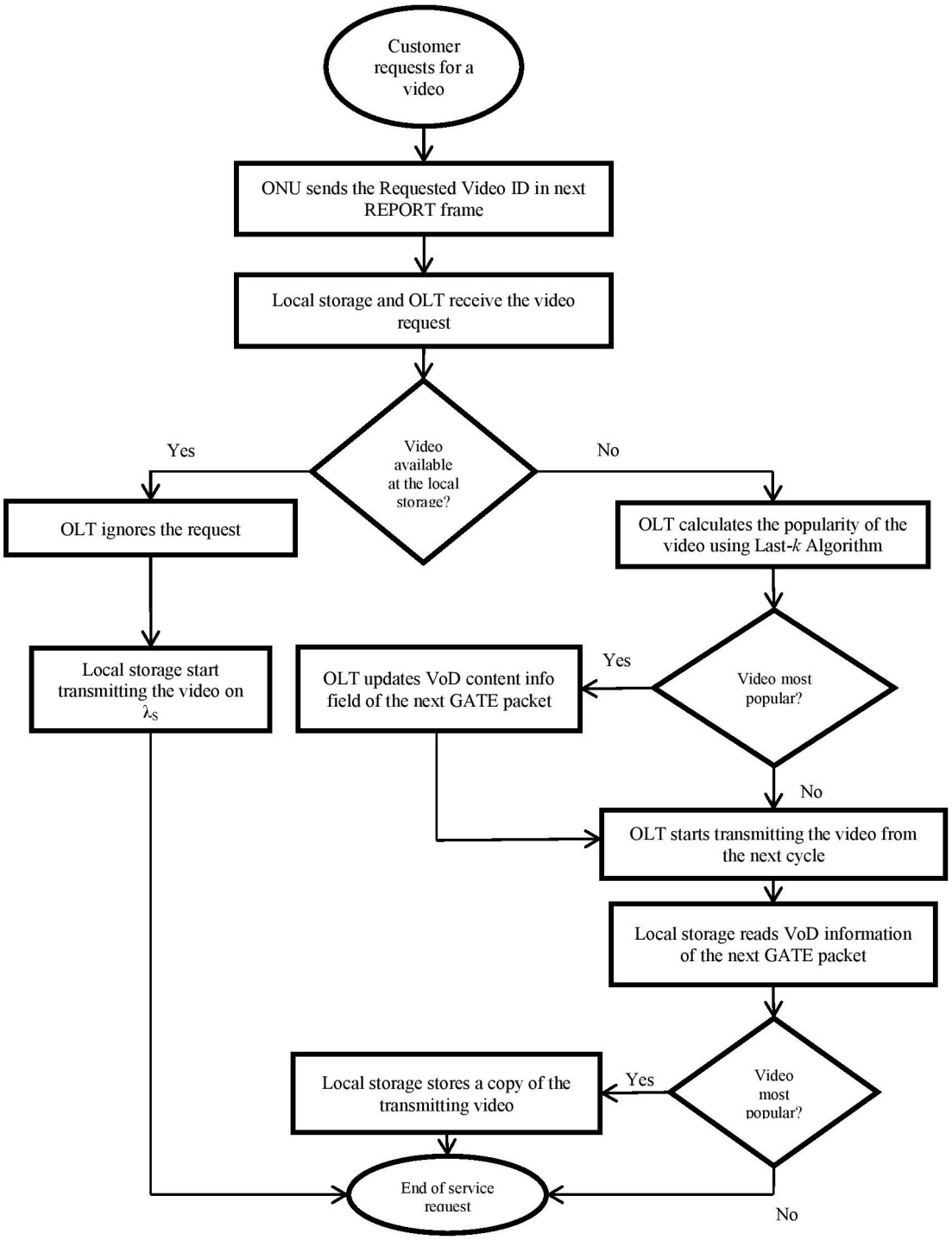


Figure 2.13: Flow chart of the DR-DBA algorithm.

Algorithm 1 Pseudo code for DR-DBA algorithm

1: ONU_i gets request for video V in cycle j
2: ONU_i updates *Requested_Vid_ID* in REPORT$_i$ frame
3: ONU_i sends REPORT$_i$ on λ_U in cycle $j+1$
4: CO and LS receive the REPORT$_i$
5: **if** V is available in LS **then**
6: **repeat**
7: LS transmits V on λ_S from cycle $j+2$
8: **until** ONU_i is VoD active
9: ONU returns to normal operation
10: **else**
11: CO calculates popularity of V
12: **if** V is popular **then**
13: CO updates next GATE$_i$ frame with popularity rating
14: CO sends adjusted GATE$_i$ frame in cycle $j+1$
15: CO transmits V from cycle $j+2$
16: LS retains a copy of V
17: **else**
18: CO transmits V from cycle $j+1$ on λ_D

Algorithm 2 illustrates the pseudo code explaining the functions of SR-DBA algorithm. The MPCP is used to govern communication between the CO and ONUs as specified by the IEEE 802.3ah standard [102]. Here, the formats of REPORT and GATE control frames are adjusted to communicate VoD related instructions as previously discussed in subsection 2.5.1. When a customer requests a video, the requested video ID is sent to the CO using the immediate REPORT control frame (lines 1-3 in Algorithm 2). Then, if the video is available in the LS, the CO instructs the ONU to re-tune its receiver to λ_S using flag bits allocated for VoD control in the GATE frame of the following cycle (line 6 of Algorithm 2). The ONU acknowledges the received instructions and tunes its receiver to λ_S from the third cycle (line 9 of Algorithm 2). The ONU continues receiving content on λ_S until the VoD stream is finished, as presented by lines 10 - 12 of the algorithm. The ONU is considered to be in *VoD active* state while it is receiving VoD content on λ_S as mentioned in the algorithm.

Conversely, if the requested video is not present in the LS library, the CO will distribute the VoD content following similar steps to CO based VoD delivery, as discussed previously in subsection 2.5.1. These steps include calculating the video popularity rat-

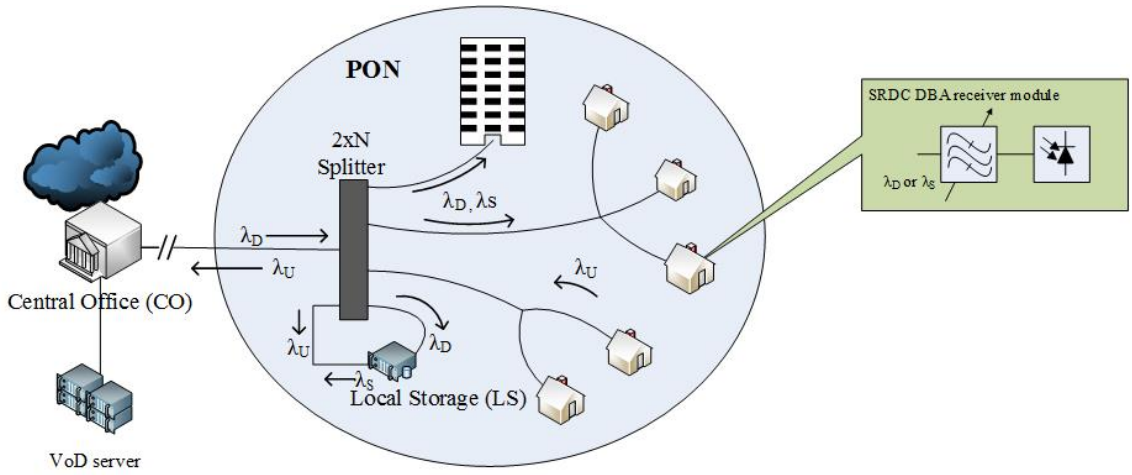


Figure 2.14: Single receiver configuration for SR-DBA algorithm.

ing using the Last-k caching algorithm [4] and if the video is a popular video, the CO instructs the LS to create a copy in its library for future use.

The main drawback of the SR-DBA algorithm lies in the fact that while an ONU is actively receiving a VoD stream from the LS, the communication between the CO and the receiving ONU is disrupted. That is, the CO extracts the $Ack\_activate\_\lambda_S$ flag bit information from the received REPORT control frame (line 8 in Algorithm 2) and considers the ONU to be inactive on λ_D. When the ONU is finished with its VoD usage on λ_S, it will follow auto discovery MPCP protocols to reactivate its presence on λ_D thus receiving a GATE frame. During the assigned window by this GATE frame, the ONU communicates finishing of VoD download with the CO by setting the *Finish_VoD* flag of the next REPORT frame. The additional time taken for EPON auto discovery protocols is another limitation of the algorithm.

2.5.3 Single-Receiver Dual-Channel Dynamic Bandwidth Allocation (SRDC-DBA) Algorithm

The SRDC-DBA algorithm is aimed at addressing the above-mentioned limitations of SR-DBA while maintaining the previously explained single receiver based architecture shown in Fig. 2.14. Here, we consider tuning the receiver rapidly between both wavelengths within each cycle time thus making it possible to stay active on both wavelengths.

Algorithm 2 Pseudo code for SR-DBA algorithm

1: ONU_i gets request for video V in cycle j
2: ONU_i updates *Requested_Vid_ID* in REPORT$_i$ frame
3: ONU_i sends REPORT$_i$ on λ_U in cycle $j+1$
4: CO and LS receive the REPORT$_i$
5: **if** V is available in LS **then**
6: CO updates *Activate_*λ_S flag of next GATE$_i$ frame
7: ONU_i receives GATE$_i$ in cycle $j+1$
8: ONU_i updates *Ack_activate_*λ_S flag in REPORT$_i$ frame (cycle $j+1$)
9: ONU_i tunes receiver to λ_S from cycle $j+2$
10: **repeat** LS transmits V on λ_S from cycle $j+2$
11: **until** ONU_i is VoD active
12: ONU_i follows auto discovery protocols
13: ONU_i updates *Finish_VoD* flag of next REPORT$_i$ frame
14: CO returns to normal operation
15: **else**
16: CO calculates popularity of V
17: **if** V is popular **then**
18: CO updates next GATE$_i$ frame with popularity rating
19: CO sends adjusted GATE$_i$ frame in cycle $j+1$
20: CO transmits V from cycle $j+2$
21: LS retains a copy of V
22: **else**
23: CO transmits V from cycle $j+1$ on λ_D

Therefore, the SRDC-DBA algorithm effectively allows the ONU to receive data and VoD contents on the two wavelengths, λ_D and λ_S, respectively. The algorithm calculates the time duration that each ONU is tuned to λ_S in each cycle in order to receive the transmitted VoD data segment on λ_S. More importantly, the SRDC-DBA algorithm ensures the two data blocks addressed to a specific ONU on λ_D and λ_S do not overlap with each other.

Algorithm 3 illustrates the pseudo code explaining the functions of the SRDC-DBA algorithm. Similarly to the SR-DBA algorithm, SRDC-DBA uses the standardised control frames of GATE and REPORT to allocate and request bandwidth in the network, respectively. When a user requests a video, the ONU will adjust its next REPORT frame to carry the *Requested_Video_ID* to the CO. Once the REPORT frame is received at the CO and LS, the CO checks the requested video with the available videos in the LS. In the instance where the requested video is available in the LS library, the duration of time (T_{idle}) that

Algorithm 3 Pseudo code for SRDC-DBA algorithm

1: ONU_i requests for video V in cycle j
2: ONU_i updates *Requested_Vid_ID* in REPORT$_i$ frame
3: ONU_i sends REPORT$_i$ on λ_U in cycle $j+1$
4: CO and LS receive the REPORT$_i$
5: **if** V is available in LS **then**
6: CO calculates T_{idle_i}
7: **repeat**
8: CO updates GATE$_i$ frame with T_{idle_i}
9: CO transmits DATA$_i$
10: CO transmits GATE$_i$ frame
11: ONU_i tunes to λ_S for T_{idle_i} duration
12: LS transmits V during T_{idle_i}
13: **until** ONU_i is VoD active
14: ONU_i updates *Finish* flag of next REPORT$_i$
15: CO returns to normal operation
16: **else**
17: CO calculates popularity of V
18: **if** V is popular **then**
19: CO updates next GATE$_i$ with popularity rating
20: CO sends adjusted GATE$_i$ in cycle $j+1$
21: CO transmits V from cycle $j+2$
22: LS retains a copy of V
23: **else**
24: CO transmits V from cycle $j+1$ on λ_D

the receiver spends idle on λ_D within a cycle is calculated at the CO server (line 6 of Algorithm 3). Let T_{win} be the time duration that a specific ONU receives data addressed to itself. Then, the idle time can be calculated by subtracting the downstream time window (T_{win}) and tuning delay (T_{tuning}) of the tunable wavelength filter by the full cycle time as presented in Eq. 2.2 below.

$$T_{idle} = T_{cycle} - T_{win} - 2 * T_{tuning} \tag{2.2}$$

Once T_{idle} is obtained, the CO updates the next GATE control frame with the idle time and transmits it to the ONU in the next cycle. The ONU, thus receiving the T_{idle} value will re-tune its receivers to λ_D immediately after it finishes receiving the downstream VoD content on λ_S for a duration of T_{idle}. This will enable the ONU to receive VoD data on λ_S during the idle time of the receiver (lines 8 - 12 of Algorithm 3).

When compared with the DR-DBA and SR-DBA algorithms, the main difference of SRDC-DBA is the use of a single receiver to receive data and video on both λ_D and λ_S simultaneously. However, similar to SR-DBA, SRDC-DBA will also utilise an extra cycle when initialising the VoD downstream to communicate the channel details from CO to the ONU.

2.5.4 Model Evaluation

This section presents the methods used to evaluate the performance of the proposed VoD delivery algorithms. We measure QoS attributes such as delay, jitter and throughput, the network power requirements, and the deployment cost of the network considering the proposed algorithm variants and their respective receiver configurations. The QoS attributes were measured using packet level simulations while the network power requirement and deployment costs were estimated with formulated mathematical models. We compare the above mentioned attributes of LS based VoD delivery against a baseline architecture where VoD is delivered without a local storage server to evaluate the performance of the proposed algorithm variants.

A. Simulation

A packet level simulator was developed using C# and executed on a computer which runs .NET Framework 3.5 as the run-time environment. The network and protocol parameters used in the simulations are listed in Table 2.2. Simulations were carried out to measure the QoS performance of the VoD delivery. Packet delay, jitter, and network throughput parameters were measured for the four main cases of our study, namely the SR-DBA case, SRDC-DBA case, DR-DBA case, and the case where the CO delivers VoD without the presence of a LS in the PON. Further, to improve the granularity of the results, the number of active ONUs using the VoD service at a given time is varied from 1 to 32. The results thus represent how packet delay varies under different load conditions. The simulation time for each run is chosen to be 5 seconds which covers 2500 cycles (cycle time = 2ms). This duration is sufficient for delay values to reach final average values. We assume that the VoD requests to the VoD server follow a Poisson distribution model in

Parameter	Value
No. of ONUs in PON	32
Length from ONU to Splitter	1 km
Length from CO to Splitter	20 km
PON Bandwidth	10 Gbps
Cycle time	2 ms
Simulation time	5 s
Inter-frame Gap Time	12 bytes
Maximum Ethernet packet size	1518 bytes
Minimum Ethernet packet size	64 bytes

Table 2.2: Network and Protocol Parameters.

order to simulate the request bursts of a typical requesting service. Further, we consider data and video packet sizes to be uniformly distributed between the standard Ethernet packet size of 64 to 1518 bytes.

B. Power Consumption Model

In [104], authors have presented the energy savings achievable at the ONU by the use of 10 Gbps vertical cavity surface emitting lasers (VCSELs). They propose sleep and doze mode operations where transmitter (TX) and receiver (RX) blocks are powered down to save the power consumption at the ONU. It is shown that a 15% power saving is achieved when ONU is equipped with a VCSEL transmitter compared to a traditional DFB based ONU. Further, the authors have formulated a power consumption model for a distributed storage solution for VoD delivery whereby videos are stored at the CO. Equation 2.3 describes the power consumption per customer for downstream VoD delivery over a PON.

$$\begin{aligned} P_{customer}(\mathrm{W}) &= \frac{P_{storage}}{N} + \frac{P_{server}}{N} + \frac{KP_{OLT}}{N} + P_{ONU} \\ P_{customer}(\mathrm{W}) &= \left(\frac{V \times M \times 17pW/bit}{N}\right) + \left(\frac{B \times N_{active} \times 70W/Gbps}{N}\right) + \\ &\quad \frac{KP_{OLT}}{N} + P_{ONU} \end{aligned} \tag{2.3}$$

In Eq. 2.3, parameter V = 28.8 Gbit is the HD video size assuming an average length of 60 minutes per video. Parameter M is the number of videos stored where parameter N represents the number of ONUs supported. Parameter K is the number of OLT line cards required to support N ONUs and B = 0.008 Gbps is the bit rate for HD video stream. The terms on the right hand side of the equation represents the power requirement of the CO storage arrays, CO video server, OLT chassis and ONU, respectively.

The LS equipment placed within the PON comprises a video server and the storage arrays that contain video data. Due to different equipment requirements and receiver configurations of considered cases as mentioned in previously, the power requirement of the PON may vary depending on the size of the network. An overview of the considered receiver configurations for the three proposed DBAs is shown in Fig. 2.15. Here, we formulate a generic mathematical model to estimate power consumption per customer for different network sizes. Table 2.3 lists the equipment specifications used for formulating power consumption models. The power consumption per customer of the proposed PON architecture (see Fig. 2.15) is given by:

$$
\begin{aligned}
P_{customer}(\mathrm{W}) &= \frac{P_{Cstorage}}{N} + \frac{P_{Cserver}}{N} + \frac{KP_{Lstorage}}{N} + \frac{KP_{Lserver}}{N} + \frac{KP_{OLT}}{N} + P_{ONU} \\
P_{customer}(\mathrm{W}) &= \left(\frac{V \times M_{central} \times 17pW/bit}{N}\right) + \left(\frac{B \times N_{central} \times 70W/Gbps}{N}\right) + \\
&\quad K\left(\frac{V \times M_{local} \times 17pW/bit}{N}\right) + K\left(\frac{B \times N_{local} \times 70W/Gbps}{N}\right) + \\
&\quad \frac{KP_{OLT}}{N} + P_{ONU} \qquad (2.4)
\end{aligned}
$$

Equipment	**Power/ Capacity**
Video server (Nexus 5010, 2 x UCS 6100 Fabric Interconnect, UCS 5108 Blade server chassis) (P_{server})	70 W/Gbps
Storage arrays (EMC vMax SE) ($P_{storage}$)	17 pW/bit
OLT line card (Alloptic edge 10) (P_{OLT})	80 W
VCSEL-ONU (P_{ONU})	3.984 W

Table 2.3: Equipment Specifications.

In Eq. 2.4, parameter V is the average size of a HD video. We consider this value to

be 28.8 Gbit assuming an average duration of 60 minutes. The parameters $M_{central}$ and M_{local} are the numbers of videos stored in storage arrays at CO and LS. The parameter K is the number of 10 Gbps optical line terminal (OLT) line cards used at the CO (maximum 32 line cards). The parameter N is the number of ONUs supported (maximum N =1024 = 32 OLT line cards $\times$ 32 ONUs). The parameters $N_{central}$ and N_{local} are number of active ONUs using video downstream from CO and from LS respectively. The parameter B = 0.008 Gbps is the bit-rate for the HD video stream. The terms on the right hand side of the equation account for different components of power consumption per customer arising from CO storage arrays, CO video server, LS storage arrays, LS video server, OLT chassis and VCSEL-ONU [104], respectively.

The parameter P_{ONU} includes the power consumed by the transmitter and the receiver at the ONU. The power consumed for the tunable liquid crystal filter is considered marginal and hence ignored in the power calculations as the filter operates in lower mW scale ($<$100 mW) [103]. Further, we consider the LS to be dynamically updated with the most popular video content. As discussed previously, the Last-k caching algorithm [4] is used to determine the most popular set of videos based on the past video request arrival statistics recorded at the CO . In our analysis, we consider M_{local} as 5% of the $M_{central}$ value to maintain simplicity. That is, we consider 1000 videos are to be stored at CO ($M_{central}$ = 1000) and 50 most popular videos(5% out of 1000 videos) are continuously maintained at LS. Comparative results amongst different cases are less dependent on the chosen base value for $M_{central}$ since it is common across the three considered cases.

C. Cost Model

In this subsection, we develop a network cost model to estimate the cost of LS-based passive optical network deployment from a network providers standpoint. The model is used to estimate the total network cost in two deployment scenarios, namely *duct reuse* scenario, and *greenfield* scenario. The duct reuse scenario assumes reusing of already available ducts and only considers costs that include blowing fibre into such already installed ducts. Conversely, the greenfield scenario assumes the network to be deployed from scratch, thus includes fibre costs related to trenching, laying, and blowing of fibre

into new ducts. The total deployment cost per customer of the LS based PON is calculated by:

$$C_{customer} = \frac{C_{CO}}{N} + \frac{KC_{LS}}{N} + \frac{C_{fibre}}{N} + \frac{KC_{splitter}}{N} + C'_{ONU} \tag{2.5}$$

In Eq.2.5, parameters C_{CO} , C_{LS}, and C'_{ONU} represent the cost contributions arising from the network equipment at the central office, local storage, and optical network unit at the customer location, respectively. The parameter C_{fibre} represents the fibre deployment costs depending on the considered deployment scenario (i.e. duct reuse or greenfield) in each case. The parameter $C_{splitter}$ accounts for the cost of the power splitter which connects the ONUs and the LS to the central office. Further, the parameter N represents the number of ONUs supported, and the parameter K is the number of OLT line cards required to support N ONUs. The parameters in the right hand side of the Eq.2.5 are further elaborated as shown next.

The costs arising from the equipment at the central office (C_{CO}) is calculated by:

$$C_{CO} = K \times C_{OLT} + C_{server} + C_{storage} \tag{2.6}$$

where, parameters C_{OLT}, C_{server}, and $C_{storage}$ account for the cost contribution from the OLT line cards, the VoD server machine, and the VoD storage arrays located at the central office.

The costs arising from the equipment at the local storage (C_{LS}) is calculated by:

$$C_{LS} = C_{LSserver} + C_{LSstorage} + C_{TX} + 2 \times C_{RX} \tag{2.7}$$

where, parameters $C_{LSserver}$ and $C_{LSstorage}$ account for the cost contributions from the VoD server machine, and the storage arrays, which are located at the LS, respectively. Further, the parameters C_{TX} and C_{RX} represent the costs arising from the additional light source (transmitting on λ_S), and the receiver module, respectively. Here, two photo-detectors have been considered to facilitate receiving on both upstream and downstream transmissions of the PON simultaneously.

The costs arising from the equipment at the ONU (C'_{ONU}) depends on the considered

receiver configuration due to additional equipment (RX blocks, couplers, filters, etc.) being used (as shown in Fig. 2.15). Eq. 2.8 presents the adjusted cost calculation for the considered two receiver configurations.

$$C'_{ONU} = \begin{cases} C_{ONU} + C_{coupler} + 2 \times C_{RX} + 2 \times C_{sta\_filter}, & \text{for dual-receiver configuration} \\ C_{ONU} + C_{RX} + C_{tun\_filter}, & \text{for single-receiver configuration} \\ C_{ONU} & \text{no LS case} \end{cases} \tag{2.8}$$

In Eq. 2.8, the parameter C_{ONU} account for the cost of the ONU equipment located at the customer premises. The parameters C_{RX}, represents the cost arising from the additional photo-detectors at the ONU. Futher, parameters $C_{coupler}$, C_{sta_filter}, and C_{tun_filter} account for the cost contribution from the optical couplers, static filters, and tunable filters, respectively.

We use the above formulated network cost models to estimate the deployment costs of a LS based PON for the network provider. Here, we consider a network of 32 PONs each supporting 32 ONUs each (1024 ONUs). In our analysis, we consider all equipment costs with respect to the cost of a GPON ONU, which is considered as a cost unit (CU). The considered equipment and infrastructure deployment costs are listed in Table 2.4.

The proposed LS based VoD delivery algorithms are evaluated using the discussed packet level simulations, power consumption models and cost models. Results are presented in the next section.

2.5.5 Results and Discussion

This section presents the results from the packet-level simulations, power consumption models, and cost models to study the performance of LS based VoD delivery. Fig. 2.16 plots the average VoD packet delays against different network load conditions. The observed sudden increment of delay in the SRDC-DBA curve where the number of active ONUs is one, could be explained by the dynamic nature of the DBA. When a single ONU is active, the DBA assigns the bandwidth on λ_S entirely to the active ONU. However, the

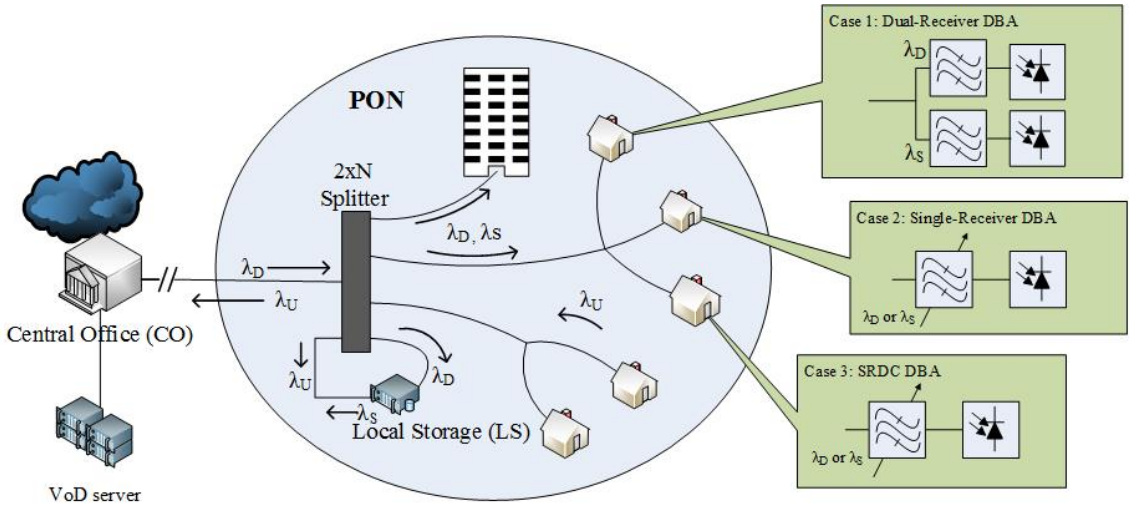


Figure 2.15: Receiver configurations used for DR-DBA, SR-DBA, and SRDC-DBA.

Component	**Cost [CU]**
OLT (C_{OLT})	463
ONU (C_{ONU})	3.1
VoD server (C_{server}, $C_{LSserver}$)	288
Storage arrays - 12TB (at CO) ($C_{storage}$) - 1TB (at LS) ($C_{LSstorage}$)	 55.6 6.5
Power splitter ($C_{splitter}$) - 1:32 - 2:32	 6.6 6.8
Filter - static (C_{sta_filter}) - tunable (C_{tun_filter})	 1.5 3
Coupler ($C_{coupler}$)	1
TX block (C_{TX})	1.5
RX block (C_{RX})	1.5
Fibre (CU/km) (C_{fibre}) - duct reuse - greenfield	 300 904

Table 2.4: Costs of considered components.

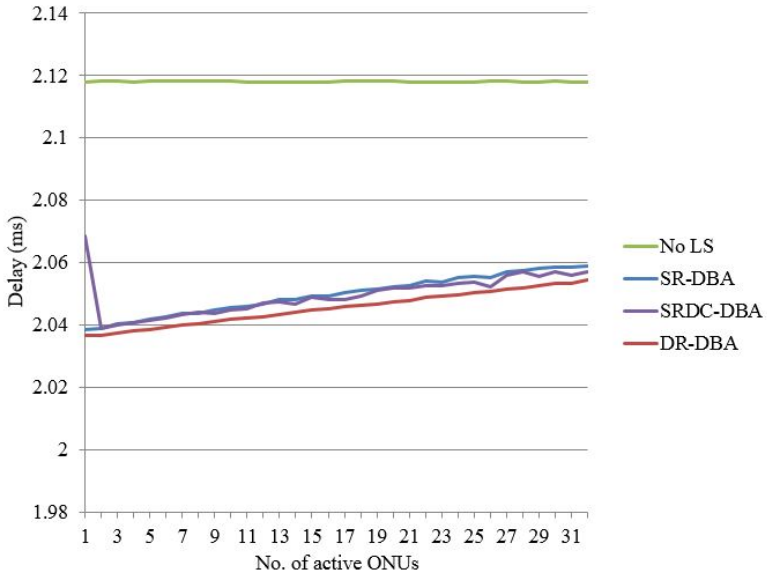


Figure 2.16: Delay as a function of number of active ONUs.

full assigned time window of the entire cycle cannot be used because the receiver should be re-tuning to λ_D in each cycle to receive its downstream data from the CO. This will introduce additional queuing delay for the VoD data packets. Also, all three algorithms SR-DBA, SRDC-DBA and DR-DBA show delay increments with increase of number of active ONUS. This is due to the packet delay introduced when the LS time window size dynamically varies with the change of active ONUs. Delay values in SR-DBA and SRDC-DBA curves indicate marginally higher delays compared to the DR-DBA case irrespective of the network load. This is due to the utilisation of an extra time cycle at the VoD delivery initialisation of the both SR-DBA and SRDC-DBA algorithms. More importantly, results in Fig. 2.16 indicate that all SR-DBA, SRDC-DBA and DR-DBA cases show lower delays than the architecture without LS due to the use of a LS to cache VoD data closer to the end user. The observed improvement in delay performance is marginal (4% to 5%) with respect to the no LS case since the VoD server is considered to be placed at the CO which is only 20 kms from the ONUs. Comparatively, better delay performances could be expected where the VoD server is placed farther away from the end user in architectures such as long-reach PONs [105].

Fig. 2.17 shows average packet jitter curves against different network loads. Jitter

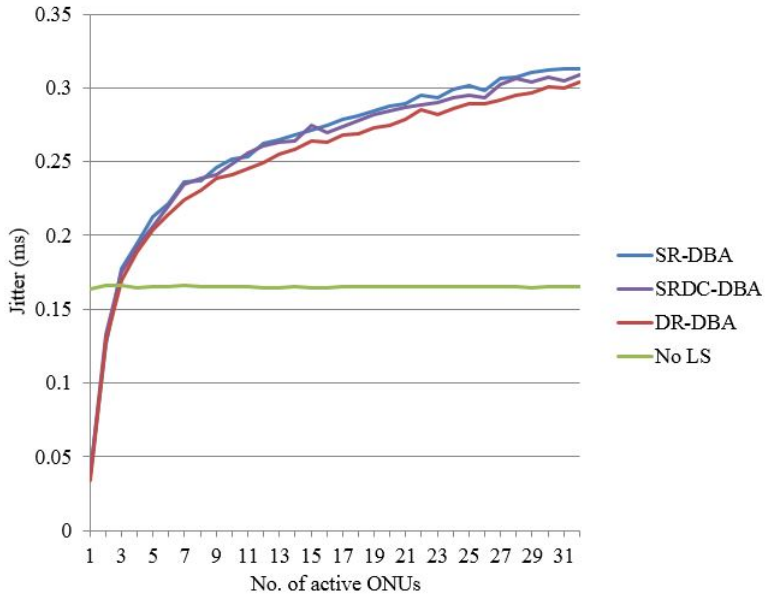


Figure 2.17: Jitter as a function of number of active ONUs.

is considered as the variation of packet delays. For cases SR-DBA, SRDC-DBA and DR-DBA, jitter levels increase with the number of active ONUs due to the additional packet delay variations caused by the change of LS window cycle every time an ONU joins or leaves. Jitter levels are indicated to be lower than that of no LS case when the number of active ONUs are very low. This is due to the low number of ONU additions, which in turn will cause less packet delay volatility on λ_S. Jitter values in the architecture without LS are constant around 0.16 ms irrespective of the network load. This is due to the time division multiple access (TDMA) nature of the λ_D transmission. Curves indicate that when the number of active ONUs increases beyond 3 ONUs, average jitter levels of SRDC-DBA and DR-DBA increase above the jitter value of the without LS case. However, we can use jitter buffers at the ONUs to offset such delay variations, even in the worst case where the jitter is still $<$0.3 ms.

Fig. 2.18 plots the maximum achievable aggregated throughput levels. Aggregated throughput is considered as the total combined throughput received on both λ_D and λ_S. As illustrated in Fig. 2.18, the achievable throughput levels in architectures with the LS are higher than that of the architecture with no LS. This is due to the use of the additional wavelength λ_D to transport the high priority VoD traffic to the ONUs. Further, since λ_S is

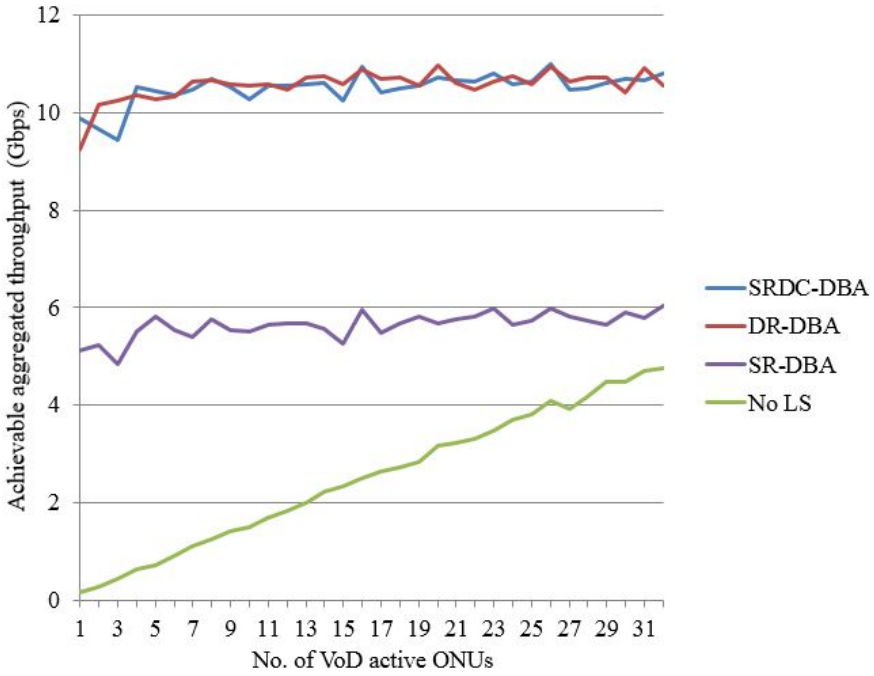


Figure 2.18: Maximum achievable aggregated throughput on λ_D and λ_S.

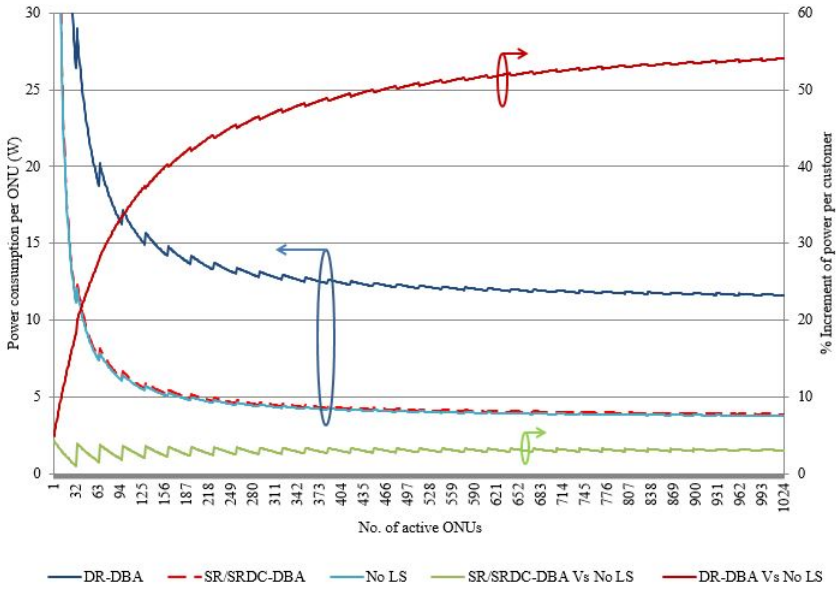


Figure 2.19: Power consumption per customer and percentage power increment as a function of number of active ONUs.

shared only amongst the active VoD ONUs in the PON, high throughput could be maintained throughout different network loads. Due to the TDMA nature of downstream transmission on λ_D, achievable bandwidth will depend on the number of active ONUs. This behaviour is exhibited by the aggregated throughput curve of the architecture without LS in Fig. 2.18. The throughput values for the SR-DBA algorithm are maintained around half of the throughput levels achieved by SRDC-DBA and DR-DBA algorithms. This is due to the mentioned limitation of the SR-DBA algorithm where data could only be received on a single wavelength at a given time (either on λ_S or λ_D). Our results clearly suggest that both SRDC-DBA and DR-DBA both show similar maximum achievable throughput levels. Results also show that the penalty in utilising two receivers in the DR-DBA algorithm compared to a single receiver in SRDC-DBA algorithm is negligible. This is due to the relatively fast switching speed of the liquid crystal wavelength filter ($\approx$ 10 μs) with respect to the cycle time (2 ms) of the simulation.

The power consumption per customer values estimated from the mathematical models discussed in Section 2.5.4 are plotted in Fig. 2.19. Here, we consider a network that connects 1024 ONUs (CO housing 32 OLT line cards that connect 32 ONUs per line card) to estimate the power consumption per customer in a large scale VoD system. Results show that the power consumption per customer for all three cases reduces as the number of ONUs increases. This is due to the equipment power requirement being shared amongst more ONUs as the network scales. Both architectures in SR-DBA and SRDC-DBA are similar in terms of the equipment considered. According to our power consumption model, the power estimation curves for SR-DBA and SRDC-DBA is identical due to the use of identical equipment. Therefore, both SR-DBA and SRDC-DBA cases are represented by the curve (in red) which is labelled as SR/SRDC-DBA in Fig. 2.19. The architecture without LS shows the lowest power consumption values. The DR-DBA architecture shows the highest power consumption values as two receivers are used to receive on λ_D and λ_S. This increment of power consumption is attributed to the power consumed by the additional receiver at each ONU. Therefore, the power consumption for SR/SRDC-DBA case lies in between the power consumption of DR-DBA and the no LS cases. Further, Fig. 2.19 also plots the percentage of power increment for the two

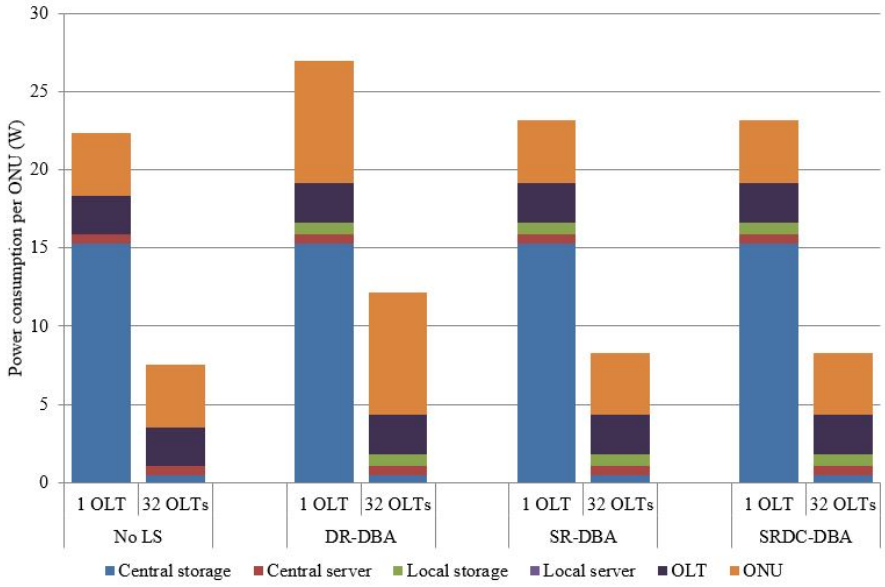


Figure 2.20: Power consumption per ONU. Equipment-based contribution break-down for 1 OLT line card servicing 32 ONUs and 32 OLT line cards servicing 1024 ONUs.

proposed algorithms as compared to the architecture without LS. The SR/SRDC-DBA algorithms show a 4% increment over the no LS case where the DR-DBA records a 54% increment in power consumption per ONU.

Fig. 2.20 illustrates the power consumption estimations for each of the four architectures as a composition of equipment contributions. The power consumption per ONU from each portion of the architecture at two network sizes, namely 1 OLT line card servicing 32 ONUs and 32 OLT line cards servicing 1024 ONUs, are plotted. The results on Fig. 2.20 show that the power consumption per ONU decreases as the network size is increased. This is due to the power consumed by common equipment i.e, central storage, central server etc., being shared amongst more ONUs. Amongst the four architectures, the DR-DBA architecture consumes the most power due to the use of two receivers at the ONU. Further, results clearly show that as the network scales, the power consumption per ONU is mainly contributed by the power requirement of the ONU itself. This value is significantly contributed by the power drawn by the receiver. Therefore, when a network size of 1024 ONUs with 32 OLT line cards is considered, results suggest that single-receiver based algorithms present minimum power consumption per ONU.

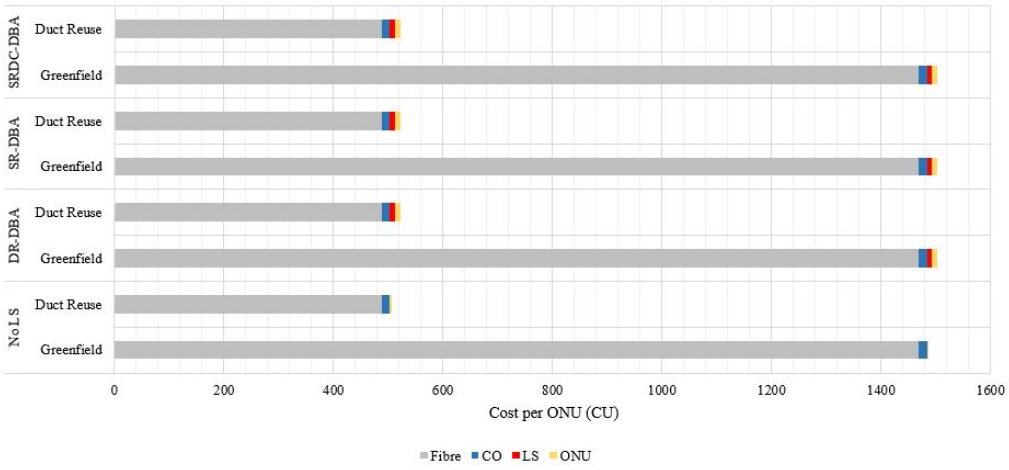


Figure 2.21: Cost per ONU. Equipment-based cost breakdown for 32 OLT line cards servicing 1024 ONUs.

Fig. 2.21 presents a breakdown of network deployment cost based on the considered equipment and network infrastructure in each algorithm. We considered a deployment of a LS based PON network to support VoD delivery for 1024 ONUs (32 OLT line cards at the CO, supporting 32 ONUs each), under two deployment scenarios (greenfield and duct reuse). The greenfield scenario considered deployment of the network in a area where current network infrastructure is not available. The duct reuse scenario considered blowing fibre into already installed ducts, thus reducing the fibre-related costs compared to the greenfield scenario. Results clearly highlight that in both greenfield and duct reuse deployment scenarios, the network cost is mainly dominated by the fibre deployment costs accounting for at least >93% of the total deployment cost in all considered cases. Further, comparing the duct reuse deployment against greenfield deployment, duct reuse scenario records significantly lower deployment cost due to reduced fibre-related costs, that is 300 CU/km compared to 903 CU/km in the greenfield scenario. Comparing the algorithm variants against the baseline architecture, the DR-DBA algorithm shows the highest deployment cost per ONU due to the considered additional equipment at the ONU. The SR-DBA and SRDC-DBA algorithms present slightly lower deployment costs compared to the DR-DBA algorithm thanks to the use of a single tunable receiver. The lowest deployment cost is recorded for the baseline architecture where the local storage is not considered in the PON. However, irrespective of the algorithm variant or the deploy-

Algorithm	Greenfield	Duct Reuse
SR-DBA	1.07%	3.13%
SRDC-DBA	1.07%	3.13%
DR-DBA	1.1%	3.23%

Table 2.5: Percentage of incremental deployment cost considering duct reuse and greenfield deployment of LS based PON.

ment scenario considered, the contribution from the local storage equipment towards the network deployment cost is relatively insignificant. This is mainly due to high cost of the fibre-based infrastructure deployment. Another reason is the LS-related cost being shared amongst multiple connected ONUs.

Table 2.5 presents the percentage of incremental deployment cost of LS based VoD delivery with respect to the baseline architecture considering the two deployment scenarios. The incremental costs are attributed to introduced LS equipment and receiver configurations considered for each proposed algorithm variant. The results clearly show that the incremental cost of LS based VoD delivery is marginal, that is approximately 1% and 3% increments in deployment cost in greenfield and duct reuse deployment scenarios, respectively. The SR-DBA and SRDC-DBA algorithms record slightly lower deployment cost increments compared to DR-DBA algorithm. This is due to the use of a single tunable receiver at the ONU. This incremental deployment cost is the cost trade-off of enabling improved VoD delivery over PONs using the LS.

2.6 Summary

In this chapter, we proposed three dynamic bandwidth allocation algorithms to optimise VoD delivery over passive optical networks by exploiting a strategically placed local storage server maintaining the set of most popular video content closer to the customer. The three algorithms, DR-DBA, SR-DBA, and SRDC-DBA are compared against the CO-based VoD delivery. The simulation results show that the delay and achievable throughput levels are improved when a LS is utilised. Nonetheless, jitter levels are increased with the use of the LS. However, the increment could be easily alleviated by using jitter buffers at the ONUs. Comparing the single receiver based algorithms (SR-

DBA and SRDC-DBA) against the dual receiver based DR-DBA algorithm, results show that both algorithms produce almost similar delay improvements and jitter variations. However, when throughput is considered, both DR-DBA and SRDC-DBA algorithms outperforms SR-DBA through the use of dual channel downstream transmission. With the introduction of active elements (local storage server and storage arrays) to the access network, an increase in network power consumption and deployment cost is inevitable. The results from the power consumption models show that the single receiver based algorithms, namely, SR-DBA and SRDC-DBA, only result in a power increment of 4% per ONU where as DR-DBA shows a high power increment of 54% per ONU compared to the no LS case. Therefore, comparatively, the SRDC-DBA algorithm delivers similar QoS and throughput improvements as the dual receiver based DR-DBA algorithm, but consumes a substantially less amount of power. Moreover, when the deployment cost is considered, the results from the cost model shows that the deployment cost of the local storage based PON is mainly contributed by the fibre-related costs and the cost spent for LS equipments is comparatively insignificant ($<$2% of total deployment cost per ONU).

In the next chapter, we take our discussion towards the wireless mobile access networks. We propose two approaches, a heuristic-based algorithm and a linear program to solve the transmit power control problem in an effort to improve the energy-efficiency of the wireless mobile access network.

Chapter 3

Green Heterogeneous Cellular Networks

This chapter focuses on the wireless mobile access network. We propose an energy-efficient transmit power control mechanism which dynamically adjusts the transmit power of base stations of a heterogeneous cellular network. Power saving is achieved by exploiting cell zooming techniques for the small cell base stations based on their hourly demand variations. The chapter proposes a heuristic-based algorithm and a linear program to solve the transmit power control problem and discusses power saving opportunities using simulation results.

3.1 Introduction

In this chapter, we focus on the energy-efficient power control solutions for wireless mobile access networks. The wireless mobile access networks use radio waves in the ultra high frequency (UHF) spectrum of the electromagnetic spectrum to connect mobile users to the cellular base stations (BSs). Reports on mobile data traffic highlight that mobile data traffic has grown 4,000-fold over the past 10 years and almost 400-million-fold in 15 years [106]. The same report also highlights that global mobile data traffic will increase nearly eight-fold between 2015 and 2020. For these reasons, network providers are beginning to deploy large scale mobile access network infrastructure to meet the impending growth in mobile traffic. However, due to large numbers of BSs being deployed, the increasing operational network power consumption has become a challenge for network operators. The power consumption of BSs is identified as the main contributor to the network power consumption, contributing to about 60% of the network power usage [5]. Therefore, energy-efficiency becomes an important parameter in the design of

next-generation wireless mobile access networks.

In general, the energy-efficiency of wireless mobile access networks is a well-studied and researched subject. Out of the many approaches proposed, a relatively novel concept named cell zooming (also known as cell size breathing) is considered in detail in this chapter. Cell zooming involves dynamically adjusting the cell coverage radius by controlling the transmission power of the small cell BSs. Our aim in this chapter is to propose a novel approach that exploits dynamic transmit power control based on the temporal variations of the hourly mobile traffic demand during a day to design an energy-efficient wireless mobile access network. In particular, we present two methods to solve the optimal transmit power control problem considering a heterogeneous wireless access network architecture.

The chapter is organised as follows. In Section 3.2, we discuss the wireless access network and the most commonly recognised state-of-the-art wireless access network technologies. We then discuss the heterogeneous networks architecture followed by a discussion on the key challenges of heterogeneous network deployments such as the coexistence of small cell BSs with existing macro cell networks in a heterogeneous network, self-organising capability and backhauling of the small cells. Then, in the Section 3.3, we discuss the energy-efficiency of heterogeneous networks with a classification of commonly used approaches to improve energy-efficiency of heterogeneous access networks. The concept of cell zooming is also discussed in detail here. Next, in the Section 3.4, a comprehensive review of the most notable efforts towards energy-efficiency of heterogeneous networks is presented. In Section 3.5, we present a heuristic-based transmit power control algorithm to solve the optimal power control problem of the heterogeneous mobile access network. Then, in Section 3.6 we take on a different approach to solve the same problem by formulating a linear program (LP).

The LP approach produces a global optimal solution whereas the heuristic-based algorithm produces a sub-optimal solution. However, the heuristic-based approach is computationally straightforward compared to the LP. As a result, the heuristic approach is able to model larger network segments than the LP. A network operator could initially model a large wireless mobile access network scenario to gain an insight of the

network operational power consumption with the heuristic-based sub-optimal solution. Subsequently, the LP could be used to model smaller and more specific segments of the network for a more detailed and optimal transmit power solution. The heuristic-based algorithm and the optimisation program proposed in this chapter present opportunities to save energy in heterogeneous wireless access networks. The chapter concludes with a summary of the results gathered from both approaches in Section 3.7.

3.2 Wireless Access Networks

Standardising bodies such as the Institute of Electrical and Electronics Engineers (IEEE) and the Third Generation Partnership Project (3GPP) have set forth various wireless access technologies towards effective operation of wireless access networks. A typical wireless access network architecture is shown in Fig. 3.1. In the wireless access network segment, wireless end users are connected to BSs which are in turn connected to the wireless core network through a wireless backhaul network. The wireless core network consists of gateways that provide connections to other networks, and the management entities that are responsible for managing the wireless users connected to the network.

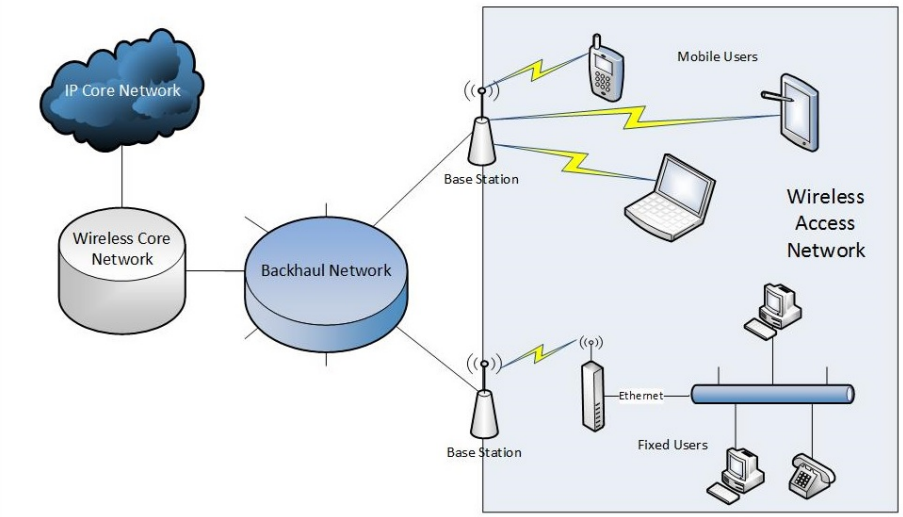


Figure 3.1: Wireless access network architecture.

3.2.1 Wireless Access Network Technologies

Different wireless technologies use different sets of operations and QoS mechanisms to efficiently serve its end users. Next, we elaborate on these aspects focusing on three state-of-the-art wireless broadband access network technologies namely long-term evolution (LTE), mobile WiMAX and WiFi. In particular, our discussion will be focusing on the LTE network, the most preferred wireless broadband access network technology today [107]. However, irrespective of the technology in use, carriers are still striving to find a cost-effective solution to serve the ever-increasing bandwidth demand in their wireless access networks.

A. Long-Term Evolution (LTE)

LTE is an all-IP network [108]. The core network segment of the LTE is referred to as evolved packet core (EPC) which consists of packet gateways and a mobility management entity (MME). The access network of LTE consists of BSs (which are termed as evolved-nodeBs (eNBs)), and user equipment. The eNBs are capable of allocating radio resources to its connected users without the help of the EPC [108]. An interface referred to as a *S1* is used to manage communications between the eNBs and EPC. Further, LTE facilitates communication amongst neighbouring eNBs by introducing an additional interface referred to as *X2*. The *X2* interface is a fairly new concept in relation to the wireless access technologies. Therefore, the implementation of a backhaul network that can efficiently facilitate both the *S1* and *X2* interfaces has been considered a crucial requirement for the deployment of LTE and also for its future releases.

Orthogonal-frequency division multiplexing (OFDM) is used as the multiple access technology in the downlink of the LTE network. However, in the uplink, the single-carrier frequency division multiplexing (SC-FDMA) scheme is used [109]. The LTE uses a concept called *evolved packet system bearer* to facilitate recommended QoS parameters for diverse next-generation broadband services [110]. These LTE bearers can be mainly classified into two types: guaranteed bit rate (GBR) and non-GBR. In GBR bearers, a specific set of parameters are maintained for the entire duration of the bearer's lifetime [111]. Each of these bearers is accompanied by a unique QoS class identifier (QCI) and

Type	QCI	Priority	Delay budget	Example services
GBR	1	2	100ms	Conversational voice
	2	4	150ms	Conversational video (live streaming)
	3	3	50ms	Real time gaming
	4	5	300ms	Non-conversational video
non-GBR	5	1	100ms	IMS signaling
	6	6	300ms	Buffered streaming
	7	7	100ms	Voice, Live streaming video, Interactive gaming
	8	8	300ms	Video (buffered streaming), TCP-based applications (e.g., www, e-mail, chat)
	9	9		

Table 3.1: LTE quality of service (QoS) classes.

an allocation and retention priority (ARP). While the ARP is used for the call admission control purposes, nine different QCIs are used to facilitate required QoS for different services. Each of these QCIs is characterised by its priority, packet delay budget and acceptable packet loss rate. These parameters of QCIs and their example services are listed in Table 3.1.

B. Mobile WiMAX

WiMAX is based on the IEEE 802.16 family of standards. Similar to the LTE, WiMAX is also an emerging wireless broadband access technology for high speed mobility access. Mobile WiMAX was first standardised as IEEE 802.16e [112] and the latest release is standardised as 802.16m. Mobile WiMAX uses orthogonal frequency division multiple access (OFDMA) in both the uplink and downlink [113]. Moreover, this technology supports two architectural models, namely point-to-multi-point (PMP) mode and mesh mode [114]. In PMP mode, the BS acts as a centralised controller to handle the wireless channel allocations for its active subscriber stations (SSs). In the mesh mode, traffic can be routed via SSs and bandwidth allocation decisions are also distributed amongst SSs.

In WiMAX, the medium access control packet data unit (MACPDU) is used to encapsulate data that traverses within the network. The QoS classes supported by WiMAX are connection-oriented where five different QoS classes are defined to support diverse broadband services [115]. These classes of services are universal grant service (UGS), extended real time pooling service (ertPS), real time pooling service (rtPS), non-real time pooling service (nrtPS), and best effort (BE). Depending on the latency, jitter, and bandwidth requirements of a service, one of these QoS classes can be selected. Similar to the

LTE, the resource scheduling mechanisms in the WiMAX play a major role in guaranteeing required QoS for the end users and hence, many studies have investigated methods to improve the resource handling in the WiMAX [116].

C. WiFi

WiFi is a wireless local area network (WLAN) technology which is based on the IEEE 802.11 family of standards [117]. Even though WiFi was introduced to facilitate wireless access in private environments such as in a house or in an office, now it is used in public places to provide broadband access [118]. The integration of WiFi and previously discussed wireless access network technologies are proposed in many studies, e.g., [119, 120]. Moreover, despite the fact that the reach of a WiFi network is small in comparison to the LTE and WiMAX networks, it is identified as an efficient solution to provide higher bandwidth to the end users [121]. The latest version of WiFi uses the OFDM-based transmission scheme. The WiFi network supports QoS differentiations for diverse services by classifying the traffic into eight different priorities. Depending on the network requirements, WiFi has the ability to facilitate different implementation architectures such as independent basic service set and infrastructure basic service set [122]. These architectures define how the WiFi stations communicate with each other in the network.

3.2.2 Heterogeneous Networks and Their Challenges

The evolution of mobile communication technologies is currently progressing towards the fifth-generation [123–125]. Each wireless access technology uses different multiplexing, modulation and coding techniques to improve the efficiency of radio channels [126–128]. Moreover, to increase the bandwidth by utilising available spectrum more efficiently, a mechanism called *dynamic spectrum management* is also investigated [129]. In order to facilitate this operation, a new intelligent radio network called *cognitive radio* in which the radio can change its parameters according to the channel conditions, is introduced in [130].

However, these approaches do not solve the problems arising from the bandwidth

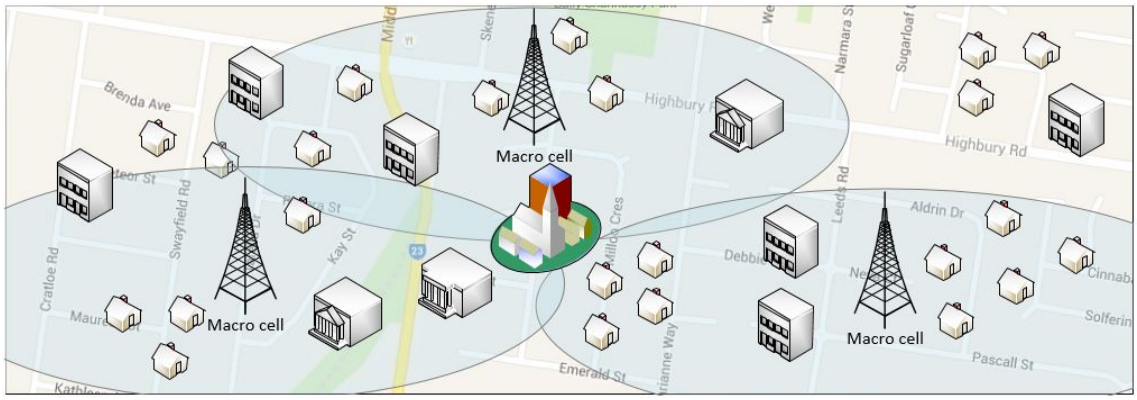
requirements completely as the radio spectrum is a finite resource. Therefore, the growth rate of technology lags behind the massive growth of the mobile traffic. However, technology upgrades alone cannot fulfil the requirements in relation to ubiquitous access. In order to facilitate both high bandwidth and ubiquitous access, the deployment of heterogeneous network is considered as an effective strategy today [131,132]. In this approach, small cells are deployed to supplement existing macro cell network. A small cell is basically a BS that provides lower reach compared to typical macro-cellular BS and hence uses a lower transmission power. These small cells are expected to be deployed inexpensively in light poles or sides of buildings. Such heterogeneous network architecture is considered in our work and presented in this chapter.

A typical macro cellular network and a heterogeneous network that has both macro and small cells are shown in Fig. 3.2 (a) and (b), respectively. As shown, some of the areas that were not properly covered by the macro cellular network can be covered by deploying small cells. Moreover, in comparison to the network shown in Fig. 3.2(a), the number of users per cell is lower in the heterogeneous network shown in Fig. 3.2(b), which increases the per user capacity. It is clear that the deployment of inexpensive small cells can be used to increase per user bandwidth and also the network coverage. In addition to these major benefits, another advantage from a small cell network (SCN) from the end user's point of view is that it can improve the battery life of mobile devices. This is because mobile devices can be operated in low transmit power mode as the BSs are now closer to the end users. When transmit power is decreased, the battery life of a mobile communication device is increased.

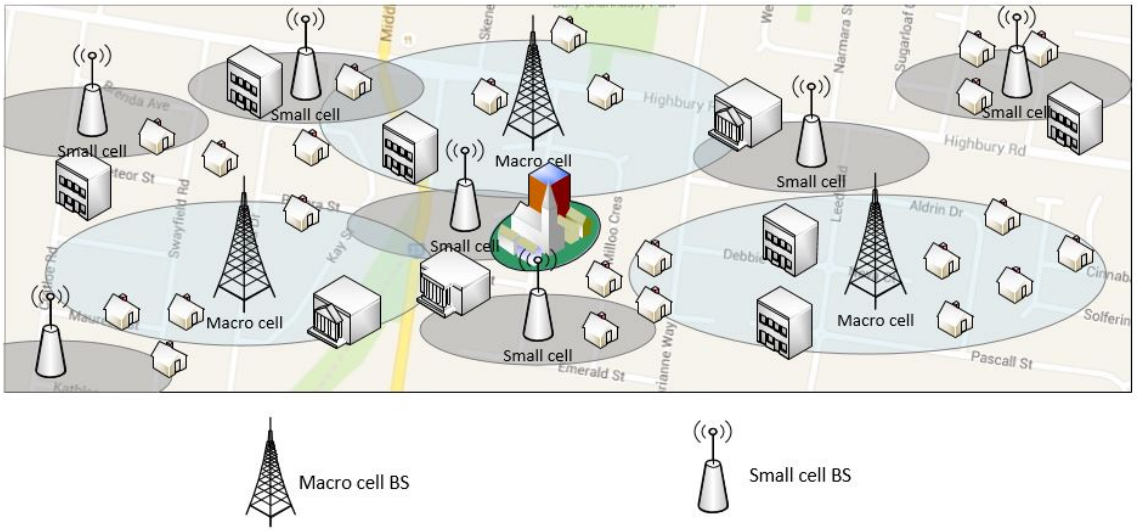
Despite these benefits, due to the high number of small cell deployments, a new set of challenges are introduced [131]. In the following subsections, we will discuss these major challenges and the approaches proposed to overcome those.

A. Coexistence of Small Cells with Existing Macro Cellular Networks

One of the key challenges of deploying small cell deployment is the method in which the small cells cooperate with the already existing macro cell BSs in a heterogeneous network. In particular, the most challenging aspect of this is when the same set of users



(a)



(b)

Figure 3.2: (a) Macrocellular network (b) Heterogeneous network.

is serviced by both macro and small cells. The aforementioned scenario is graphically presented in Fig. 3.2(b). As shown, the network consists of two tiers of cells. In this case, the coordination between the two tiers of cells is important to mitigate interference [133]. This is due to the typical mobile devices being set to communicate with the BS from which it can receive the highest signal strength. In a general environment, this would be the high power macro cell. As a result, most users will tend to connect with the macro cells. Thus, even distribution of load amongst the multiple tiers to take advantage of the full capacity provided by the heterogeneous network becomes an important challenge. Interference cancellation techniques in the user equipment and the coordination between

two tiers have been studied in order to overcome this challenge [133].

Another challenge is the usage of a single frequency spectrum amongst different tiers of heterogeneous networks [134]. The authors of [134] discussed different approaches such as frequency planning and implementation of power control mechanism to mitigate cross-tier interferences in the heterogeneous network. Their simulation results indicated that the frequency planning techniques can potentially be one of the most effective approaches to reduced interference issues in data channels. Moreover, a theoretical framework to analyse the performance of a multi-tier heterogeneous network that consists of macro and pico cells is discussed in [132]. In particular, its authors use a random spacial model to analyse the coverage performance. Their simulation results showed that the user experience can be significantly increased by introducing the multi-tier network in the wireless access domain.

Furthermore, the standardisation bodies have also taken heterogeneous network operations into account when standardising wireless technologies. These bodies have included different mechanisms in their latest releases to cope with the interference issues related to the SCN deployment. For example, 3GPP included the enhanced inter-cell interference coordination (ICIC) mechanism for the heterogeneous network operation in their LTE standard [135].

B. Self-Organising Capability

As large numbers of small cells are expected to be deployed, managing each and every small cell individually would be a complex task. The capability of self-configuring, planning and optimising the parameters up to some extent is important in reducing the operational costs of large scale deployments. Therefore, the self-organising capability of SCNs is considered an important deployment requirement. Depending on the traffic load, channel conditions, and the performance of neighbouring cells, a small cell BS is expected to independently optimise its parameters such as transmit power and frequency channels. In [136], the authors described how different techniques can be implemented to overcome some of the challenges related to small cell deployments. These techniques include self-organising network (SON), multi-input multi-output (MIMO), carrier aggre-

gation, and inter-cell interference coordination (ICIC). The authors further described the SON by categorising it into two types: distributed and centralised.

The importance of using self-organising to overcome complications related to small cell planning is discussed in [131]. The same study reviews the issues arising from frequent handovers due to the low radii of small cell BSs. In a subsequent publication [137], the authors further investigate how large scale SCNs could be efficiently analysed considering parameters such as path loss, BS cooperation, inter-cell interference, and limited backhaul. In their work, the authors used random matrix theory to analyse the performance of SCNs.

C. Backhauling Small Cells

In a hierarchical telecommunications network, the backhaul portion of the network comprises the intermediate links between the core network and the small cells. Backhaul cost is estimated to be 30% of the overall mobile network cost [138]. With the deployment of high numbers of small cells, the cost of backhaul links is predicted to further increase rapidly. Therefore, providing a cost-effective backhaul that also satisfies the requirements of the diverse mobile services become a challenge in realising the SCN [139].

Different backhaul technologies for SCNs are discussed in [140, 141]. These technologies can be mainly categorised into two types: wired and wireless. Direct fibre, xDSL, and passive optical networks (PONs) are examples of wired access technologies discussed in these studies. However, microwave (6 - 60 GHz), millimetre wave (60 - 80 GHz) and satellite technologies are considered as wireless backhaul options. These millimetre wave and microwave radio frequency (RF)-based backhaul technologies can be again categorised into line of sight (LOS), non line of sight (NLOS), and near line of sight (nLOS) technologies, depending on their propagation capabilities. In [141], the authors proposed the use of a combination of different technologies for backhauling small cells where they focus on the delay, reach, and bandwidth that can be achieved through each of these different options.

Amongst the backhaul technologies studied for SCNs, optical fibre technology is identified as the most suitable for the cellular networks as it provides high bandwidth

and fault tolerant access [142]. However, due to the high initial costs and longer roll-out times in greenfield scenarios, deploying new fibre networks to backhaul small cell is still questionable. Nevertheless, to provide cost-effective fibre backhaul for small cells deployment, several solutions have been investigated in literature. The work reported in [142] pointed out that the backhauling cost can be saved by deploying wireless backhaul links as an interim solution, until fibre infrastructure is in place.

In addition, leveraging existing wire-line networks is another smart approach to provide fibre backhaul closer to the small cells. The usage of VDSL2 and Gigabit passive optical network (GPON) for small cell backhauling is discussed in [143]. The authors pointed out that using these two most popular wire-line access networks, service providers will be able to provide proper backhaul for small cells while reducing the total cost of ownership. In particular, PON is considered as a long-term solution for small cell backhauling as per the higher bandwidths offered by PON technologies. This is because, depending on the bandwidth requirement of a small cell or any other user, the split ratio of a PON can be adjusted. Therefore, approaches to leverage existing fibre resources and to use PON for the purpose of backhauling small cell warrant comprehensive research investigations.

3.3 Energy-Efficiency of Heterogeneous Networks

With increasing awareness of global warming and environmental consequences of information and communications technology (ICT), researchers have been seeking solutions to reduce energy consumption [144, 145]. Cellular networks, as the main contributor to the network power consumption, have drawn considerable attention in literature [146–150]. The deployment of heterogeneous network with low power small cells is considered to be a potential approach to reduce the overall power consumption of the radio access network (RAN) [151, 152]. However, as an abundant number of small cells are expected to be deployed, the energy consumption arising from a SCN can be significant although it is lower than a typical macro cell deployment. Therefore, improving the energy-efficiency of the heterogeneous networks is considered as one of the most impor-

tant aspects. As a result, an innovative new research area called *green cellular networks* concentrating on environmental influences of cellular networks has been formed and has attracted considerable research interest. For mobile operators in particular, another motivation and objective of *green* approaches is to gain extra commercial benefits, mainly by reducing operating expenses that arise from energy costs [148, 150]. Various distinctive approaches to reduce energy consumption in a mobile cellular network have been proposed. Next, we classify these approaches broadly into five different domains, as detailed below.

A. Green Hardware Components

This approach aims to improve hardware components (e.g., power amplifiers) with more energy-efficient designs. In [153], the authors present an envelope tracking wideband code division multiple access (WCDMA)/ WiMAX BS amplifier based on high-voltage heterojunction bipolar transistors (HV-HBTs) that satisfies the linearity requirements of WCDMA and WiMAX standards. In [154], the authors propose a highly efficient feed-forward amplifier using a RF high power Doherty amplifier which utilises a single push-pull packaged LDMOS FET, which generally has a low power added efficiency (PAE) of 6 - 10%. In [155], the authors developed a broadband class-AB GaAs HBT power amplifier that demonstrates high linear output power and low distortion at very low operating currents for the third-generation CDMA and fourth generation LTE applications. However, the performance of most components used in current cellular network architecture is unsatisfactory from the energy-efficiency stand point. Considering, for example, the power amplifier, the component consuming the largest amount of energy in a typical cellular BS, more than 80% of the input energy is dissipated as heat. Generally, the useful output power is only around 5 - 20% of the input power [151]. The authors in [151] also showed that the potentially optimised ratio of output power to input power for power amplifiers (power efficiency) could be as high as 70%. Accordingly, a substantial amount of energy savings can be achieved if more energy-efficient components are adopted in the cellular networks. However, the implementation costs of these approaches are high when considering larger coverage and the need for high power efficiency. Therefore,

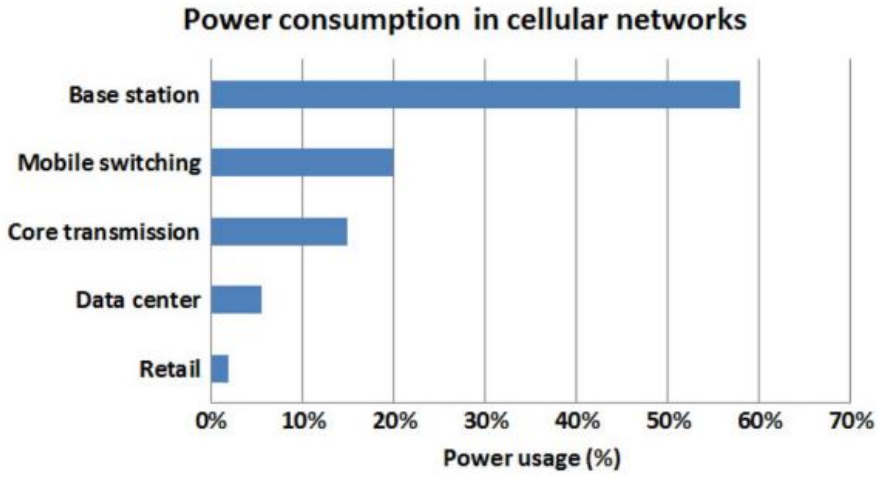


Figure 3.3: Breakdown of energy consumption in cellular networks (Source: Vodafone [5]).

careful consideration in both operational and economical aspects by network operators is required before decisions on hardware replacement are made.

B. Sleep-mode for BSs

Sleep-mode approaches aim to selectively turn off some resources in existing network architecture during non-peak traffic hours [156–160]. Approaches in this category generally try to save energy by monitoring the traffic load in the network and then decide whether to turn off (or switch to *sleep-mode*, also referred to as *low power mode* or *deep idle mode* in some literature), or turn on (or switch to *active mode*, *ready mode*, or *awake mode*) certain elements in the network. Unnecessary energy consumption contributed by, for example, air conditioning under-loaded BSs, can be avoided by adopting such sleep-mode mechanisms. These approaches generally involve switching certain elements including (but not limited to) power amplifiers, signal processing unit, cooling equipment, the entire BS, or the whole network between *sleep-mode* and *active mode* [161]. Most often, sleep-mode techniques focus on turning off BSs during the off-peak traffic hours. This is due to the fact that BSs contribute to the highest proportion of energy in wireless networks [5]. Fig. 3.3 presents a breakdown of energy consumption in cellular networks.

We will now discuss some of the most effective sleep-mode mechanisms that have

been proposed in the literature. In order to reduce the energy required to transport a single bit, the authors in [156] introduced sleep-mode operation to the small cell network. In sleep-mode operation, the cells that do not have any active users are switched off. Their simulation results indicated that using these methods a significant amount of energy can be saved in the RAN. Moreover, authors also studied the implications of cell size reduction on the energy consumption of the RAN. Their results showed that the reduction of cell size leads to low energy consumption in comparison to typical macro cell network.

In [157], the authors adopt an analytical approach towards designing optimal sleeping parameter and transmit power (and thus service rate) jointly for energy-delay tradeoffs considering non-real time user requests arrival at the BSs. In [159], the authors introduce sleep-mode transmission for BSs in a LTE cellular network, termed as discontinuous transmission (DTX). The simulation results show that cell DTX is most efficient when the traffic load is low in a cell. The technology potential for a metropolitan area is shown to be 90% reduced energy consumption compared to no use of cell DTX [159].

Furthermore, a new class of sleep-mode algorithms was introduced to improve energy-efficiency of SCNs in [160]. In these algorithms, the hardware components of the BSs are completely switched off in the idle mode. Additionally, the BSs enter to the sleep-mode depending on their traffic conditions. The authors further discussed three different strategies to control sleep-mode operation. These strategies are core network driven, small cell driven and user equipment driven operations. With the help of simulations, authors showed that these algorithms can be said to effectively reduce the energy consumption of the heterogeneous wireless access network.

In [162], the authors investigate the energy savings that are possible for existing mobile networks by exploiting daily variations in network traffic. This saving is obtained by introducing a feature that dynamically puts low loaded cells into sleep-mode, during which radio equipment is effectively powered down, reducing the networks energy consumption. Throughout this investigation, the variation in network traffic is achieved by changing the number of users being simulated.

C. Radio Transmission Process Optimisation

Radio transmission process optimisation approaches work on the physical or MAC layer. Advanced techniques including MIMO technique, cognitive radio transmission, cooperative relaying, channel coding, and resource allocation for signalling have been studied to improve energy-efficiency of telecommunication networks (e.g., [5, 163–167]). A variety of approaches have been proposed to efficiently utilise resources in time, frequency, and spatial domains to achieve energy saving. Similar to approaches based on sleep-mode, this family of approaches does not generally require upgrade of hardware components in the system. However, trade-offs between energy-efficiency and other performance metrics, such as delay and jitter of the network are inevitable. Based on information theory, four fundamental trade-offs related to energy-efficiency on wireless networks have been acknowledged, namely deployment efficiency-energy efficiency, spectrum efficiency-energy efficiency, bandwidth-power, and delay-power [168, 169].

D. Planning and Deployment of Cells

As the title implies, these approaches aim to increase energy-efficiency by deploying small cells, including micro, pico, and femto cells, in the cellular network [170]. As discussed previously in Section 3.2.2, these smaller cells serve small areas with dense traffic with low power consuming cellular BSs, which are affordable for user-deployment and usually support plug-and-play feature. In contrast to conventional homogeneous macro cell deployment, such heterogeneous deployments reduce energy consumption in the network by shortening the propagation distance between nodes in the network and utilising higher frequency bands to support higher data rates. The major constraint of these approaches is that the extra small cells incur additional radio interferences as compared to conventional homogeneous macro cell networks, which might negatively affect user experience. Meanwhile, if too many micro, pico or femto cells are deployed, the benefit of energy savings may even be reversed because of extra embodied energy consumed by newly deployed cells as well as overhead introduced in transmission. Therefore, the number of extra smaller cells, as well as their locations, needs to be carefully planned in order to reduce total energy consumption. It has also been noticed that integrating

heterogeneous network deployment with previously discussed sleep-mode schemes can potentially achieve significant gains in terms of energy saving [171–173].

In [171], the authors propose a vertical sector antenna deployment scheme and demonstrate that this approach can yield significant operational energy and transmission efficiency savings. The authors also present a comprehensive performance sensitivity analysis considering a traditional hexagonal homogeneous cell deployment scenario, which yields insight into both the validity of results and the parameters which can significantly influence future development. In [173], the authors propose a clustering-based power saving algorithm (CPSA) for self-organised sleep-mode in femto cell networks. The CPSA algorithm first builds a leader-member cluster framework, in which each femto cell BS (FBS) is either a FBS leader (FL) or a FBS member (FM). Then, the FL acts as an autonomous entity and is responsible for detecting active calls in the cluster coverage, whilst the FMs without the active user can entirely shut down pilot transmissions and the related processing all the time. Furthermore, the authors provide some guidelines for deploying energy-efficient femto cell networks.

E. Renewable Energy Powered Cells

Compared to current widely used energy resources such as hydrocarbon which produces greenhouse gases (GHGs), renewable resources such as hydro, wind, and solar power stand out for their sustainability and environmental friendliness [174]. In underdeveloped areas, solar power operated cellular BSs have been deployed by the network providers. For example, in Nigeria, where roads are in poor condition and unsafe, delivering traditional energy resources such as diesel for off-grid BSs cannot be guaranteed [175]. Energy harvesting techniques, namely exploiting available energy from such renewable resources to complement existing electric-operated infrastructure, would probably be the long-term environmental solution for the mobile cellular network industry. Especially for those areas without mature network infrastructure, deploying energy harvesting networks would be ideal. For developed countries with completed infrastructure, however, the same question of embodied and replacement cost arises as the component-based approaches, as discussed previously in subsection A. When service migrates from

the obsolete electric-operated BSs to the new energy harvesting BSs, it becomes technically challenging to preserve fault tolerance and data security without any service interruption.

A comparison of advantages and disadvantages of the above discussed energy saving approaches is summarised in Table 3.2. Out of the five commonly investigated energy saving approaches, we recognise sleep-mode techniques as the best way forward due low cost and ease of implementation (as no hardware upgrades are required). Therefore, our work in this chapter is mainly focused on introducing sleep-mode techniques to the BSs. Cell zooming is a relatively novel sleep-mode technique whereby the BSs are enabled to self-configure their configuration including the coverage radius. In the next subsection, we discuss the cell zooming concept in detail.

Approach	Advantages	Disadvantages
Green hardware components	Largest reported savings, direct and intuitive	Upper limit for improvements, high cost for hardware replacement
Sleep-mode techniques	Easier and less costly for testing and implementation	Trade-off between performance and saving, current modelling not accurate enough
Radio transmission process optimisation	Low cost, various applications	Trade-off between performance and saving, errors due to uncertainty issues
Planning and Deployment of Cells	Low cost to implement, user-oriented, high potential savings	Introduces new issues such as radio interference
Renewable energy powered cells	Long-term solution for off-grid BSs	High replacement cost and limited gain for existing on-grid BSs

Table 3.2: Comparison of green cellular network approaches.

Cell Zooming for Heterogeneous Networks

In this section, an overview of the concept called cell zooming (also referred to as cell breathing) is presented. A technique named *self-organising network* (SON) is introduced in the 3GPP standard (refer 3GPP TS 32.521 for definition and 3GPP TS 36.902 for cases and solutions) [176], to be gradually implemented in BSs along with the fourth generation standards including LTE and WiMAX. It introduces automatic network management

as well as intelligence features to the system and thus reduces costs, improves performance and increases flexibility of the cellular system through network optimisation and reconfiguration process. SON enables the BSs to adjust their own configuration when necessary without human intervention. Therefore, automated operations such as timed sleep-mode, user location prediction, and reverse channel sensing are possible in the system [177]. Sleep-mode in BSs is one of the various applications of SON, where BSs are enabled to act collectively to save energy by redistributing traffic and sharing traffic information amongst BSs.

Cell zooming or *cell size breathing* is a similar concept to SON but provides a higher level of flexibility. Cell zooming is a network layer technique adaptively adjusting the cell size according to traffic conditions by adjusting antenna tilt angles, height, or transmit power [162, 178]. It is much simpler than switching a BS off entirely from an implementation perspective [178]. It can be applied to balance the traffic load and reduce the energy consumption. When the traffic load in a certain cell increases, the cell will zoom in to reduce the coverage area and therefore avoid possible congestion. The *service hole* created by this will be taken care of by the neighbouring cells with less traffic, which are supposed to zoom out. A cell zooming server, which can be implemented virtually at the gateway , or distributed in the BSs, controls the procedure of cell zooming. It sets the zooming parameters based on the traffic load distribution, user requirements, as well as channel state information. In fact, zooming the cell coverage to zero is equal to switching off the entire BS. Therefore, cell zooming can be perceived as a generalisation of BS sleep-mode. In our work, we exploit the use of cell zooming capability of the small cell BSs to achieve optimal transmit power control throughout the day.

3.4 Related Work

In this section, we discuss the related work in literature that focuses on energy-efficient BS sleep-mode techniques for heterogeneous networks. As discussed in Section 3.2.2, heterogeneous network deployment schemes were originally designed to improve the spectral efficiency in the cellular networks by offloading traffic from macro cells to smaller

and easier to handle cells, and may lead to increase in energy consumption because of the large number of small cells deployed. Nevertheless, with the introduction of the sleep-mode in BSs, heterogeneous cellular networks can now outperform traditional macro cell-only counterparts in terms of energy-efficiency. During peak traffic hours, more energy-efficient smaller cells can be deployed, replacing some of the macro cells in the macro cell-only network. Then, those smaller cells are transitioned to sleep during light traffic hours with the remaining macro cells being able to maintain throughput and coverage.

In [160], the authors studied the application of dynamic sleep-mode in BSs with pico-cell deployment. A pico-cell is a small mobile BS connected to the cellular network via the Internet, typically used to improve coverage indoors and considered to be smaller than a small cell BS. Heterogeneous network planning can improve the coverage of the cellular network, but will likely result in even more severe over-provisioning, thus consuming more energy if the cells are unable to adapt to traffic load. The solution proposed in [160] is to introduce the dynamic sleep and wake modes in pico cells. The result shows that a network with both macro and pico cells, where dynamic sleep-mode algorithm is applied in pico cells, consumes less amount of energy than the network with only macro cells. In [179], the authors proposed an energy model for heterogeneous cellular network and a cross layer optimisation method. Several pico cell BSs (lower layer) are in the coverage area of one macro cell BS (upper layer). The authors solve the problem of how to associate users to the group of macro cell and pico cells, so that energy consumption is minimised after lightly loaded pico cells are switched to sleep-mode. Another similar model is presented in [180].

In [181], the authors particularly addressed the inter-tier interference amongst multi-tier heterogeneous cells. In their work, macro BSs are modelled by a Poisson point process (PPP) whilest users are distributed according to different stationary point processes. A Bernoulli trail-based random sleeping technique and a strategic sleeping technique aimed at maintaining coverage are examined for both homogeneous and heterogeneous network architectures. The strategic sleeping based on activity of macro cell users is designed to maintain or improve coverage probability of users as in the non-sleeping case.

Numerical results presented in [181] showed that random sleeping is detrimental to the energy-efficiency. The authors also showed that the gain in energy-efficiency saturates as the density of smaller cells reaches a certain level.

In [182], the authors investigated how small cell access points (SAPs) can play a role in enhancing energy-efficiency of heterogeneous cellular networks. Sleep-mode of SAPs corresponds to the trade-off between energy consumption and false alarm rate. The authors note that bursty traffic from macro cell traffic, due to mobility of users, makes duty cycling of sleep-mode in smaller cells more complicated. Similar to [181], a PPP is also used to model the locations of SAPs and macro BSs.

In [183], the authors proposed a similar sleeping strategy by which the small cells are switched off when the cell is not heavily loaded and the macro cell can serve the overall traffic without deteriorating the QoS. Based mainly on queuing theory, the work utilises a continuous time markov decision process (CT-MDP), in which states represent load status of each BS. Every user brings a certain load to its connected BS. Each possible action for the state and transition probabilities is assigned a value of rewards/cost. The cost function is defined as an increasing function of energy consumption and a decreasing function of target throughput, a QoS measure. The switching operation is added to the state space as a new dimension. Apart from the straightforward case that BSs have complete information of its associated traffic, optimal solutions have also been found for partial traffic information and delayed information (by transforming their MDPs into equivalent MDPs without delay).

In another study [173], the authors proposed a clustering-based power saving algorithm for self-organised sleep-mode in femtocell networks. Femtocells are semi-autonomous thus they sense the best frequency and radio parameters to use from the immediate environment. They are installed, powered and connected by the end user or business with less active remote management by the network operator. In the cluster construction process in [173], the leader of each cluster is first elected based on the sum of received pilot signal power and the distribution density, and then other femto BSs determine whether they are to be attached to the leader on the basis of the pilot signal power. The leader and members in the same cluster will then exchange information collected by a sniffer

installed with each femtocell BS. The member will only turn on the pilot transmission and the processing if notified by the leader and the received signal energy rise above a certain threshold within a predefined period of time.

The locations of BSs, either macro or micro, in heterogeneous networks are often assumed to follow a Poisson point process (PPP) [181, 184, 185]. In this regard, capacity extension by additional micro cells (increase density of BSs) and energy saving by BS sleeping (decrease density of BSs) can be generalised into a single optimisation problem on BS density based on the cost per micro BS. In [184], the authors illustrated numerical calculation to obtain the optimal BS density for both homogeneous and heterogeneous networks.

In another study [185], the authors considered a scenario where users in macro and micro cells have different traffic patterns. They assume that micro cells serve hotspots with higher traffic volume. The authors investigated three energy saving approaches including micro cell BS sleeping and expansion/shrinking coverage of micro cells (similar to cell zooming). The coverage and power consumption of macro cells are held constant. It is shown that each approach is effective under different traffic conditions. The crucial factor affecting the performances of different approaches is highlighted as the traffic rate ratio, namely the ratio of traffic rate per unit area in hotspots to that in non-hotspots.

The above mentioned approaches all use simulations or stochastic geometry methods which are computationally complex to model the energy-efficient heterogeneous network with cell zooming capabilities. In our work, we take a novel and a relatively straightforward approach in solving this problem. Our aim is to calculate the optimal power levels mathematically using two approaches: a heuristic-based algorithm, and an optimisation problem. In both these approaches, we focus on dynamically changing the transmission power of small cell BSs in accordance with the required capacity variations during the 24-hour span of a day. Such an investigation had never been performed in the above mentioned energy saving approaches.

In the next section, we present a novel heuristic-based transmit power control algorithm to estimate the optimal transmission powers for the small cell BSs in a heterogeneous network environment.

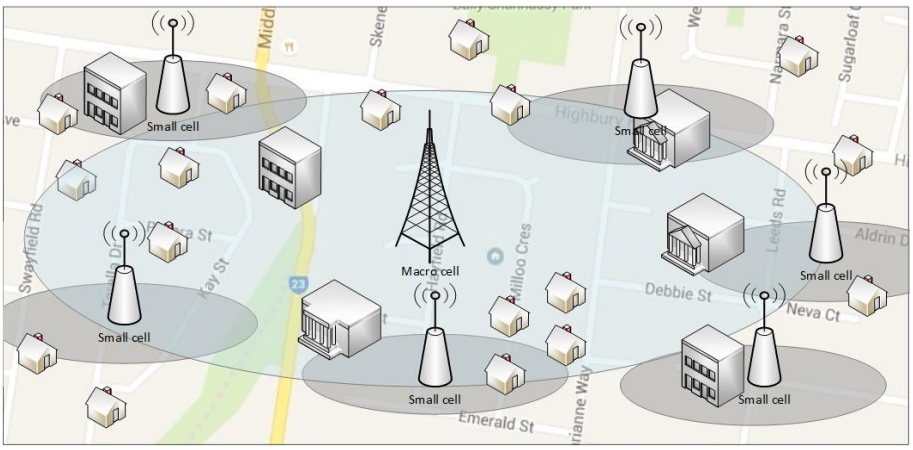


Figure 3.4: Considered single macro cell access network segment.

3.5 Heuristic-based Transmit Power Estimation for Small Cell Networks

This section presents a heuristic approach to solve the transmit power control problem for heterogeneous networks. In this work, we consider a single macro cell system in a cellular network, which contains a single macro cell BS and a set of N small cell BSs as shown in Fig. 3.4. However, the heuristic transmit power control algorithm presented is easily scalable to model a larger wireless access network segment which consists of multiple macro cell systems.

The heuristic-based power control algorithm presented in this section aims to derive the optimal transmission power for the small cells ensuring a data rate threshold and a signal strength. Note that the proposed approach controls the transmit powers of the small cell base stations of the SCN and makes no adjustments to the functionality of user equipment, thus can be implemented in the current heterogeneous network environment. Generally, user equipment is designed to identify and connect to the base station that provides the highest SINR at the user location. Moreover, we assume the user locations to be static to maintain the simplicity of the transmit power calculations. The algorithm estimates the optimal operational transmit power by following an iterative process. We will now discuss the centralised small cell power control algorithm in detail.

3.5.1 Centralised Small Cell Power Control Algorithm

The heuristic-based centralised power control algorithm is presented in this section. The algorithm takes the location coordinates of users, small cells, and macro cell as inputs and produces the estimated optimal operational transmit powers of the small cells in order to ensure a capacity threshold and a signal to interference and noise ratio (SINR) for the user.

As shown in Fig. 3.4, we consider a single macro cell system in a cellular network, which consists of a set of N small cells. The macro BS is considered to be located at the centre of the considered 1 km x 1 km area where the small cells are randomly placed in the region. The algorithm operates in a centralised manner, where the macro BS calculates the set of optimal transmit powers for the small cells to maximise the energy-efficiency. These values are then communicated to the small cells periodically. Here, we consider that all BSs transmit a reference signal in both a working state and a sleep state so that a sleeping cell could be activated on demand by the macro cell BS [186]. When a BS is in sleep-mode, it transmits its reference signal and system information less frequently than during regular operation, thus maintaining the sleep-mode until an *wake up* indication is received.

The heuristic-based power control algorithm consists of three steps as listed below.

- Step 1: User assignment and load balance
- Step 2: Iterative cell zooming
- Step 3: Toggle cells to sleep-mode

Step 1: User assignment and load balance

In this step, the algorithm considers all users to be assigned to their nearest small cell. Here, we do not assign any users to the macro cell. Second, a load balancing sequence is run in order to check if the small cells have enough resources to transmit to all the assigned users. In the case where a particular small cell is over-loaded with more users than it can serve, the users are handed over to the next nearest cell with available resources. The pseudocode of this step is presented in Algorithm 4.

Algorithm 4 Initial user assignment and load balance

```
1: user assignment:
2: for each user u ϵ U do
3:     Find nearest small cell sc
4:     Assign u to sc
5: load balancing:
6: for each small cell sc ϵ S do
7:     Set P_t to P_sc,max
8:     if number of assigned users > available resource blocks then
9:         while number of assigned users ≥ available resource blocks do
10:            Find furthest user u_f
11:            Assign u_f to next nearest small cell with available resources
12:    for each assigned user u_sc do
13:        Calculate received data rate (RDR)
14:        if received data rate < threshold data rate then
15:            Assign u_sc to next nearest small cell with resources
16:            if RDR requirement not satisfied with any small cell then
17:                Assign u_sc to macro cell
```

Algorithm 5 Iterative cell zooming

```
1: Sort small cells by distance to macro cell (ascending)
2: repeat
3:     for each small cell sc ϵ S do
4:         Find furthest connected user u_f
5:         Calculate transmit power that ensures threshold data rate (P_RDR_th,u_f)
6:         Set small cell transmit power to P_RDR_th,u_f
7: until P_t of all small cells are converged
```

Step 2: Iterative Cell Zooming

The pseudocode of the iterative cell zooming phase of the algorithm is presented in Algorithm 5. In this step, the algorithm iteratively adjusts the transmit power of the small cells to ensure that the received data rate (RDR) for each connected user is maintained above the received data rate threshold (RDR_{th}). The aim is to derive the minimum transmission power required to serve the assigned users. Optimally, when the optimal transmit power of a small cell is reached, the user that is located the furthest from the cell receives exactly the threshold data rate (RDR_{th}).

In Algorithm 5, first, the set of small cell BSs are sorted according to their distance

Algorithm 6 Toggle cells to sleep-mode

```
1: Sort small cells by the number of connected users (ascending)
2: for each small cell sc ϵ C in sorted list do
3:     if number of users < number of available resource blocks in macro cell then
4:         Check if RDR_th is ensured from macro cell for all users
5:         if RDR_th is ensured then
6:             Assign users to macro cell
7:             Switch sc to sleep-mode
```

from the macro cell BS. Second, starting from the small cell which is located closest to the macro cell, the algorithm reduces the transmission power of the small cell such that all the users assigned to that particular small cell receives at least the threshold data rate. Here, we assume that when the furthest connected user is ensured a threshold data rate, other users who are located closer to the BS are also guaranteed to receive the threshold data rate. When the transmit power of a particular BS is changed, that will in turn affect the signal noise for the users connected to the neighbouring small cells. As a result, the SINR for the users connected to the neighbouring cells is changed. Therefore, the algorithm will repeat the transmit power adjustment process until the change in transmit powers of all small cells between two iterations is negligible. The transmit power of the BSs are considered to have been converged to their final values at this point. This way, the sub-optimal transmit power that guarantees the threshold data rate for all the connected users is calculated for each of the small cells.

Step 3: Toggle cells to sleep-mode

In this step, the algorithm searches for small cells serving a low number of users in an attempt to determine if a small cell could be switched to sleep-mode to improve energy-efficiency of the heterogeneous network. The small cells with low loads (fewer number of assigned users) are considered first. If all the users can be reassigned to either a nearby small cell or the macro cell, the algorithm determines that small cell as a candidate to be turned to sleep-mode. Here, we assume that the macro cell BS is able to maintain coverage throughout the entire considered area even when a set of small cell BSs are toggled to sleep-mode. A pseudocode of this step is presented in Algorithm 6.

Here, the algorithm aims to toggle as many small cell BSs as possible to derive the most energy-efficient cell configuration. The small cells are first sorted by the number of connected users in an ascending order. Then, starting from the least loaded small cell, the algorithm reassigns the users from the SCN to the macro cell BS. The small cell BSs are then toggled to sleep-mode. These optimal transmit power values generated from the algorithm are then communicated to the set of small cells via a power control reference signal by the macro cell station. In the following section, we evaluate the performance of the presented algorithm.

3.5.2 Model Evaluation

We evaluate the heuristic-based centralised power control algorithm proposed above using a C# simulation. We simulate a heterogeneous network segment of one macro cell BS with multiple small cell BSs deployed around it. The simulated area is 1 km x 1 km. The macro cell BS is located in the centre of the area of concern. The small cell locations are derived from a uniform random variable sequence. The users are considered to be scattered through the area and their locations are also derived from a uniform random variable sequence. Population density of the simulated area is controlled by adjusting the considered number of users in the fixed geographical area of 1 km^2. It is also ensured that two neighbouring small cells are at least 50 m away from each other and at least 100 m away from the macro cell.

Further, as we consider LTE technology in our heterogeneous network simulations, we consider orthogonal frequency division multiple access (OFDMA)[1] and single carrier frequency division multiple access (SC-FDMA)[2] to be used for downlink and uplink multiple access schemes, respectively. The macro cell station and each small cell station are considered to have N_M and N_S resource blocks, respectively. Each user can be allocated

[1]OFDMA is a multi-user version of a digital modulation scheme orthogonal frequency division multiplexing (OFDM). OFDMA is a modulation and access technique that combines both time division multiple access (TDMA) and frequency division multiple access (FDMA) technologies. In OFDM, the signal is first split into independent sub-carriers and these closely spaced orthogonal sub-carriers are used to carry the data. The data is divided into several parallel data streams or channels, one for each sub-carrier. [187, 188]

[2]SC-FDMA is a hybrid modulation format that combines the low peak to average ratio offered by single carrier systems with the multipath interference resilience and flexible sub-carrier frequency allocation that OFDM provides.

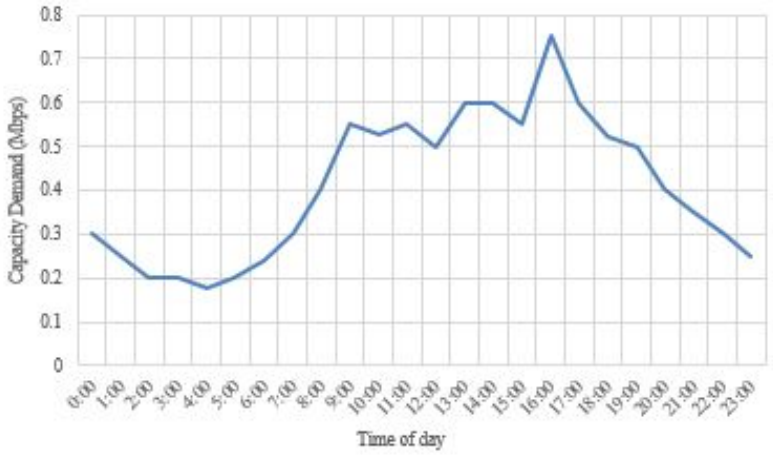


Figure 3.5: Hourly capacity demand variation per user during a day [6].

only one resource block. The macro cell station is considered to be always in a working state, while each small cell BS can regulate it,s transmit power to adjust the coverage provided according to the control messages received from the macro BS. The capacity demand threshold (C_{th}) is considered to vary with the time of the day as presented in Fig. 3.5. This data was derived from a study that estimated network load of voice calls and data demands over a mobile access network with traffic data from a mobile operator [6]. Furthermore, in this study, we consider the users to be static, thus not accounting for cell handovers. Movement of the users would increase the computational complexity of the algorithms and is therefore considered out of scope in this work. The other model parameters used for the simulations are listed in Table 3.3. We define three simulation scenarios whereby urban, sub-urban, and rural environments are modelled by considering various population densities (as listed in Table 3.3) in order to evaluate the performance of the proposed heuristic-based algorithm.

Total transmission power

We define a performance parameter, *total transmission power*, in order to evaluate the algorithm performance. Here, the total transmission power consists of individual transmission powers of all BSs considered in our access network segment. The total transmission power is defined as follows.

Parameter	Values
Area	1 km x 1 km
Bandwidth (B)	$1.5x10^6$ Hz
Antenna gain and carrier frequency factor (K_0)	10^3
Path loss exponent (η)	4 (urban) 3 (sub-urban) 2 (rural)
Population density (inhabitants per km^2) [189]	>400 (urban) 150 - 400 (sub-urban) <150 (rural)
Interference noise power (σ^2)	10^{-10} W
$P_{sc,max}$, $P_{mc,max}$ [190]	6.3 W, 80 W
$P_{sc,avg}$	5W
Resource blocks in macro cell (N_M)	100
Resource blocks per small cell (N_S)	25
Number of iterations	100

Table 3.3: Parameters considered in the evaluation of the heuristic-based algorithm.

$$Total\ transmission\ power = P_{MC} + \sum_{i \epsilon N} P_i \tag{3.1}$$

where P_{MC} is the power of the macro cell BS and P_i represents the power of the i^{th} small cell. Note that in this model, only the transmitting powers of the BSs are considered in the problem formulation. Other power consuming elements such as power amplifiers and digital signal processing are not considered. Our focus is on the energy savings derived by optimising the small cell transmit power. We present the results obtained from the simulations in the next section. All result values are average values from 100 simulation cycles for each instance.

3.5.3 Results and Discussion

In this section, we present the results gathered from the simulations. First, we evaluate the proposed centralised small cell power control algorithm under various user densities in an effort to model a central business district (CBD) environment, where a high number of users come into the metropolitan area (to work during the daytime), and leave the area during the later hours of the day. Second, we classify three scenarios; urban, sub-urban,

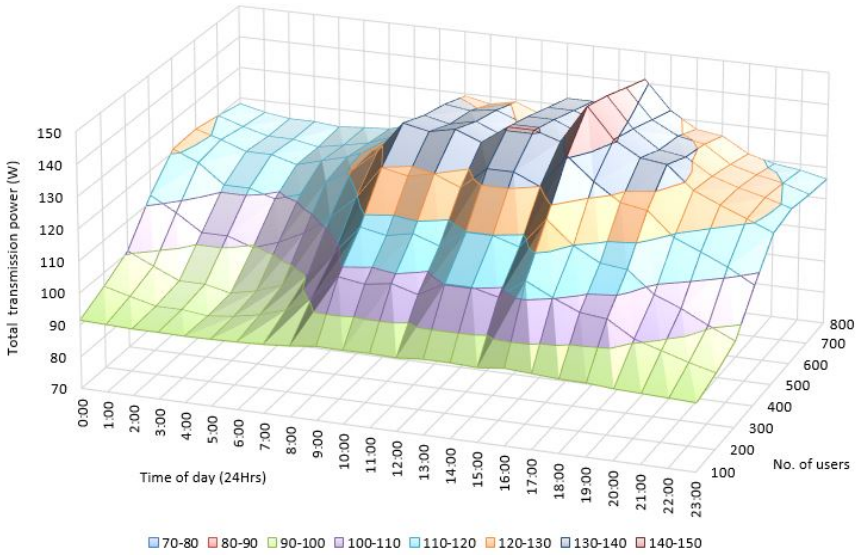


Figure 3.6: Total transmit power versus number of users under varying capacity demand throughout the day.

and rural environments to evaluate the solution for residential purposes. Further, we present results where the number of small cell BSs deployed in a unit area is varied in an effort to investigate the impact on the energy-efficiency from the deployed cell density.

Fig. 3.6 presents the total transmission power of the considered wireless access network segment under varying loads throughout the day. The network load is varied by adjusting the number of users connected to the network at a given time. The capacity demand per user is modelled as shown previously in Fig. 3.5. In the case shown in Fig. 3.6, the number of deployed small cell BSs is fixed at 20. The results show the highest transmission power recorded at the peak traffic hour (at 1600 hours) for all the user densities considered. When the network load is increased by increasing the number of connected users, the transmit power also increases, thus increasing the coverage of the network. However, when the number of users is low, the required transmission power has also reduced considerably allowing space for realisation of power savings. This could be exploited to send small cells to sleep-mode during the low traffic hours, and also in a CBD environment where the number of users vary considerably when users go home after work.

Fig. 3.7 shows the transmission power variations during peak traffic hour (at 1600

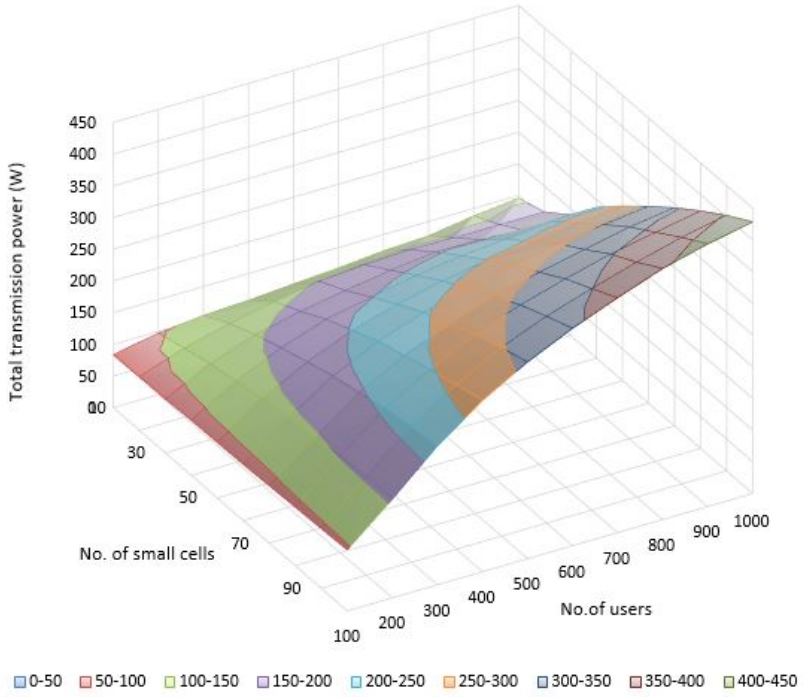


Figure 3.7: Total transmission power versus number of users and number of deployed small cells during peak traffic hour.

hours) as a function of number of users and number of deployed small cell BSs in the considered area. Results clearly suggest that the total transmission power increases gradually when the number of users is increased. This is due to more traffic being offloaded to the SCN. Moreover, when the number of deployed small cell BSs in the 1 km x 1 km area is increased, the total transmission power has also increased. When the user density is low, the transmission power remains low irrespective of the number of deployed small cells. This is due to the switching of the small cell BSs with low traffic load to sleep-mode.

Fig. 3.8 shows the total transmit power of small cells against the hourly capacity demand variations throughout the day. Here, we consider a network segment with 20 small cell BSs. Results indicate peak transmit power corresponding to the peak capacity demand at 1600 hours. Amongst the three scenarios, urban, and sub-urban environments indicate increased transmit power compared to the rural environment. This is due to more users being served by the small cells. The trend of the transmission power variation correlates to the required capacity thresholds enforced. The dotted horizontal lines in

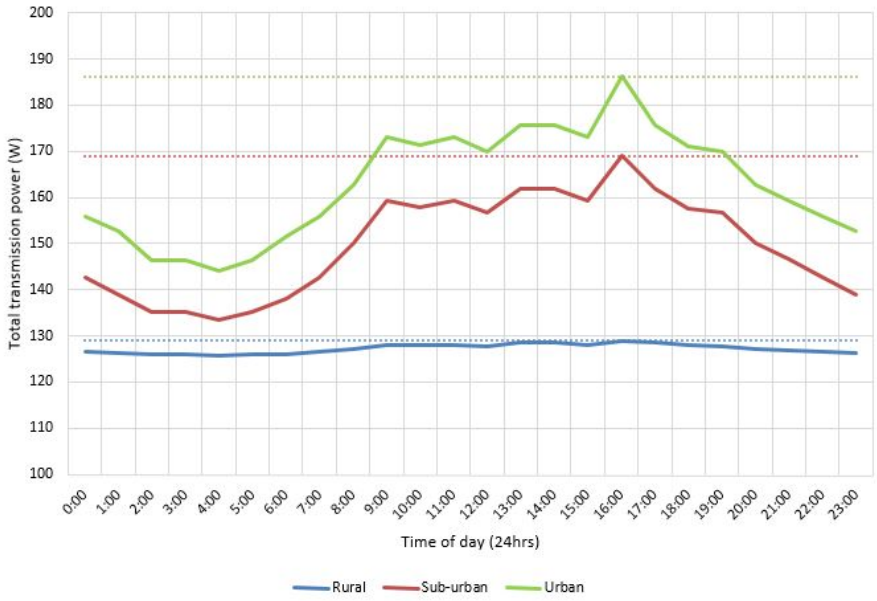


Figure 3.8: Total transmission power under urban, sub-urban and rural cases during the day.

Fig. 3.8 indicate the peak transmit power in each scenario. Similar to the case of the linear problem, we consider this power level as a baseline where the small cells operate continuously at this peak transmit power throughout the day. We evaluate the power savings of the considered heterogeneous network segment with respect to this baseline in Fig. 3.9. The power savings are averaged over the three population density cases. Results show maximum savings during low traffic periods and minimum savings during the peak traffic hour. The results show a 12% power saving compared to the baseline throughout the day.

Fig. 3.10 presents the average transmit power per small cell as a function of number of small cells deployed at the peak traffic hour (1600 hours). Results show that as the number of small cells increases, the average transmit power decreases due to users being shared amongst more cells and the increased freedom to choose a cell with higher signal strength. The average transmission power per small cell tends to saturate when the number of deployed small cells increases due to the total transmission power being shared amongst many small cells. This is evidence that over-densification of cells does not improve power savings though bandwidth capacity per user is increased. Further, we

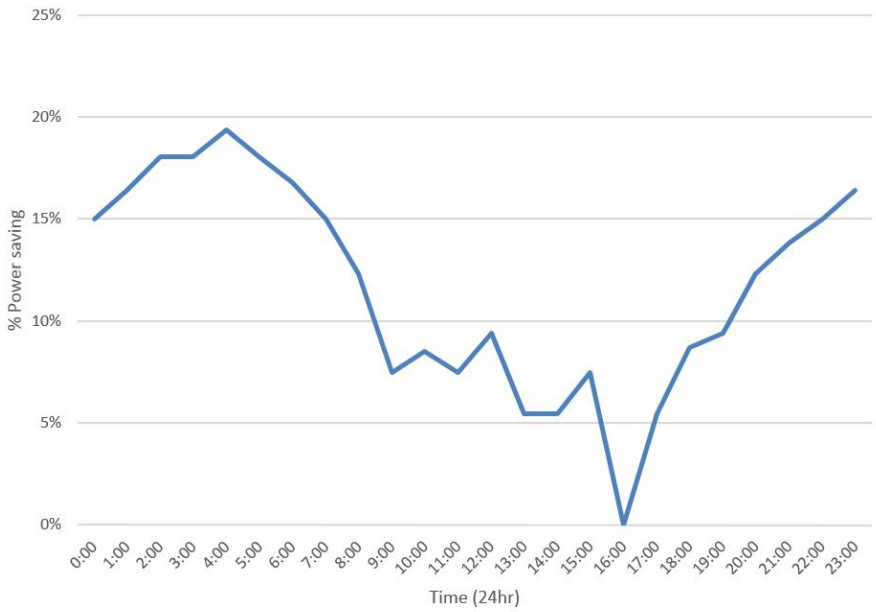


Figure 3.9: Percentage transmission power saving.

observe that the increase in population density will increase transmit powers. However, as the population density is increased, the average power per small cell becomes less dependent on it, as observed by the small difference between the urban and sub-urban curves.

Due to the iterative nature of the algorithm, we are able to derive a near-optimal solution for the problem considered. Further, due to the simplicity of the heuristic algorithm, we are able to simulate larger network segments. However, an optimisation program could be formulated to solve the same problem to find a global optimal solution.

In the next section, we formulate an optimisation program to solve the individual power levels that small cells should operate in order to ensure a minimum data rate to users with acceptable SINR. We will simplify the problem to mathematically model the problem in a LP. We will then discuss the results obtained from the heuristic algorithm and the LP, and summarise the findings in Section 3.7.

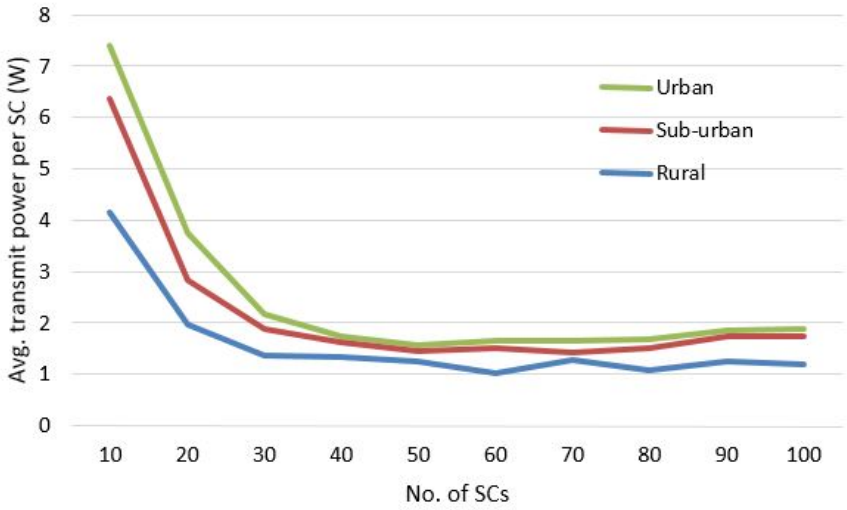


Figure 3.10: Average transmit power per small cell as a function of number of cells deployed.

3.6 Transmit Power Estimation for Small Cell Networks: Linear Program

In this section, we present our formulation of a linear program (LP) to obtain the optimal operating power of small cells while constrained by the power and capacity requirements. We calculate the optimal transmit power for the small cell antenna while guaranteeing a threshold data capacity for each user. We will now discuss the formulation of the LP along with the considered network structure and the assumptions made.

3.6.1 Problem Formulation

In the formulation of the LP, we focus on a segment of the wireless access network which includes one macro cell BS and multiple randomly distributed small cell BSs.

Objective Function

Our objective is to minimise the total transmitting power of the considered network segment. Let N be the set of small cells in the network segment. Then, the objective function can be expressed as:

$$minP_{MC} + \sum_{i \epsilon N} P_i \tag{3.2}$$

where P_{MC} is the power of the MC BS and P_i represents the power of the i^{th} small cell. Note that in this model, only the transmitting powers of the BSs are considered in the problem formulation. Other power consuming elements such as power amplifiers and digital signal processing are not considered. In this work, our focus is on the energy savings derived by optimising the transmit power.

Constraints

The following capacity and transmit power constraints are enforced on the model.

(a) Capacity demand constraint

We enforce a capacity constraint to ensure that the capacity demand requirements of the users connected to a specific BS are met. Let M_i be the set of users connected to the BS i, then the capacity constraint can be shown as:

$$\sum_{m \epsilon M_i} C_{m,i} \geq n_i C_{th} \quad \forall i \epsilon N \tag{3.3}$$

where $C_{m,i}$ denotes the capacity from the i^{th} small cell to the m^{th} user, C_{th} denotes the capacity demand threshold per user and n_i denotes the number of users connected to the i^{th} cell.

For an additive white Gaussian noise channel, the capacity of the channel is given by Shannon Hartley capacity formulae [191]. Shannons formula is shown in 3.4.

$$C = Blog_2(1 + SINR) \tag{3.4}$$

where C denotes the channel capacity, B denotes the bandwidth of the channel and $SINR$ is the signal to interference and noise ratio observed at the receiver. When SINR is much larger than 1, 3.4 could be further approximated to:

$$C = Blog_2(SINR) \tag{3.5}$$

In this work, we use 3.5 in order to represent channel capacities in terms of bandwidth and received SINRs. Let $n, o \epsilon N$ and $m \epsilon M$, then the SINR received at the m^{th} user is given by:

$$SINR_m = \frac{P_n \ G_{n,m}}{\sum_{o \epsilon N, o \neq n} P_o \ G_o + \sigma^2} \tag{3.6}$$

where, P_n is the transmitting power of the BS that the user m is connected to and P_o denotes transmitting power of any other surrounding BS. $G_{n,m}$ denotes the slow fading gain between the m^{th} user and the n^{th} BS. σ^2 represents the interference noise power.

The slow fading gain for the SINR calculation in Eq. 3.6, is modelled as in [192], and is given by:

$$G_{n,m} = K_0 \times 10^{\frac{\beta_{n,m}}{10}} \times d_{n,m}^{-\eta} \tag{3.7}$$

where, K_0 is a factor accounting for the effects of antenna gain and carrier frequency, $\beta_{n,m}$ is a Gaussian random variable with zero mean and standard deviation of 4 dB to account for log-normal shadowing effects, $d_{n,m}$ is the distance between the m^{th} user and the n^{th} BS, and η is the path loss exponent.

We calculate the SINR received at the user equipment to determine the cell that offers the best signal strength to the user. When doing so, the noise at the receiver is dependent on the transmitting powers of surrounding BSs making the optimisation problem hard to solve. Therefore, in order to simplify the constraint to a linear constraint, we assume that the surrounding BSs operate at an average transmitting power $(P_{sc,avg})$ in the noise calculation. This assumption is discussed later in the formulation. This modified constraint could be shown as:

$$P_i \geq 2^{\left[\frac{C_{th}}{B} - \frac{1}{N_i} log_2 \left[\frac{\Pi_{m \epsilon M_i} \ G_{i,m}}{\Pi_{m \epsilon M_i} \left[\Sigma_{j \epsilon N, j \neq i} \ P_{avg} \times G_{j,m} + \sigma^2 \right]} \right] \right]} \tag{3.8}$$

where N_i is the number of users connected to the i^{th} small cell. Note that the revised constraint will now also enforce a lower bound on the transmitting powers. If N_i becomes zero for a certain cell, as we consider the lower bound of transmitting power to be zero,

hence representing a cell which is turned off.

(b) Power constraint

Transmit power of the small cell antenna is bounded by a maximum threshold due to physical equipment limitations. Therefore, we enforce a maximum power constraint as shown below:

$$P_i \leqslant P_{sc,max} \tag{3.9}$$

where $P_{sc,max}$ is the maximum allowable transmit power for a small cell. Further, we consider the macro cell BS to operate at a constant transmit power level of $P_{mc,max}$ continuously.

Assumptions

In the problem formulation, since the noise at the receiver is dependent on the transmitting powers of surrounding BSs, the capacity constraint becomes a DC constraint (difference of convex functions) thus making the problem non-convex and complex to solve. Here, we assume that the surrounding BSs operate at an average transmitting power $(P_{sc,avg})$ in the noise calculation. The effect on this assumption is minimal because for a specific user, gains from the surrounding BSs except for the one which it is connected to are in the order of 10^{-6}, thus making the SINR value relatively insensitive to the transmitting powers of surrounding cells. The assumption allows us to simplify the capacity demand constraint into a linear constraint.

Further, we assume the small cells and users are randomly distributed in the area considered. Therefore, the coordinates of small cells as well as the users are derived from a uniform random variable sequence.

The main limitation of the LP approach is the computational complexity. As the modelled network size increases to consider multiple macro cell BSs, the optimisation problem becomes complex and hard to solve using optimisation methods. In particular, this is due to the noise calculation involving the increased set of neighbouring cells. However, the optimisation program approach is effective to model a smaller network segment to

Parameter	Values
Area	1 km x 1 km
Bandwidth (B)	$1.5x10^6$ Hz
Antenna gain and carrier frequency factor (K_0)	10^3
Path loss exponent (η)	4 (urban) 3 (sub-urban) 2 (rural)
Population density (inhabitants per km^2) [189]	>400 (urban) 150 - 400 (sub-urban) <150 (rural)
Interference noise power (σ^2)	10^{-10} W
$P_{sc,max}$, $P_{mc,max}$ [190]	6.3 W, 80 W
$P_{sc,avg}$	5W

Table 3.4: Parameters considered in the evaluation of the linear program.

obtain a more accurate and optimal solution. Therefore, the LP can be used in conjunction with the previously proposed heuristic-based algorithm to thoroughly investigate the energy-efficient transmission power control for heterogeneous wireless access networks.

3.6.2 Model Evaluation

In this section, we evaluate our model to obtain the optimal power levels considering a segment of the wireless access network. In order to maintain comparability with the heuristic-based approach, we consider a network segment which includes one macro cell BS surrounded by several randomly placed small cell BSs. Parameters considered in our evaluation are listed in Table 3.4.

Further, we derive small cell BS locations from a uniform random sequence with the condition that they are placed at least 100 m away from the macro cell BS and 50 m apart from each other. Users' locations are obtained from a uniform random sequence to satisfy different user densities to simulate dense and sparse populated scenarios. Each user is considered to be connected to the BS which records the highest SINR. The capacity demand threshold (C_{th}) is considered to vary with the time of the day as presented previously in Fig. 3.5 in Section 3.5.2.

3.6.3 Results and Discussion

In this section, we present the results obtained from the above formulated LP. Similar to the evaluation of previously discussed heuristic algorithm, first, we evaluate the LP under various user densities in an effort to model a CBD environment, where a high number of users come into the metropolitan area (to work during the daytime), and leave the area during the later hours of the day. Second, we classify three scenarios; urban, sub-urban, and rural environments to evaluate the solution for residential purposes. Further, we present results where the number of small cell BSs deployed in a unit area is varied in an effort to investigate the impact on the energy-efficiency from the deployed cell density.

Fig. 3.11 showed the total transmission power of the small cell BSs and how it varies as the guaranteed capacity threshold (C_{th}) varies with time and subjected to different user densities. The number of small cells is fixed at 20 in this instance. As the results show, the highest transmission power is recorded at the peak traffic hour (at 1600 hours) for any user density considered. When the number of users is increased, the transmission power increases, correspondingly increasing the small cell coverage area. However, it is observed that when the number of users is low (<400) the required transmission power tends to reduce considerably. This allows room for power savings in terms of transmission power in the network where the number of users varies, e.g., in a CBD where users go home after work.

We compare the results generated from the LP against the results from the previously proposed heuristic-based approach, presented in Section 3.5.3. Comparing the resulting total transmission power shown in Fig. 3.11 against that of the heuristic approach (Fig. 3.6 in Section 3.5.3), the results clearly show that the LP presents a more energy-efficient solution. For example, in the presented case where 20 small cell BSs are considered, at the peak traffic hour (1600 hours), when 800 users are connected to the network, the LP shows a total transmission power of 112 W where the heuristic-based approach shows a total transmission power of 140 W. The average gap observed between the two surface graphs in Fig. 3.11 and Fig. 3.6 is 14 W. That is, in an average instance, the LP will present a transmission power control solution that consumes 14 W less power than the heuristic-based algorithm generated solution.

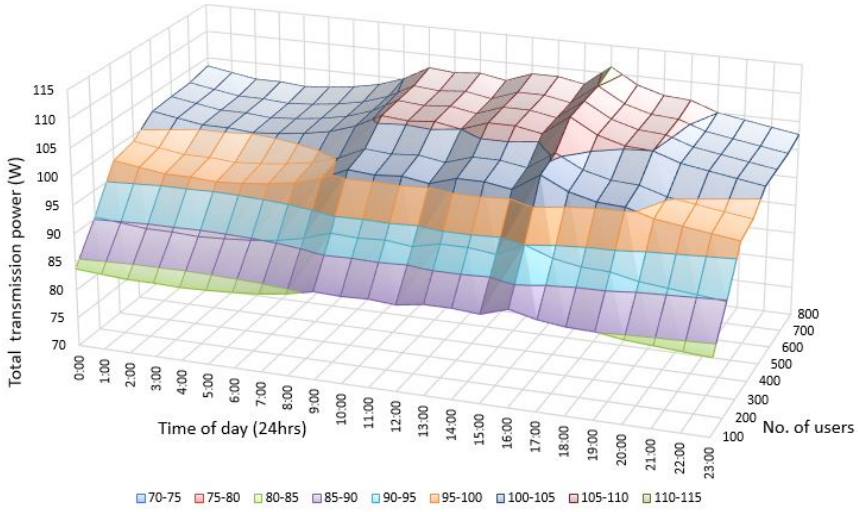


Figure 3.11: Total transmission power versus number of users under varying capacity demand throughout the day.

Fig. 3.12 shows the transmission power variations during peak traffic hour as a function of number of users and number of deployed small cells in the considered area. Results suggest that the total transmission power gradually increases when the deployed small cell density is increased. However, the transmission powers tend to be insensitive to the number of users above a certain value of approximately 400, because once the small cell coverage is wide enough to cover the entire area, the cells can cater for the increased capacity demand while operating at the same power levels. Therefore, in order to derive the best power savings, such a power control scheme could be most suited to an area with a wide variation in the number of users. Further, when the number of users is increased, the total transmission power increases linearly. This suggests a linear correlation between the transmission powers of the SCN and the traffic load of the network.

Comparing the results in Fig. 3.12 against the results derived from the heuristic-based algorithm (Fig. 3.7), we observe that there is a clear resemblance between the two graphs. However, the LP presents lower transmission power requirements for the small cell BSs compared to the heuristic approach irrespective of the number of small cell BSs, or the number of users considered. This is due to the near-optimal nature of the heuristic-based algorithm solution.

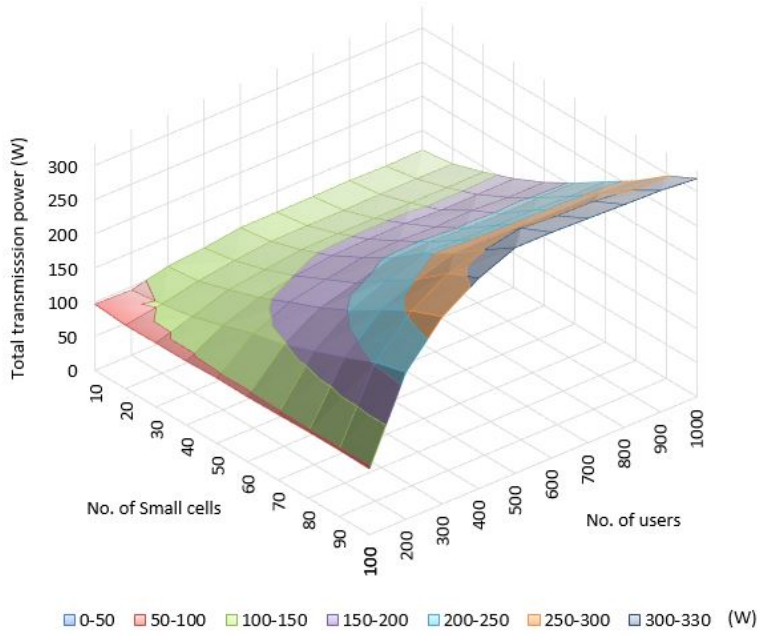


Figure 3.12: Total transmission power versus number of users and number of deployed small cells during peak traffic hour.

Next, we evaluate the LP in three scenarios, namely, urban, sub-urban and rural depending on the number of users connected to the network. Fig. 3.13 shows the total transmission power of small cells against the hourly capacity demand variations throughout the day. Here, to maintain comparability with the similar results from the heuristic (previously shown in Section 3.5.3, Fig. 3.8), we consider a network segment with 20 small cell BSs. Results indicate peak transmission power corresponding to the peak capacity demand at 1600 hours. Amongst the three scenarios, urban, and sub-urban environments indicate increased transmit power compared to the rural environment. This is due to more users being served by the small cells. The trend of the transmission power variation correlates to the required capacity thresholds enforced. Comparing the results against the results from the heuristic-based algorithm (in Fig. 3.8, we observe that when the rural environments are considered, both methods approach similar solutions. However, when the population density is increased (when considering the sub-urban and urban environments), the resulting total transmission power increment from the heuristic approach is larger than the increment observed in the LP. This is due to the sub-optimality

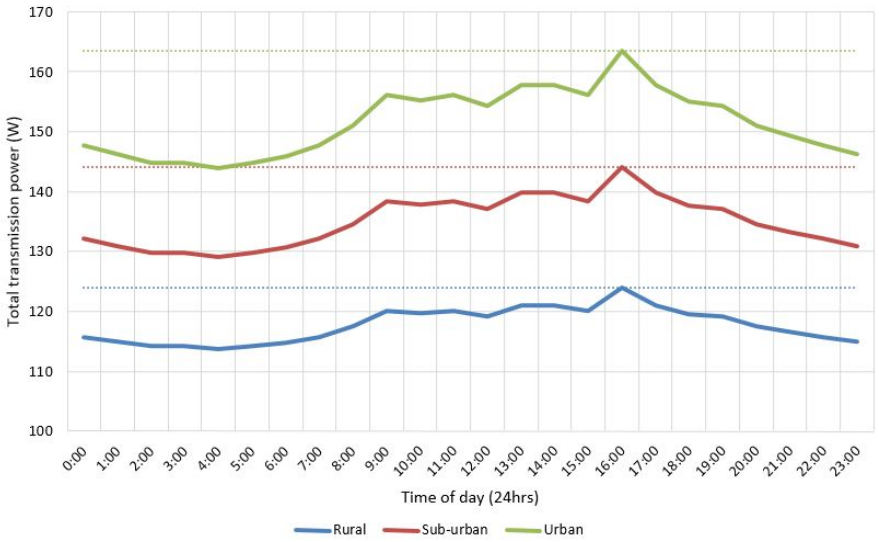


Figure 3.13: Total transmission power under urban, sub-urban, and rural cases during the day.

of the heuristic-based approach. However, in rural environments, the sub-optimality is shown to be insignificant.

The dotted horizontal lines in Fig. 3.13 indicate the peak transmit power in each scenario. We consider this power level as a baseline where the small cells operate continuously at this peak transmit power throughout the day. We evaluate the power savings of the considered heterogeneous network segment with respect to this baseline in Fig. 3.14.

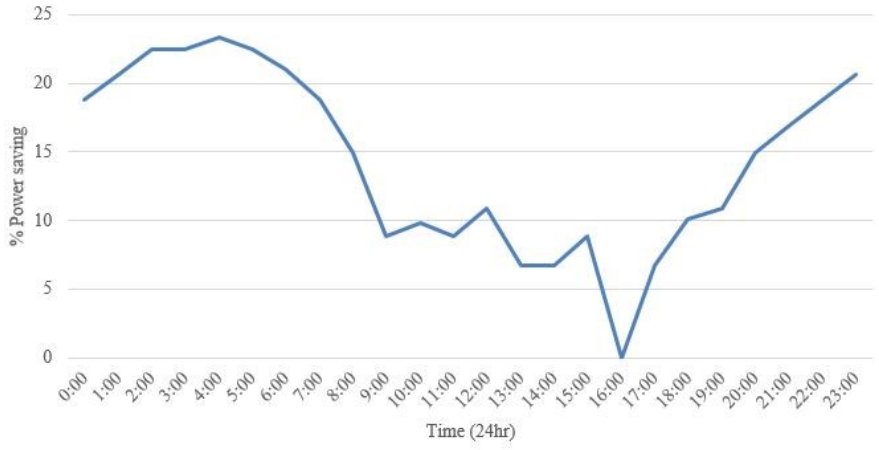


Figure 3.14: Percentage transmission power savings.

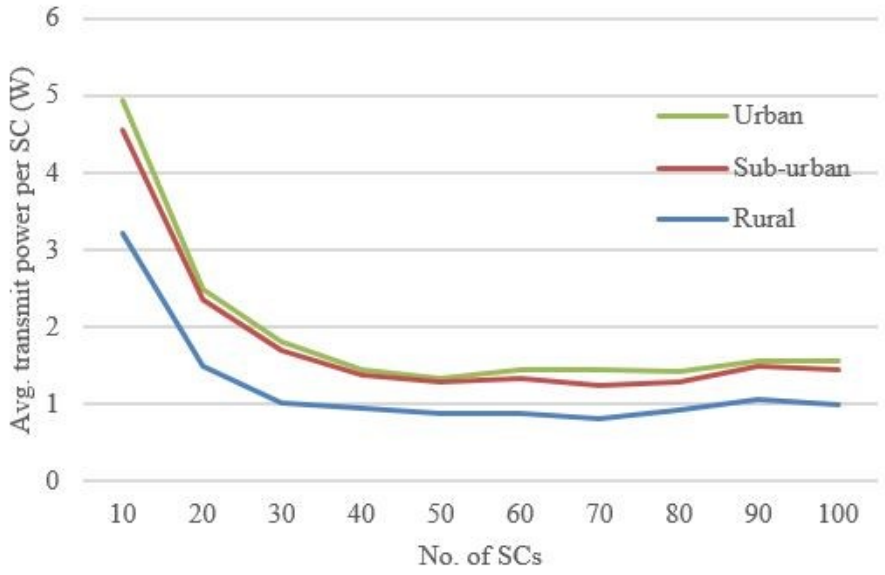


Figure 3.15: Average transmit power per small cell as a function of number of cells deployed.

The power savings are averaged over the three population density cases. Results show maximum savings during low traffic periods and minimum savings the peak traffic hour. The LP approach presents a 14% power saving on average throughout the day compared to the 12% power saving observed from the heuristic-based approach in Fig. 3.9.

Fig. 3.15 presents the average transmit power per small cell as a function of number of small cells deployed at the peak traffic hour. Results show that as the number of small cells increases, the average transmit power decreases due to users being shared amongst more cells and the increased freedom to choose a cell with higher signal strength. We observe that the average transmit power per small cell tends to saturate when the number of small cells increases due to the total transmission power being shared amongst many small cells. This trend suggests that over-densification of cells does not improve power savings though bandwidth capacity per user is increased. Further, we observe that the increase in population density will increase transmit powers. However, note that as population density is increased, the average power per small cell becomes less dependent on it, as observed by the small difference between the urban and sub-urban curves. We will now summarise the findings from the proposed transmission power control approaches

for the heterogeneous wireless access network presented in this chapter.

3.7 Summary

In this chapter, we presented two approaches for energy-efficient power control of heterogeneous networks. Our approaches mainly focused on adjusting the transmission power of the small cell BS antenna such that the overall power consumption of the heterogeneous network is minimised. In the first approach, we developed a heuristic-based transmit power control algorithm to estimate the optimal transmission power for the small cell BSs considering the traffic variations during a day. In the second approach, we presented the formulation of an optimisation problem to derive the optimal transmit power for the small cell BSs. The main difference between the two approaches is the scalability of the approach towards different network sizes. Due to the complexity of the optimisation problem, modelling large networks becomes computationally infeasible. However, the relatively straightforward heuristic-based algorithm could easily be scaled to model large network segments. The two methods could be used subsequently (use heuristic-based approach to model a large network first and then, use optimisation program to further model smaller segments) for a thorough analysis of traffic-based transmission power control of heterogeneous wireless access networks.

When the two approaches were simulated considering a single macro cell system, results showed opportunities to improve energy-efficiency of the wireless access network by considering transmission power control for the small cell BSs. In the heuristic approach, results presented a near-optimal solution for transmit powers of the small cell BSs. The total transmission power of the network was reduced by 12% on average during the 24-hour time span of a day whereas the LP showed a 14% reduction of total transmission power on average. As far as the optimality of the solutions is concerned, the LP presents a highly accurate and globally optimal solution, however, it has greater computational complexity. In the heuristic-based approach, even though the results are sub-optimal, due to the simplicity of the heuristic-based approach, it could model complicated network architectures with more flexibility. However, considering the results,

we can conclude that effective transmission power control which considers traffic demand variations of the network, can improve the energy-efficiency of the future wireless access networks.

In the next chapter, we take our discussion towards the core network segment. We propose a series of path protection schemes that exploit an architectural feature of the access network segment, referred to as the dual-homing capability. The next chapter is an effort to increase the core network survivability against probable failures in fibre links and network hardware, while identifying opportunities to increase the energy-efficiency of the optical core networks.

Chapter 4
Survivable Core Networks

This chapter focuses on the core network segment of next generation telecommunication networks. We propose a series of dedicated path protection schemes that exploits dual-homing architecture in the access network to improve the core network survivability against fibre link and/or network hardware failures. The chapter also proposes a power consumption model that estimates the core network power consumption followed by a detailed discussion of power saving opportunities from dual-homing based core network survivability.

4.1 Introduction

In this chapter we discuss the core segment of telecommunications networks and their survivability against possible failures in fibre links and networking hardware. The topic of survivable core networks has become a crucial area of study within the research community, especially due to large amounts of data that could potentially be failed to be delivered in an event such as network equipment failure. For the network service providers, even a small down time of the network could result in huge data loss, as well as financial liabilities that could arise from not being able to uphold the service level agreements (SLAs) made with the clients.

At present, with the emergence of bandwidth intensive Internet based services such as video-on-demand (VoD), internet protocol television (IPTV) and voice-over-IP (VoIP) services the network traffic over core networks has risen sharply. Forecasts published by Cisco visual networking index (VNI) on global traffic highlight that globally, Internet traffic will grow 3.2 fold from 2014 to 2019, at a compound annual growth rate (CAGR) of 26%. The same forecast also highlight that, the metro-only traffic will account for 66% of all IP traffic in 2019, up from 52% in 2014. The metro-only traffic reports a compound

annual growth rate of 29%, where the core traffic report a 14.4% of CAGR [35].

Optical wavelength division multiplexing (WDM) networks have been identified as a solution to cater for this booming traffic demand. WDM networks have been widely deployed especially due to their capacity to deliver high data rates, high capacities, and most importantly, long reach making it possible to deliver services over a large geographical area. However, delivering high data rates over the core network gives rise to other concerns such as effective resource utilisation and reliability of the data transmission. When dealing with high data rates, reliability becomes an important concern for the network operators. A failure of a link or a node which could translate into a network malfunction and even a small down time of a single fibre link could result in loss of a massive amount of data. That would ultimately translate into financial liabilities from a network operators' point of view. Therefore, ensuring uninterrupted service even in the event of failure has become increasingly important. Thus we focus our attention in this chapter towards network survivability.

In this chapter, we propose a series of routing and wavelength assignment (RWA) algorithms to provide dedicated path protection for data communication through the core networks while utilising core network resources effectively. These algorithms are designed to exploit the dual-homing capability of access networks to provide protection to the data transmission through the core segment of the network. In a dual-homed optical access network architecture, a local exchange (LE) is connected to two different metro/core (M/C) nodes, protecting the access network against feeder fibre failures. This concept is illustrated in Fig. 4.1. The end-to-end primary connection thorough the core network is protected by provisioning network resources to another end-to-end backup path between the source and destination LEs. Due to the dual-homing architecture in the access network, we have several options for the backup path as shown in Fig. 4.1. The RWA algorithms presented in this chapter solve the routing and wavelength assignment problem considering the dual-homing capabilities at the access network to improve core network survivability.

The remainder of the chapter is organised as follows. First, we discuss WDM networks in detail in Section 4.2. In particular, in this section, we comprehensively dis-

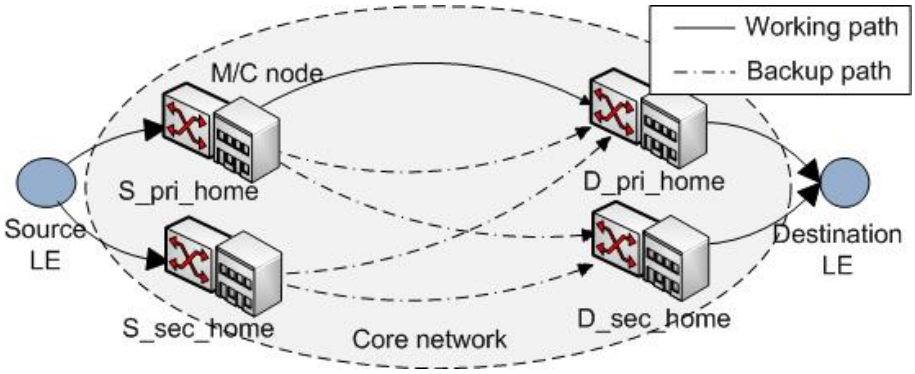


Figure 4.1: An example of dedicated path protection over a dual-homed network.

cuss WDM mesh networks focussing on the network architecture, technologies used, and commonly practised fault management techniques. The dual-homing concept and architecture is discussed next in Section 4.3. Then, we present a literature review of related work that focuses on protection based core network survivability in Section 4.4. Next, in Section 4.5, we propose a series of heuristics algorithms to solve the RWA problem in core networks where the dual-homing capability is available in the access segment, followed by a discussion of the simulation results gathered. Further, we analyse the impact to the energy-efficiency of the core network and present an analysis highlighting the energy trade-off of our proposed protection approach. Finally, Section 4.6 summarises the chapter.

4.2 Wavelength Division Multiplexing (WDM) Networks

Wavelength division multiplexing (WDM) is an approach that can drastically increase bandwidth capacity of a transmission channel by multiplexing multiple wavelengths onto a single fibre. In WDM, the optical transmission spectrum is separated into a number of non-overlapping distinct wavelength bands (or frequency bands), with each wavelength supporting a single communication channel operating at whatever rate one desires, e.g., peak electronic speed. Thus, by allowing multiple WDM channels to coexist on a single fibre, one can tap into a huge fibre bandwidth, with corresponding challenges being the design and development of appropriate network architectures, protocols, and algorithms.

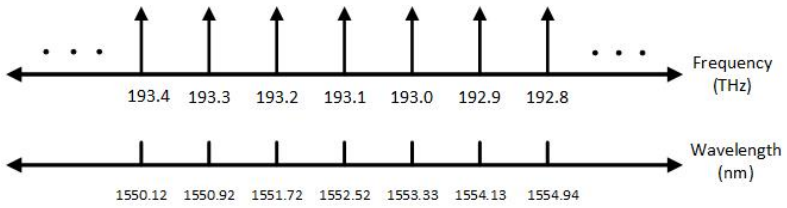


Figure 4.2: The ITU wavelength grid with 100 GHz channel spacing [7].

Research and development on optical WDM networks have matured considerably over the past decade. These are networks widely deployed by telecom network operators all over the world. Currently, the Internet employs WDM-based optical backbone. In such a network, end users connect to the network through a wavelength sensitive switching/routing node (the details of which will be discussed later in the chapter). An end user in this context need not necessarily be a terminal equipment, but the aggregate activity from a collection of terminals including those that may possibly be aggregating traffic from other regional and/or local sub-networks. Therefore, the end user's aggregate activity on any of its transmitters is close to the peak electronic transmission rate.

4.2.1 ITU Wavelength Grid

There is a strong requirement for the standardisation of WDM systems so that WDM components and equipments from different vendors can inter-operate with one another. Thus, industry standards for wavelengths have been developed under the leadership of International Telecommunications Union (ITU) [7]. A standard set of wavelengths, called the ITU grid, has been defined to coincide with the 1550 nm low-loss region of the fibre. Specifically, this grid is anchored at a frequency of 193.1 THz (which corresponds to a wavelength of 1552.52 nm). There is a 100 GHz grid, which means that spacing between adjacent channels is 100 GHz, which corresponds approximately to 0.8 nm wavelength channel spacing around the anchor frequency. A few channels of this grid around the anchor channel are shown in Fig. 4.2.

For denser packing of channels, a 50 GHz grid has also been defined around the same reference frequency of 193.1 THz [7]. The 50 GHz grid is obtained by adding a channel exactly half way between two adjacent channels of the 100 GHz grid. Continuing this

process, a 25 GHz grid can also be defined, and it can support up to 600 wavelengths [7].

4.2.2 The Evolution of WDM Networks

The first-generation of WDM networks provides only the point-to-point physical links, which are either static or manually configured. Fig. 4.3 presents a four-channel point-to-point WDM transmission system with amplifiers. The technical issues of the first generation WDM networks include design and development of WDM lasers and optical amplifiers [193].

The second-generation of WDM is capable of establishing connection-oriented end-to-end lightpaths in the optical layer by introducing optical add/drop elements (WADM or OADM) and optical crossconnects (OXC). The ring and mesh topologies can be implemented using these OADMs and OXCs. The lightpaths are operated and managed based on a virtual topology over the physical fibre topology, and the virtual topology can be reconfigured dynamically in response to traffic changes. The technical issues of second generation WDM includes the development of OADM and OXC, wavelength conversion, routing and wavelength assignment (RWA), interoperability amongst WDM networks, network control and management and so on.

The third-generation of WDM supports connectionless optical networks. The key issues include the development of optical access network such as passive optical network (PON), and optical switching technologies, referred to as optical "X" switching (OXS),

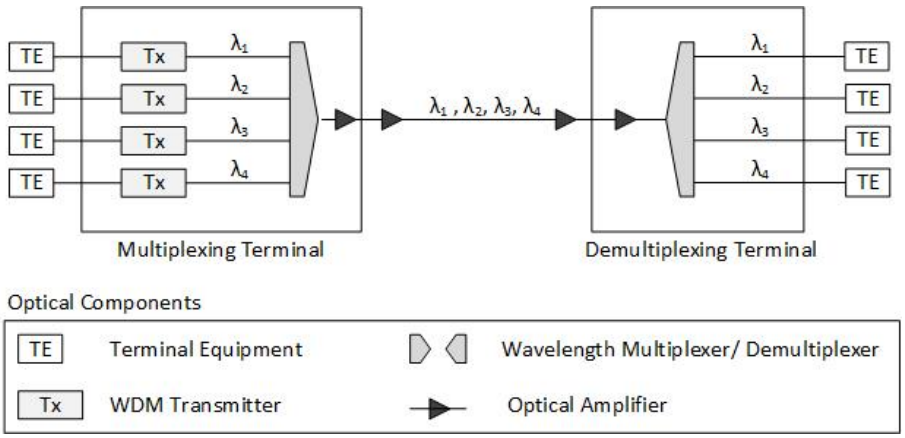


Figure 4.3: A four-channel point-to-point WDM transmission system [8].

where X can represent P (for packet), B (for burst), L (for label), F (for flow), C (for cluster or circuit), etc.

There are mainly three switching schemes for WDM networks: wavelength routing (or switching), packet switching, and burst switching. Wavelength routing (WR) is quite popular in today's WDM networks. However, as the data traffic requirements continue to grow, WR by itself might not the best WDM strategy. A long-term strategy for the network evolution is explored in optical packet switching (OPS) and optical burst switching (OBS) technologies.

The main idea of OPS is the migration of certain switching functionalities from electronics to optics to remove the incumbent layers that impose unnecessary optical-electrical-optical (O-E-O) conversions and unnecessary signal processing. Incoming packets are switched all-optically without being converted to electrical signals. This technology has been considered as the most flexible and also the most demanding switching scheme [8]. One of the main challenges for OPS is that there is no practical and cost-effective optical equivalent of the random access memory (RAM) and logic devices for optical signal processing. Optical burst switching has been proposed as a compromise between optical circuit switching and optical packet switching, while avoiding their shortcomings [194].

The latest WDM networks combine the most advanced optical transmission technologies with the switching of lightpaths and creating packets of information into a multifunctional packet-optical transport network. Its principal elements are the (a) transponders and muxponders that allow traffic to enter and leave the optical network, and (b) the optical filters, multiplexers/demultiplexers and reconfigurable optical add drop multiplexers that multiplex and send wavelengths of light in different directions as directed by the controlling management system. Fig. 4.4 shows a typical WDM optical network covering the access network to the long haul.

A WDM optical network provides circuit-switched end-to-end optical channels or lightpaths between network nodes and their users, the clients. A lightpath is made up of a wavelength between two network nodes that can be routed through multiple intermediate nodes. As we progress our discussion towards the core network survivability, it is important to discuss lightpath communication, WDM mesh network architecture, and

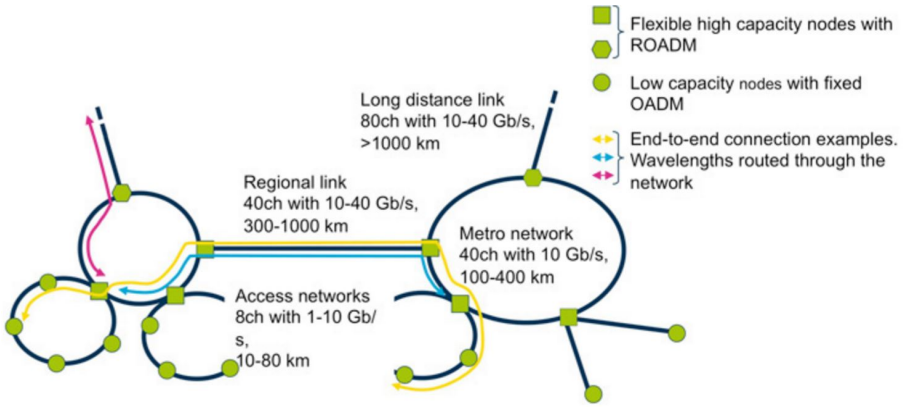


Figure 4.4: A typical WDM optical network covering access to long haul.

the routing and wavelength assignment. These topics are discussed in the next subsections.

4.2.3 Lightpaths

A *lightpath* is the basic communication approach in a wavelength-routed network [195]. Lightpaths in a WDM network are end-to-end connections, and should be considered as the equivalents of uninterrupted *wires*, stretching from one point in the network to another while passing one or several nodes. This is a significant difference from the principles of classical TDM optical transport networks, such as SDH and SONET, where the signals are regenerated at each node. The equivalent uninterrupted *wire* stretches only between two nodes. Hence, a WDM network requires careful wavelength planning to define where each wavelength starts and ends, in comparison, a SDH/SONET network makes all signals available in every node passed. The end-to-end aspect also affects how the power budget (i.e. signal attenuation) is calculated. In a WDM network, the optical transmission characteristics for a wavelength has to be calculated for the complete distance the light path traverses. In a SDH/SONET network, a new power budget is calculated for each hop between two adjacent nodes.

Figures 4.5 (a) and (b) show the layout of a SDH/SONET node and WDM node, respectively. In the SDH/SONET node (Fig. 4.5 (a)), all traffic signals are regenerated and switched, making them available for add and drop. In the WDM node (Fig. 4.5 (b)), only

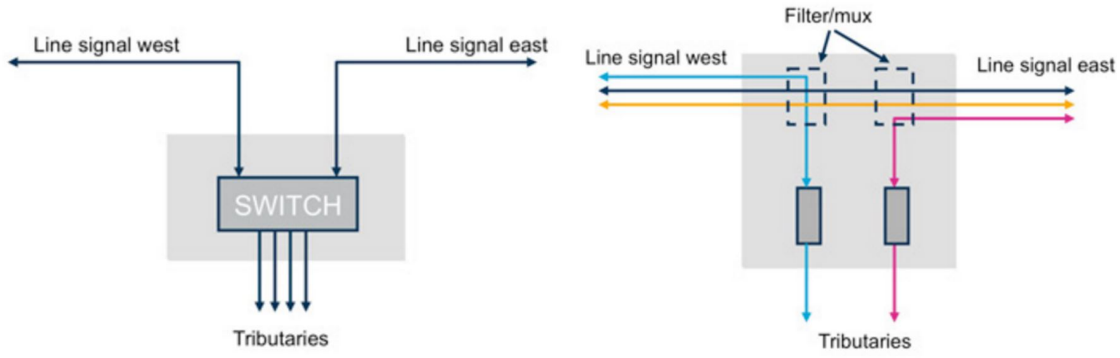


Figure 4.5: (a) SDH/SONET node (b) WDM node (adapted from [8]).

selected signals (wavelengths) are available for add and drop, the rest are passed through (*glassed through*) without change.

The lightpaths of the WDM optical network have several important characteristics.

- The lightpaths are transparent, i.e. they can carry data at various rates, with different protocols etc. This enables the optical layer to support a variety of higher layer protocols concurrently.
- Wavelength and data rate used are set by the terminating nodes. Hence an individual lightpath may be updated to higher capacity by simply changing traffic units in the start and end nodes, without affecting any equipment in the intermediate nodes. This is a fundamental difference to SDH/SONET networks as well as networks of interconnected Ethernet switches.
- Lightpaths can be set up and taken down on demand, equivalent to the establishment of circuits in a circuit switched network.
- Alternative lightpaths can be configured and kept in *standby mode* so that in the event of a failure, traffic may be re-routed and the service maintained.
- Wavelengths can be reused. If a lightpath which uses a particular wavelength ends in one node, the same wavelength can be reused in another lightpath heading in another direction.

- The concept of WDM and lightpaths is based on analog optical transmission techniques, making parameters such as dispersion, signal attenuation, optical signal to noise ratio and interference over the whole length of the path able to controlled as required to optimise data transmission.

A lightpath could be unidirectional or bidirectional. Note that in the context of this thesis, we have considered lightpaths to be unidirectional.

4.2.4 WDM Mesh Networks

Compared to traditional SONET-based networks [196] which usually takes the form of rings or interconnected rings, OXC-based wavelength routed networks can have a general mesh topology. A wavelength routed optical WDM mesh network is shown in Fig. 4.6. The network consists of ten OXCs connected by fibre links to form an arbitrary mesh topology. Each access station is connected to an OXC via a fibre link. Further, each access station is equipped with a set of transmitters and receivers, both of which may be wavelength tunable.

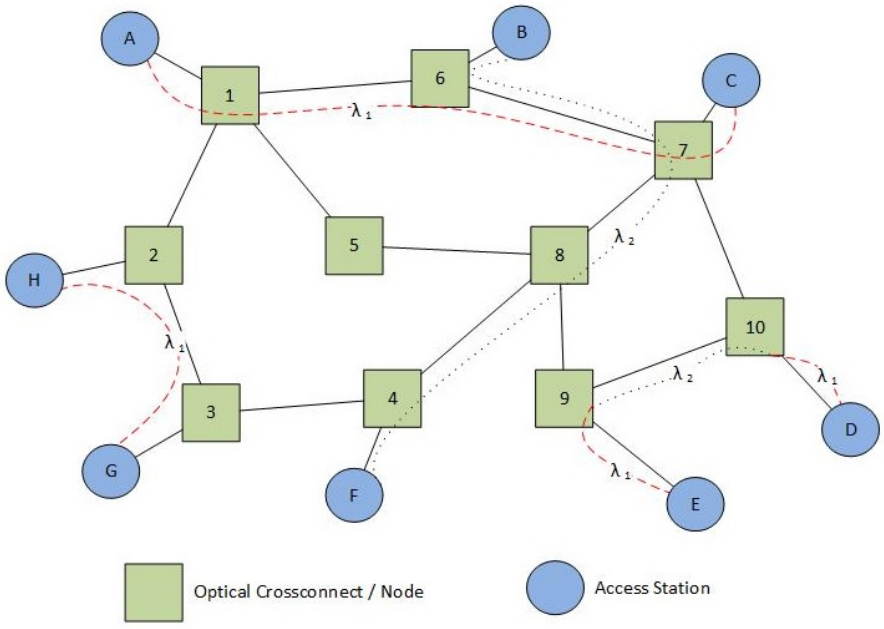


Figure 4.6: A wavelength routed WDM mesh network.

As discussed in the previous subsection, the basic mechanism of communication in a wavelength-routed network is a lightpath. A lightpath is an all-optical communication channel between two nodes in the network, and it may span over more than one physical fibre link. The intermediate nodes in the fibre path route the lightpath in the optical domain using their switches. The source and destination nodes (or the end nodes) of a lightpath access the lightpath with transmitters and receivers, respectively. If the transmitters/receivers are tunable, they must be tuned to the wavelength on which the lightpath operates. For example, in Fig. 4.6, lightpaths are established between nodes A and C on wavelength channel λ_1, between B and F on wavelength λ_2, and between H and G on wavelength channel λ_1. The lightpath between access stations A and C is routed via switches 1, 6, and 7.

In the absence of any wavelength conversion device, a lightpath is required to be on the same wavelength channel throughout its path in the network. This requirement is referred to as the *wavelength continuity* property of the lightpath. This requirement may not be necessary if we also have wavelength converters in the network. For example, in Fig. 4.6, the lightpath between the access stations D and E traverses the fiber link from access station D to switch 10 on wavelength channel λ_1, gets converted to wavelength channel λ_2 at switch 10, traverses the fibre link between switch 10 and switch 9 on wavelength channel λ_2, gets converted back to wavelength λ_1 at switch 9, and traverses the fibre link from switch 9 to access station E on wavelength channel λ_1.

The problem of routing the lightpaths in a WDM network is known as the routing and wavelength assignment (RWA) problem. A fundamental requirement in a wavelength-routed optical network is that two or more lightpaths traversing the same fibre link must be on different wavelength channels so that they do not interfere with one another. Therefore, routing and wavelength assignment (RWA) methods are essential for the operation of WDM mesh networks. In the next sub-section, we will discuss the RWA problem.

A. Routing and Wavelength Assignment

There are two main variants of the routing and wavelength assignment problem depending on the considered lightpath request method. If the lightpath requests are known in

advance, the RWA problem is referred to as the static RWA problem, or the static lightpath establishment (SLE) problem. If the lightpath requests occur dynamically, that variant is called the dynamic RWA. Our main aims in this chapter are to first, investigate the survivability improvements made possible by dual-homed access network architecture, and then, conduct an availability analysis to compare the improved availability of the connections. Our analysis focuses on a particular instant of the network where a prior known connection matrix is solved for a primary path and a backup path, protecting them over the core network. Therefore, in this chapter, we focus on the static routing and wavelength assignment problem. The static RWA problem, is thus directly referred to as the routing and wavelength assignment problem hereafter in this chapter. Formally, the RWA problem can be stated as follows. Given a set of lightpath requests that need to be established on the network, and given a constraint on the number of wavelengths, the routes over which these lightpaths should be set up and also the wavelengths which should be assigned to these lightpaths need to be determined so that the number of established lightpaths is maximised. The RWA problem has been extensively studied in literature [195, 197–200].

In this chapter, we propose a series of RWA algorithms to improve core network survivability against several types of failures. Our aim is to exploit the dual-homing architecture in the access network to find optimal primary and protection paths. In order to compare the performance of the RWA algorithms, we conduct an availability analysis using the simulation results. Next, we will discuss the availability parameter of a end-to-end optical connection and how it is calculated in our analysis.

B. Availability Analysis in WDM Mesh Networks

The availability of a system (which could be a component, path, connection, etc.) is the fraction of time the system is *up* during the entire service time. The availability of a system in a mesh network is generally analysed with the following assumptions:

- A system is always either available (functional) or unavailable (experiencing failure).
- Failure of two network components are independent events.

- For any component, the *up* time (or Time To Failure) and the *down* time (or Time To Repair) are independent exponentially distributed stochastic variables with known mean values. Therefore, we assume that mean time to failure (MTTF) and mean time to repair (MTTR) are known.

If a connection t is carried by a single path, its availability (denoted by A_t) is equal to the path availability. However, if t is dedicated or shared protected, A_t will be determined by both the primary and the backup paths. Here, note that the contribution of the reconfiguration time for switching traffic from the primary path to backup path (including signal propagation delay of control signals, processing time of control messages, and switching time at each node) towards unavailability is disregarded since it is relatively small, usually on the order of milliseconds, compared to the failure-repair time (usually on the order of hours).

Availability of a Dedicated Path Protected Connection

In path protection, connection t is carried by one primary path p and protected by one backup path b. The paths, p and b are link disjoint[1]. When a failure occurs in p, traffic may be switched to backup path b (if 1:1) or b may already be carrying the same traffic, and the receiver will switch its reception from path p to path b (if 1+1), as long as b is available. Otherwise, the connection becomes unavailable until the failed component is replaced or restored. t is *down* only when both paths p and b are unavailable. Availability of the connection t (denoted as A_t) can be calculated as:

$$A_t = 1 - (1 - A_p) \times (1 - A_b) = A_p + (1 - A_p) \times A_b \tag{4.1}$$

where, A_p and A_b denoted the availabilities of paths p and b, respectively. Note that a connection may employ multiple backup paths to increase its availability. Assuming that all backup paths are disjoint, the availability of a connection with multiple backup paths can be derived following the similar principle in Eq. 4.1.

[1] By link disjoint, we mean that the backup path for a connection has no links in common with the primary path for that connection. Node failures can also be accommodated by making the primary and the backup paths node disjoint as well.

End-to-End Path Availability

Given the route of path i, the availability of i (denoted as A_i) can be calculated based on the known availabilities of the networks components along the route. Path i is available only when all the network components along the path are available. Let a_j denote the availability of network component j. Let G_i denote the set of network components used by the path i. Then, A_i can be computed as:

$$A_i = \prod_{j \epsilon G_i} a_j \tag{4.2}$$

Network Component Availability

The availability of a network component can be estimated based on its failure characteristics. For a network component j, the availability (denoted as a_j) can be calculated as:

$$a_j = \frac{MTTF}{MTTF + MTTR} \tag{4.3}$$

In equation 4.3, the parameters $MTTF$ and $MTTR$ stand for the mean time to failure, and mean time to repair, respectively. These parameters can be derived from the historical failure statistics of the components.

4.2.5 Fault Management in WDM Mesh Networks

In a wavelength-routed WDM network, the failure of a network element, such as optical-layer hardware, fibre link, crossconnect, *etc.* can cause the failure of several lightpaths, thereby leading to a large data and hence, revenue loss. Fault recovery mechanisms can be classified into protection and restoration schemes, depending on there recovery approach.

Protection vs. Restoration

If backup network resources (routers and wavelengths) are pre-computed and reserved prior to the occurrence of a failure, such schemes are referred to as *protection* schemes [201–206]. Protection schemes, have a faster recovery time and can guarantee 100 percent recovery from the failure scenario for which they were designed, but require more energy in absolute terms. However, if another route and a free wavelength has to be discovered dynamically for each disrupted connection after the detection of the failure, such schemes are referred to as *restoration* schemes [207,208]. Generally, dynamic restoration schemes are more efficient in utilising network capacity because they do not allocate spare capacity in advance, and they provide resilience against different kinds of failures (including multiple failures). However, restoration schemes have longer recovery times and does not provide a 100 percent recovery guarantee against failure.

Protection schemes have faster recovery time and they can guarantee recovery from disrupted services they are designed to protect against. On the contrary, restoration schemes can not provide this guarantee. In the context of our work, we mainly discuss protection schemes for WDM mesh network fault management due to the above reasons.

Protection schemes can be classified as ring protection and mesh protection. Both ring protection and mesh protection can be further divided into two groups *path protection* and *link protection*. In path protection, the traffic is rerouted through a backup path (also known as backup route or protection path) once a link failure occurs on its working path (also known as the primary path). The primary and backup paths for a connection must be link-disjoint so that no single link failure can affect both of these paths. In link protection, that traffic is rerouted only around the failed link. When comparing link protection against path protection, link protection provides faster protection-switching times, however, path protection is more efficient than line protection. Path protection leads to efficient utilisation of backup resources and lower end-to-end propagation delay for the recovered route [8]. As such, for the context of the work proposed in this chapter, the main focus is on path protection schemes.

Dedicated vs. Shared Protection

Path protection schemes can be further classified into two classes: *dedicated path protection* and *shared path protection*. The two classes are briefly discussed below.

A. Dedicated Path Protection (DPP)

In dedicated path protection, spare resources are specifically allocated for a particular end-to-end connection. When a connection is brought down by a failure, it is guaranteed that there will be available resources to recover from the failure. Dedicated path protection schemes do not allow sharing of backup resources amongst multiple backup paths. Dedicated protection schemes can be implemented in a 1+1 or 1:1 fashion.

In *1+1 dedicated path protection*, the backup path is *active*, i.e., there are two live connections between the source and destination nodes. The destination node is equipped with decision circuitry to select the better of the two paths.

In contrast, in *1:1 dedicated path protection*, the backup path does not become active until after a failure has occurred on the primary path. After the failure is repaired, the connection may return to the primary path, which is called the revertive mode, or may remain on the backup path (non-revertive mode).

There are several advantages to operating dedicated path protection in a 1+1 mode. First, recovery from a failure can be almost immediate. As soon as the receiver detects that the primary path has become unsatisfactory, it can switch over to using the backup path. There is usually a small synchronisation delay due to the transmission latency of the two paths being different. However, the 1:1 mode is slower, as the failure must first be detected by the destination first. Then the source must be notified of the failure so that it can begin to transmit over the backup path. Another advantage to 1+1 is that failures on the backup path can be detected when they occur. With 1:1, a *silent failure* can occur on the backup path, such that the failure is not detected until the backup path is actually needed. One possible disadvantage to 1+1 dedicated protection is that it may require more equipment at the source and destination, to support two active paths.

The downside of dedicated protection, whether 1+1 or 1:1, is the large amount of spare capacity that it generally requires. In typical networks, the ratio of the dedicated

backup capacity to the working capacity is often on the order of 2 to 1 [209].

B. Shared Path Protection

Shared path protection addresses the inefficiency in requiring a large amount of spare capacity by allowing spare capacity to be used as protection resources for multiple working paths. The working path that share protection capacity should have no links or intermediate nodes in common so that a single network failures does not affect more than one of the paths.

Whilest sharing protection resources improves the capacity efficiency, one drawback is that contention for the resources may arise when where are multiple concurrent failures. Only one path can use the shared resources at a time, such that the other paths sharing the resources are vulnerable if a second failure occurs. Shared protection also requires greater coordination in the network so that the working paths are aware of whether the shared protection capacity is available or not. Note that shared protection usually operates in a revertive mode, such that the protection resources are released by the connection after the failures is repaired.

Shared protection is often referred to as 1:N protection, indicating there is one protection element for every N working elements (or more generally M:N protection). As N increases, the protection efficiency increases. However, the vulnerability of the scheme to multiple failures also increases. One study on shared path protection indicated that limiting N to about five provides significant capacity savings without leaving the network too vulnerable to multiple failures [210]. Ultimately, however, for a given network failure rate, it is the required availability of a connection that determines whether shared protection is suitable, and if so, what level of shearing is acceptable.

The different approaches taken to survive failures in WDM mesh networks (as discussed above) are illustrated in Fig. 4.7. As mentioned above, shared path protection cannot guarantee protection in an instance of multiple fibre failures due to the shared backup resources. However, dedicated path protection schemes (either 1+1 or 1:1) will guarantee the recovery of each end-to-end connection. Due to this reason, we consider the dedicated path protection approach when designing the dual-homing based dedi-

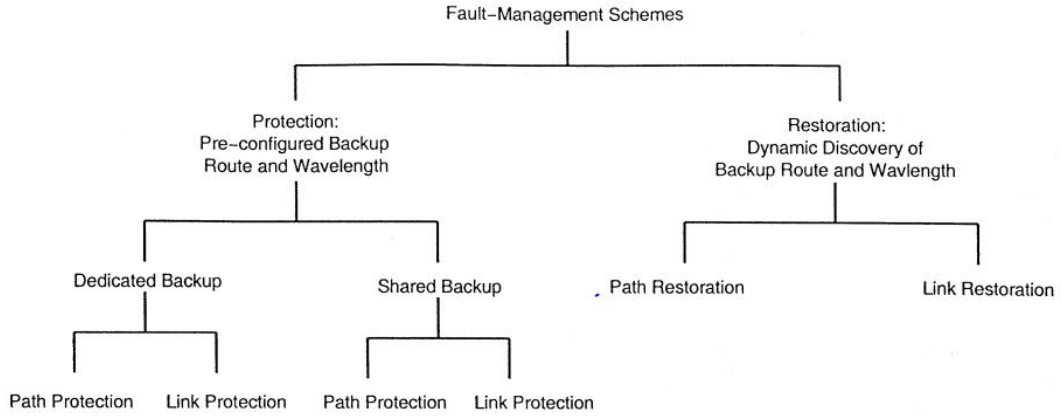


Figure 4.7: Different schemes for surviving failures in WDM mesh networks [9].

cated path protection algorithms later in this chapter. Further, the proposed algorithms exploit the dual-homing architecture of the access network to provide protection for the core network data communications. Thus, the dual-homing architecture is discussed in the next section.

4.3 Dual-Homing (DH) Architecture

Our aim is to exploit the dual-homing based architecture at the access network by proposing a series of routing and wavelength assignment solutions to improve core network survivability. We discuss the dual-homing architecture in detail in this sub-section. As the name implies, dual-homing (DH) is when a local exchange (LE), which aggregates end-user traffic from the access network, is connected to two nodes of the WDM mesh network. Initially, the aim of dual-homing was to ensure survivable connectivity between the access network and the metro/core network [211]. In a dual-homed optical access network, a LE is connected to two different metro/core (M/C) nodes, protecting the access segment against feeder fibre and/or M/C node failures. This is for example the architectural option envisaged by the FP7 project DISCUS [212], and also the network architecture considered in our work.

Fig. 4.8 presents an example of this concept, i.e., two dual-homed LEs and a core network comprising a number of M/C nodes interconnected by an optical circuit-switched

wavelength layer. Typically, the M/C node closest to the LE is referred to as the primary home for the LE in Fig. 4.8) and it is used during default operating conditions, while the second closest M/C node is chosen as the secondary home and it is used at the event of failure at the primary home.

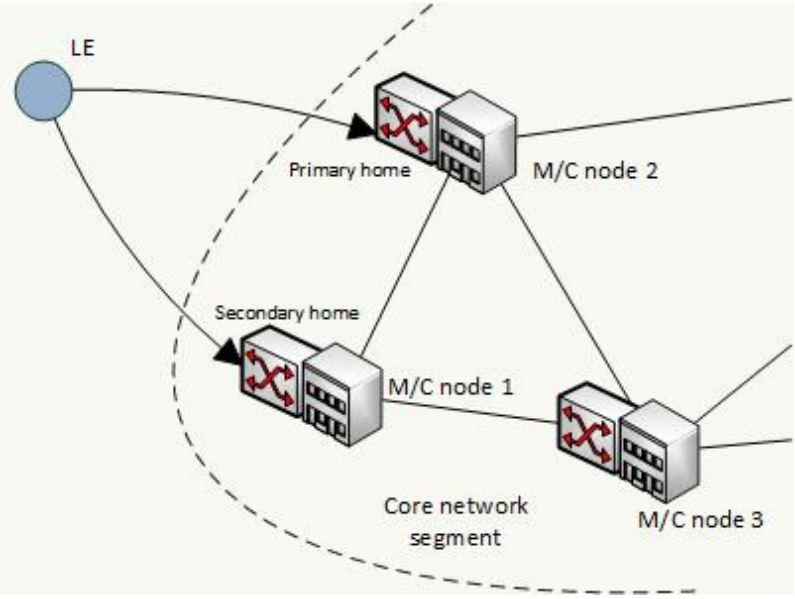


Figure 4.8: Dual-homed architecture at the access segment.

The dual-homing architecture has been widely studied in self-healing ring networks [211, 213, 214]. Some research has been conducted to provide protection for WDM mesh network architectures. We will discuss these studies in the *Related Work* section (Section 4.4). A measurement based analysis for a performance of multi-homing solutions is given in [215]. Further, dual-homing applications in wireless networks are also reported in literature [216].

Dual-Homing for Core Network Survivability

When a source/destination LE is attached to two M/C nodes, there exists two paths from the source LE to the destination LE. If these two paths are disjoint, they can provide 100 percent protection for the destination LE against a single link failure. Fig. 4.9 shows an example protected end-to-end connection between two dual homed LEs. Also, the

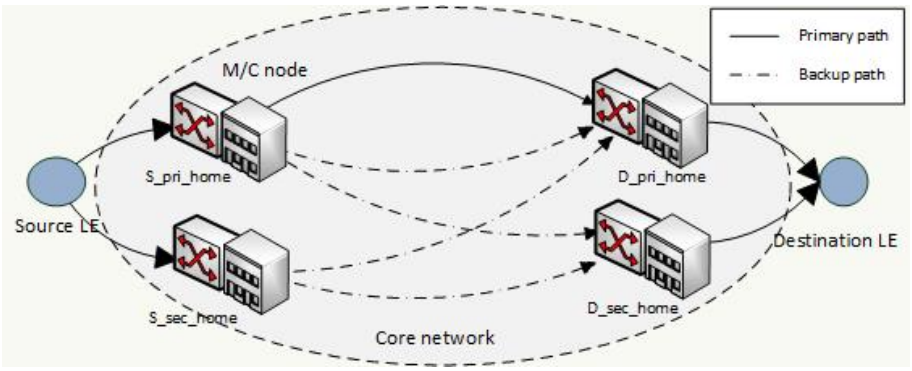


Figure 4.9: An example protected end-to-end connection showing different possible working/backup path options through the metro/core network.

primary[2] and backup path options are also presented in Fig. 4.9.

In this way, dual homing can be beneficial not only to protect from failures which affect the feeder fibre and primary metro/core (M/C) node, but also to provide more resource efficient resiliency against failures in the core part of the network. In the next section, we briefly discuss the considered design of the M/C node for the work proposed in this chapter.

Metro/Core (M/C) Node Design

In this section, the design specifications of the M/C node are presented. The design is initially proposed by the DISCUS FP7 project in [217] and a more detailed report on the architecture is later published in [10]. Fig. 4.10 shows the M/C node architecture. The main principle of the node is to have a transparent optical layer in the form of an optical switch that fibre links towards both access and core networks segments and electronic layers, i.e. Layer 2/Layer 3 switches could flexibly connect to. The involved optical switch does not necessarily distinguish between access ports and core connection ports and also enables direct connection of optical paths between the access and core segments. It could simplify the installation and operation, i.e. freely connecting the optical switch to the interfaces of access/core segments. Further, Fig. 4.10 also shows Layer 2 and Layer 3 routers, i.e. Ethernet layer and IP layer respectively, as well as a block of functions called

[2]We have considered the primary path to traverse through the primary homes for the operation simplicity.

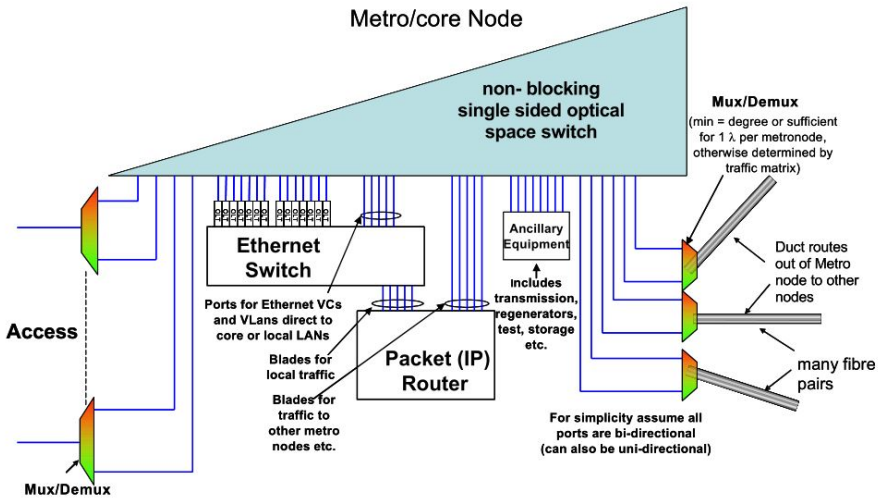


Figure 4.10: The Metro/Core node architecture proposed by DISCUS FP7 project [10].

ancillary equipment, which could be wavelength conversion, regeneration, testing and diagnostics, etc. as proposed in [10].

In the context of the our work in this chapter, we consider the DISCUS Metro/Core node as the network node in the wavelength-routed WDM mesh architecture, and is referred to as the M/C node for the remainder of this thesis. Our aim is to propose dual homing based routing and wavelength assignment solutions to improve core network survivability. In the next section, we discuss the related work published on this topic.

4.4 Related Work

In this section, we discuss the related work in the literature that studies survivability techniques for the WDM mesh networks.

In [207, 218], the authors have compared different approaches to protect WDM mesh networks against single link failures. In [218], three schemes, namely the dedicated path protection, shared path protection, and shared link protection were examined. In their work, the authors have assumed a static traffic demand and no wavelength conversion. The study compared the wavelength capacity requirements (the sum of the number of wavelengths required on each link) of the three approaches for 100 percent restoration.

The results showed that shared path protection provides the most efficient capacity utilisation over the other two methods. For example, in a 15 node mesh network, their results obtained by solving integer linear programs, showed that 59 wavelength links are sufficient if no protection is required for a 25-connection demand. The number of wavelength links required in dedicated path, shared path, and shared link protection schemes in the same example were 163, 99, and 189, respectively. Many studies have been conducted on shared path protection and restoration schemes for optical mesh networks [219–222]. In [219, 220], the authors propose efficient path selection algorithms for restoration of connections over shared bandwidth in a optical mesh network system. In [221, 222], the authors present a framework for service-guaranteed shared protection in WDM mesh networks. However, in the context of our work, our consideration is on the dedicated path protection.

In [223–225], the authors examined 1+1 protection and other WDM architectures with optical protection. In [226], the authors proposed an analytical approach to estimate the maximum capacity utilisation in optical networks that are resilient against single link failures. In [227], the authors have reviewed various protection and restoration techniques considering an IP-over-WDM network. In another study [203], the authors have presented efficient routing algorithms for computing the primary and backup routes in a WDM optical network. However, these studies do not consider the dual homing architecture in the access segment. Next, we will discuss some related studies where the dual homing capabilities have being considered to provide core network survivability.

There have been several efforts on providing protection over dual-homed networks in literature [228–231]. In [228], the authors integrated dual-homing and core network protection to handle a single link failure by considering a dual-homed IP-over-WDM architecture. In [229], Vokkarane *et al.* proposed a coordinated protection scheme for dual-homed-based IP-over-WDM networks in which users (i.e., enterprises) connect to IP routers of different service providers. The approach considered the presence of dual homing only at the source node of each communication request. The aim was to establish a working and a backup path between any of the two homes at the source side and the single home at the destination side, with the objective of minimising the total (i.e.,

working plus backup) path length. In [230], the authors have extended their work in [229], to consider dual-homing at both source and destination LEs to provide two pairs of working and backup paths to each of source dual homes to destination dual homes. The authors considered a generalised failure scenario in which independent failures may occur in the source access network, the destination access network, and the optical core network. Their work presented a wavelength cost analysis for the proposed coordinated solution. In [231], a scalable scheme for partial multicast protection based on a dual-homing architecture is presented. Their work investigated the use of dual homing to protect against one-to-many communications without making any changes to the routing algorithm in the core network. In [230] and [231], the authors assume full-wavelength conversion capability at each core node.

In our work, we investigate the benefits of utilising dual-homing at both the source and the destination LE while protecting demands against different types of failure scenarios in the core network segment. Our work differs from previous dual-homing-based survivable approaches in that it focuses on establishing a *single path pair* (i.e., one working path and one backup path) from/to either home (primary *or* secondary) of the source and destination LEs. And also, our approach ensures wavelength continuity in all established lightpaths, thus our algorithms do not require wavelength conversion capability at the M/C nodes. Unlike other studies, our aim is to present a comprehensive availability analysis, and an investigation on the network energy consumption for the proposed algorithms.

We study the survivable routing and wavelength assignment (RWA) problem with dual-homing to protect against either link and/or M/C node failures in the core segment. As discussed previously in subsection 4.2.4.A, the RWA problem is a key problem in establishing a set of all-optical connections (i.e., lightpaths) between pairs of source-destination M/C nodes over the fibre topology of the core network. The RWA problem is solved by assigning a single pair of lightpaths to each connection demand. Each lightpath consists of a physical route and a wavelength. The wavelength assignment (WA) part of the solution is subject to two main constraints: the wavelength clash constraint, which prohibits wavelength sharing amongst lightpaths traversing a common fibre; and the

wavelength continuity constraint, which ensures that each lightpath uses the same wavelength along its entire physical path. While the RWA problem itself has been shown to be NP-complete [195], the incorporation of protection further increases the RWA problem complexity, requiring heuristic approaches in order to solve realistic problem instances.

We consider two failure scenarios for the core network segment: (i) single link failures, and (ii) single failures of either a link or an intermediate node (i.e., a M/C node that is neither the source or the destination node of the considered lightpath). For each of the considered failure scenarios, we propose two heuristic algorithms to solve the RWA problem with dedicated path protection for a static traffic demand matrix with two different objectives: (i) minimising the number of used wavelengths, or (ii) minimising the total path length. Due to greater LE accessibility and to a higher number of alternatives for the protection paths, our proposed approach significantly reduces the average path length and the number of used wavelengths.

4.5 Dedicated Path Protection with Dual Homing (DPP-DH) Approach

In a dual-homed optical access network, a local exchange (LE) is connected to two different metro/core (M/C) nodes, protecting the access segment against feeder fibre and/or M/C node failures. This is for example the architectural option envisaged by the FP7 project DISCUS [212], and also the network architecture considered in this chapter. Fig. 4.11 presents an example of this concept where two dual-homed LEs and a core network comprising a number of M/C nodes interconnected by an optical circuit-switched wavelength layer. Typically, the M/C node closest to the LE is referred to as the primary home (denoted as *S_pri_home* for the source LE in Fig. 4.11) and it is used during default operating conditions, while the second closest M/C node represents the secondary home (denoted as *S_sec_home* in Fig. 4.11) and it is used when a failure takes down the primary home.

The survivable RWA problem with dedicated path protection in networks with a dual homed access segment can be formally defined as follows. Given the physical topol-

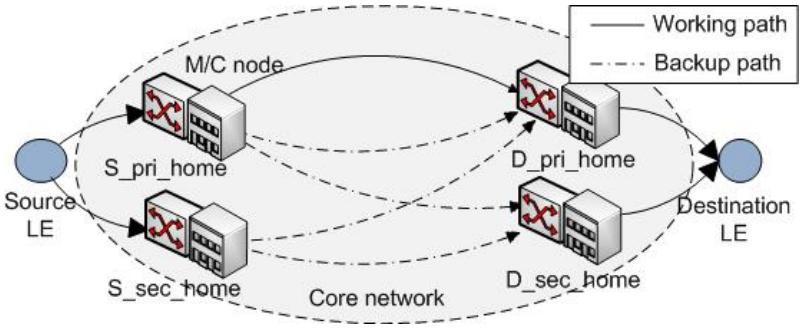


Figure 4.11: Core network architecture with dual-homed access network.

ogy of the network that comprises local exchanges and metro/core nodes interconnected by physical links (as shown in Fig. 4.11), and given a static traffic demand matrix between LEs, we must solve the RWA problem for a pair of physically disjoint working and backup paths between each source and destination M/C node pair that has traffic between them.

The problem is solved considering two failure scenarios for the core network, i.e., (i) link failures and (ii) M/C node or link failures. We assume only a single network element (i.e., link or M/C node) can fail at any point in time. We then solve the above problem considering two different optimisation objectives: (i) minimising the number of wavelengths used and (ii) minimising the total length of established lightpaths. We assume that the working path is always established between the two primary homes at each side, while the backup path can be established between any combination of the primary and secondary homes at the source and destination side. Here, we assure wavelength continuity for all established lightpaths and assume traffic grooming capability at each M/C node. Further, to maintain simplicity of the heuristic, we assume that all lightpaths to be detected at LE's without transmission errors.

We propose four variants of survivable RWA algorithm in dual-homed networks, each one considering a different failure scenario and a different optimisation objective. While the general instance of the problem is referred to as dedicated path protection with dual-homing (DPP-DH), each variant of the algorithm is denoted with an extension specifying the failure scenario it refers to and the optimisation objective it considers. The variants can be summarised as follows, where l stands for core link failures only, n for

both node and link failures, *W* denotes the minimisation of the wavelength usage, and *L* the minimisation of the lightpath length: Thus, the:

1. the DPP-DH-*l*-*W* algorithm considers core link failures only and minimises the number of used wavelengths,
2. the DPP-DH-*l*-*L* algorithm considers core link failures only and minimises the total path length,
3. the DPP-DH-*n*-*W* algorithm considers either core node or link failures and minimises the number of used wavelengths, and
4. the DPP-DH-*n*-*L* algorithm considers either core node or link failures and minimises the total path length.

4.5.1 Dedicated Path Protection with Dual Homing Considering Core Link Failures (DPP-DH-*l*)

To provide protection against a single core link failure in networks with dual homing, we propose two variants of the DPP-DH approach aimed at minimising the number of wavelengths used (denoted as DPP-DH-*l*-*W*), and minimising the total length of established lightpaths (denoted as DPP-DH-*l*-*L*).

A. DPP-DH-*l*-*W* Algorithm

The pseudocode of the DPP-DH algorithm is presented in Algorithm 7. During the initialisation phase, traffic demands are sorted in the descending order of the product of their capacity (*C*) and distance (*D*), i.e *C x D*. Here, the distance (*D*) is computed as the length of the shortest path between the source and destination node pair of the demand. This initial sorting makes sure that the demands with higher capacity requirements over greater distances are given priority when assigning resources. The sorted demands are then processed sequentially. In the first phase (steps 4-15), the algorithm finds a route and a wavelength for the working path in the following way. First, the shortest physical path between the primary homes of the source and destination LE is computed on

all of the wavelengths already in use, whose number is denoted as *nActive_Wavelengths* (steps 4-5). If all of the links on a computed shortest path do not have sufficient capacity to accommodate the current demand, the algorithm removes the most congested links of the path from consideration (step 7) and searches again by returning to step 5, thus iteratively reaching the shortest path with sufficient available capacity. The initial value of *nActive_Wavelengths* is set to 1 and is incremented only if a working path cannot be found on any of the already active wavelengths (step 11), in an effort to minimise the total number of wavelengths used. Once the shortest feasible (i.e., with sufficient available capacity) path is found for the current demand, network resources for the working path are reserved in step 15. Similarly, steps 16-27 find and reserve network resources for the shortest link-disjoint backup path.

Algorithm 7 Pseudocode for DPP-DH-*l*-*W* algorithm

```
1:  SORT demands by C x D (descending)
2:  nActive_Wavelengths = 1
3:  for each demand do
4:      for each nActive_Wavelengths do
5:          Find the shortest path between primary homes on current wavelength
6:          if capacity unavailable in path then
7:              Disconnect most congested link
8:              go to step 5
9:      if path not found then
10:         nActive_Wavelengths ++
11:         go to step 4
12:     Select the shortest path amongst all active wavelengths and assign resources
13:     for each nActive_Wavelengths do
14:         Find shortest link-disjoint backup path on current wavelength
15:         if capacity unavailable in path then
16:             Disconnect most congested link
17:             go to step 17
18:     if backup path not found then
19:         nActive_Wavelengths ++
20:         go to step 16
21:     Select shortest path and assign resources
```

B. DPP-DH-*l*-*L* Algorithm

The second variant of DPP-DH-l algorithm, denoted as DPP-DH-*l*-*L*, aims at minimising the total length of the established primary and backup paths in an effort to improve the utilisation of wavelength resources. The pseudocode of the DPP-DH-*l*-*L* algorithm is shown in Algorithm 8. The DPP-DH-*l*-*L* algorithm is similar to DPP-DH-*l*-*W*, with the exception of considering all wavelengths supported by the system when searching for the shortest path pair for each demand. The goal is to minimise the total path length even at the cost of greater wavelength utilisation.

4.5.2 Dedicated Path Protection with Dual Homing Considering Core Node and Link Failures (DPP-DH-*n*)

The DPP-DH-*l* approach provides protection against core link failures, but it is not able to guarantee protection in the event of a M/C node failure. Even if the working and backup paths found by DPP-DH-*l* are link disjoint, they might share a common intermediate node in the network which then represents a single point of failure affecting both paths. An example of such scenario is shown in Fig. 4.12.

To provide protection in the presence of a single M/C node failure, we extend the DPP-DH-*l* approach to account for node-disjointedness of the working and the backup

Algorithm 8 Pseudocode for DPP-DH-*l*-*L* algorithm

```
1:  SORT demands by C x D (descending)
2:  for each demand do
3:      for each wavelength do
4:          Find shortest path between primary homes
5:          if capacity unavailable in path then
6:              Disconnect most congested link
7:              go to step 4
8:      Select shortest path and assign resources
9:      for each wavelength do
10:         Find shortest link-disjoint backup path
11:         if capacity unavailable in path then
12:             Disconnect most congested link
13:             go to step 12
14:     Select shortest path and assign resources
```

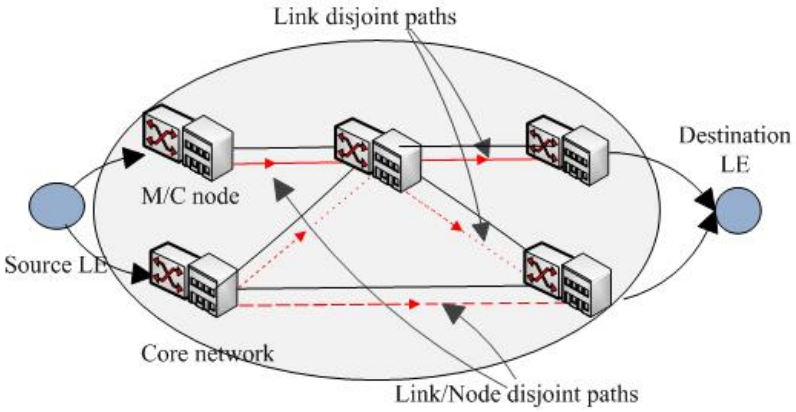


Figure 4.12: Dual-homed LE architecture with link disjoint paths and link/node disjoint paths.

path of each connection. Here, note that the condition we enforced earlier in DPP-DH-*l* approach, i.e. the primary paths are always considered to traverse through the primary homes for operational simplicity, is still enforced. Therefore, ensuring complete node-disjointness would relate to restricting the backup paths to traverse through the secondary homes at both source and destination ends. This increases the traffic congestion in the network, thus affecting the effectiveness of network resource utilisation.

The solution is denoted as DPP-DH-*n*. As it was the case for DPP-DH-*l*, we develop two variants of the DPP-DH-*n* approach aimed at minimising the number of wavelengths and minimising the total path length.

A. DPP-DH-*n*-*W* Algorithm

The objective of the DPP-DH-*n*-*W* algorithm is to minimise the number of wavelengths used in the network while establishing node-disjoint working and backup paths for each connection requests. Algorithm 9 presents the pseudocode of DPP-DH-*n*-*W*. The basic structure of the DPP-DH-*n*-*W* algorithm is similar to its counterpart which considers only link failures, i.e. the DPP-DH-*l*-*W* algorithm, with a few additional steps added to account for the node-disjointedness of the working and backup path pairs.

In the first phase, the DPP-DH-*n*-*W* algorithm searches for the shortest working path from the source LE to the destination LE (steps 4-15 in Algorithm 9). Once the working path is found, DPP-DH-*n*-*W* temporarily removes the intermediate M/C nodes included

in the working path from the network (step 17) and searches for the backup path in the resulting modified topology (steps 16-28). Finally, the shortest backup path is selected amongst all backup path options on each wavelength, while network resources are assigned to the selected backup path in step 29.

B. DPP-DH-*n-L* Algorithm

The second variant of the node-disjoint DPP-DH approach, denoted as DPP-DH-*n-L*, aims at minimising the total path length of the primary and backup paths allowing a better utilisation of the wavelength resources. The DPP-DH-*n-L* algorithm is similar to the DPP-DH-*n-W*, with the difference of considering all wavelengths when searching for the shortest path pair for each demand, without attempting to minimise their usage but minimising the total path length instead. The pseudocode of the DPP-DH-*n-L* algorithm

Algorithm 9 Pseudocode for DPP-DH-*n-W* algorithm

```
1:  SORT demands by C x D (descending)
2:  nActive_Wavelengths = 1
3:  for each demand do
4:      for e doach nActive_Wavelengths
5:          Find the shortest path between primary homes on current wavelength
6:          if capacity unavailable in path then
7:              Disconnect most congested link
8:              go to step 5
9:      if path not found then
10:         nActive_Wavelengths ++
11:         go to step 4
12:     Select shortest path amongst all active wavelengths and assign resources
13:     for each nActive_Wavelengths do
14:         Disconnect intermediate nodes of primary path
15:         Find shortest backup path on current wavelength
16:         if capacity unavailable in path then
17:             Disconnect most congested link
18:             go to step 17
19:         Re-connect intermediate nodes from primary path
20:     if backup path not found then
21:         nActive_Wavelengths ++
22:         go to step 16
23:     Select shortest path and assign resources
```

is presented in Algorithm 10.

4.5.3 Model Evaluation of DPP-DH

In this section, we describe the traffic model and network parameters used in the simulation. The DPP-DH algorithms exploit the increased accessibility of the LEs due to dual-homing architecture in the access network to provide better survivability compared to their baseline counterparts. Thus, the improvements provided by the algorithm variants depend on the size of the considered network topology where better improvements are expected in a larger network where more path options are available. In our analysis, we consider a countrywide network topology representing the optical communication network of Ireland comprising 20 M/C nodes. The M/C nodes are connected with 33 links and 1204 LEs, locations of which are shown in Fig. 4.13. The proposed algorithms were implemented in C# and evaluated via simulation on the countrywide network topology of Ireland. The network parameters used in the simulation work is listed in Table 4.1.

Algorithm 10 Pseudocode for DPP-DH-*n*-*L* algorithm

```
1:  SORT demands by C x D (descending)
2:  for each demand do
3:      for each wavelength do
4:          Find the shortest path between primary homes on current wavelength
5:          if capacity unavailable in path then
6:              Disconnect most congested link
7:              go to step 4
8:      Select shortest path and assign resources
9:      for each wavelength do
10:         Disconnect intermediate nodes of primary path
11:         Find shortest backup path
12:         if capacity unavailable in path then
13:             Disconnect most congested link
14:             go to step 13
15:         Re-connect intermediate nodes from primary path
16:     Select shortest path and assign resources
```

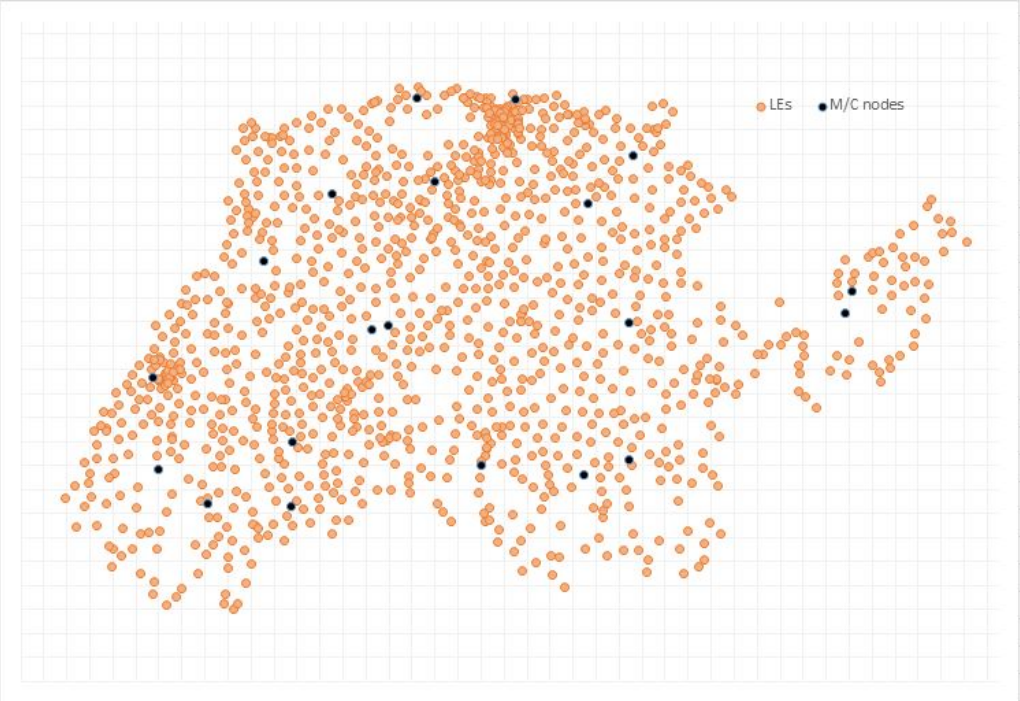


Figure 4.13: Network topology of Ireland.

4.5.3.1 Traffic Model

The traffic between each pair of LEs was generated based on the number of users served by each LE. The traffic matrix was generated using the gravity model from [232], as shown in Eq. 4.4.

$$T_{AB} = (K \times C \times N_A \times N_B)/(100 \times L^2) \tag{4.4}$$

In Eq. 4.4, T_{AB} denotes the traffic between local exchanges A and B, each one serving N_A and N_B users, respectively. K is the traffic load factor used for tuning the total traffic offered to the network, C is the peak capacity provided per user, while L denotes the Euclidean distance between the two LEs. In our simulation, K is varied to model total network traffic intensities at 68 Tbit/s (0.25 normalised traffic load), 135 Tbit/s (0.5 normalised traffic load) and 203 Tbit/s (0.75 normalised traffic load). Here, a static traffic matrix of all traffic demands between all LEs is implemented representing the worst case scenario. The physical topology of network comprising M/C nodes locations, LE loca-

Table 4.1: Simulation parameters.

Parameter	**Value**
Channel capacity	100 Gbps
Peak capacity per user (C)	10 Mbps
Number of wavelengths	80
Link capacity	8 Tbps (100 Gbps x 80 wavelengths)
Full network load	264 Tbps (100 Gbps x 80 wavelengths x 33 links)
Load factor (K)	0.25 (68 Tbps - network load of 0.25) 0.5 (135 Tbps - network load of 0.51) 0.75 (203 Tbps - network load of 0.77)

tions, set of physical links and wavelength capacity, and the generated traffic matrix, are then input to the proposed algorithms for simulation.

The proposed algorithms are compared against a baseline case, in which dual-homing capability is considered absent in the access network, that is, each LE is connected to only one (i.e., physically closest) M/C node with duplicated feeder fibres (to ensure functionality in case of a feeder fibre failure). For a fair comparison with their dual-homed counterparts, we analyse several variants of the baseline approach. The baseline approaches which consider only link failures while minimising wavelength usage, and minimising path length, are denoted as DPP-l-W and DPP-l-L, respectively. Similarly, DPP-n-W and DPP-n-L denote their respective node-disjoint variants. The obtained results are analysed in the next section.

4.5.3.2 Power Consumption Model

In this section, the formulation of the numerical model to estimate power consumption of the core network is described. In this work, we consider the total power consumption of a core network to be the sum of the power consumptions of the individual transmission links and M/C nodes. For the transmission links, we consider the use of an erbium doped fibre amplifier (EDFA) that consumes 8 W to be placed every 80 km to compensate for fibre losses [233]. Further, we consider the elements in the M/C node architecture design

as presented by FP7 DISCUS [234]. The power consumption of a M/C node i, denoted as $PM/Cnode_i$, is calculated as follows:

$$P_{M/C\_node\_i} = P_{O\_switch\_i} + P_{IP\_router\_i} + P_{E\_switch\_i} + K_i \times P_{EDFA\_line\_board\_i} \tag{4.5}$$

where $P_{O\_switch\_i}$, $P_{IP\_router\_i}$, and $P_{E\_switch\_i}$ represent the power consumption of the optical switch, packet (IP) routers, and Ethernet switches, respectively. The parameter K_i and $P_{EDFA\_line\_board\_i}$ represent the number of degrees of the i^{th} M/C node and the power consumption of the i^{th} EDFA line board. The parameter $P_{IP\_router\_i}$ is further calculated as follows:

$$P_{IP\_router\_i} = N_i \times P_{electronic\_processing} + M_i \times P_{transponder} \tag{4.6}$$

where N_i is the number of active IP routers in the M/C node, which is calculated based on the number of active wavelengths. The parameter $P_{electronic\_processing}$ is the contribution from the electronically processed traffic and fixed elements of the router, e.g. control board, fans and fabric board. The parameter M_i is the number of active wavelengths and the parameter $P_{transponder}$ denotes the power consumption of the configurable 100 Gbps transponders.

The power consumption of the network is estimated by numerical evaluation of the model. In our analysis, we consider the power consumption of the network components as listed in Table 4.2. The obtained results from the analysis are presented and discussed in the next section.

The above power consumption model is used to estimate the total power consumption of the network for a given instant by aggregating individual power consumptions of all network elements while all connection requests are provisioned with a primary and backup path pair considering dual-homing in the access network (denoted as $P_{total,dual-homing}$). Further, we also estimate the power consumption of the same demand matrix considering a baseline network architecture without dual-homing capability (denoted as $P_{total,baseline}$). In order to compare results, we introduce the parameter, *percentage of power saving* as follows.

Component	Power
Optical Switch ($P_{O\_switch\_i}$) [234]	
• 2-degree ROADM	975 W
• 3-degree ROADM	1145 W
• 4-degree ROADM	1315 W
• 5-degree ROADM	1485 W
IP Router (MPLS-TP) ($P_{IP\_router\_i}$) [234]	
• Electronic processing ($P_{electronic\_processing}$)	2840 W
• Configurable transponder ($P_{transponder}$)	100 W
Ethernet Switch ($P_{E\_switch\_i}$) [235]	2766 W
In-line EDFA [233]	8 W
EDFA line board ($P_{EDFA\_line\_board\_i}$) [234]	110 W

Table 4.2: Considered power consumption values of the components/elements of the core network.

$$Percentage\ power\ saving = \frac{P_{total,baseline} - P_{total,dual-homing}}{P_{total,baseline}} \times 100 \tag{4.7}$$

4.5.4 Results and Discussion

In our analysis, the proposed algorithms are compared to their respective baseline cases in which dual homing is not implemented in the access network. As such, each LE is connected to only the physically closest M/C node with a two feeder fibres to ensure fibre protection in case of a feeder fibre failure. As with the proposed algorithms, four baseline DPP variants where (a) only link failures while minimising wavelength usage, DPP-*l*-*W*, (b) only link failures while minimising path length, DPP-*l*-*L*, (c) node or link failures while minimising wavelength usage, DPP-*n*-*W*, and (d) node or link failures while minimising path length, DPP-*n*-*L*, are considered.

Allocation of Network Resources

Figs. 4.14 and Fig. 4.15 show the average total path length (working and backup paths) obtained using the algorithm variants which minimise the number of used wavelengths and path length, respectively. The total normalised network load considered is 0.5. Since the virtual topology of the network is fully connected for all traffic loads and the net-

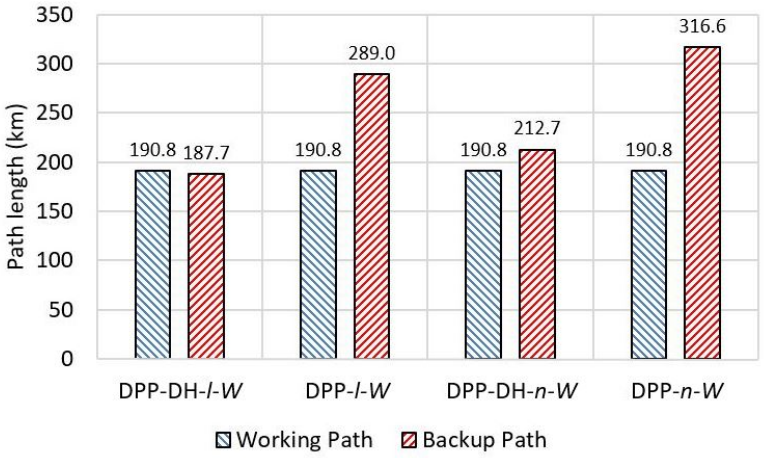


Figure 4.14: Average path lengths for dual-homed and baseline architectures using DPP-DH-*l*-*W* and DPP-DH-*n*-*W* (number of active wavelengths minimised) for normalised total traffic load of 0.5.

work is dimensioned to support the highest traffic intensity, the average path lengths are relatively insensitive to the enforced traffic intensity and are primarily dependent on the physical network architecture itself, i.e. location of M/C nodes. Thus, path length results for lower traffic loads are omitted in Fig. 4.14 and Fig. 4.15 to maintain simplicity and avoid duplication. As the working paths are always forced to use the primary homes of the source and destination LE, the length of the working path is the same for all algorithms. Further, due to the higher number of alternative routes for establishing backup paths, the selected routes are shorter compared to the baseline DPP cases with no dual homing. Shorter paths result in a reduction of resource usage, which in turn reduce network operator costs. When compared to the baseline DPP cases, the results in Fig. 4.14 and Fig. 4.15 show that utilising dual homing for protection reduces the backup path lengths by 35% when considering link disjointness (for single link protection) and 32% when considering node-disjointedness (for link or M/C node protection). When comparing the DPP-DH-*l* and DPP-DH-*n* approaches, introducing node-disjointedness between working and backup paths increases the average length of the backup paths by 9% in both variants. Therefore, protection from M/C node failures is obtained with a trade-off of a small increase in backup path length.

When comparing Fig. 4.14 against Fig. 4.15, we note that using the algorithm to op-

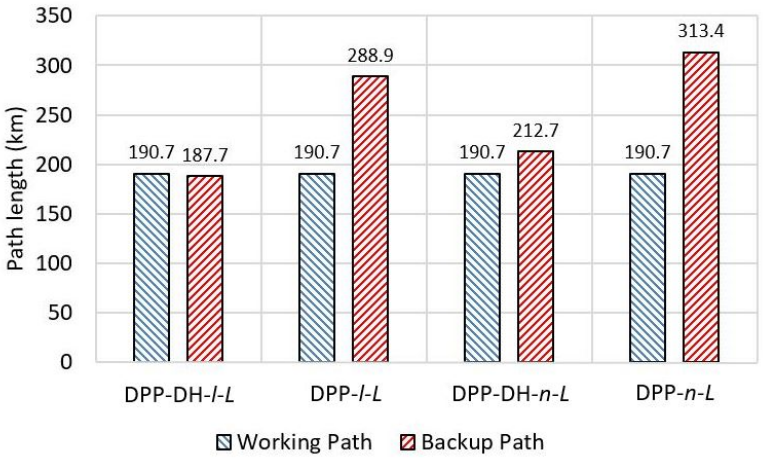


Figure 4.15: Average path lengths for dual-homed and baseline architectures using DPP-DH-*l*-*L* and DPP-DH-*n*-*L* (path length minimised) for normalised traffic load of 0.5.

timise path length (Fig. 4.14) does not significantly reduce the path lengths compared to the wavelength optimising algorithm (Fig. 4.15). This is due to the results being average values of a large number of connections (all-to-all connection matrix between 1204 LEs), where connections are between source and destination LEs that are typically located very far from each other (even 1000s of kms). Hence, the reduction is path length achieved by selecting different route options is rather little compared to the end-to-end distance between the source and destination of the connection. This is the reason for the almost non-existent difference between observed path lengths in Fig. 4.14 and Fig. 4.15.

Fig. 4.16 shows the number of active wavelengths used in the core network comparing the dual-homed and baseline network architectures. Different network loads representing low (normalised total network load of 0.25), medium (normalised total network load of 0.5), and high (normalised total network load of 0.75) traffic intensities are considered. In all variants, the number of active wavelengths used increases with the traffic load in the network. However, in all cases, DPP-DH-*l* and DPP-DH-*n* obtain a lower number of used wavelengths than the baseline DPP case, a beneficial feature which directly reduces the operational costs of the network. Comparing the variants of DPP-DH, it is evident that DPP-DH-*l*-*W* and DPP-DH-*n*-*W* use fewer wavelengths than DPP-DH-*l*-*L* and DPP-DH-*n*-*L*, respectively, which agrees with the respective optimisation objectives. Comparing the DPP-DH approach against the baseline architectures, on average,

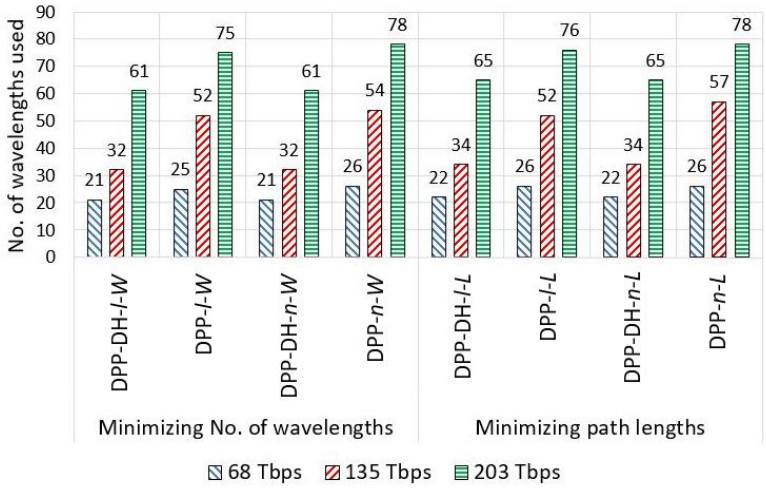


Figure 4.16: Number of wavelengths used by algorithm variants on dual-homed and baseline architecture under different network loads.

the number of used wavelengths is reduced by 22% in the path length minimising variants, and by 25% in the wavelength utilisation minimising variants. The results also suggest that introducing node-disjointedness to the algorithms does not yield a significant increase in the number of used wavelengths. On average, only a 4% increment of is observed in the number of used wavelengths between the link-disjoint and node-disjoint DPP-DH algorithm variants.

Availability Analysis

Using availability values reported in [236], we calculate the availability of dedicated path protected connections and evaluate the percentages of connections satisfying different availability categories for the network. A brief introduction to the availability calculation was previously presented in subsection 4.2.4.B.

Fig. 4.17 and Fig. 4.18 show the percentages of connections satisfying different availability categories for the network traffic load of 0.5. All variants of DPP-DH obtain significant improvements in the percentage of connections with high availability (99.999% and 99.995%), due to the path length reduction enabled by dual homing. On average, amongst all considered cases, results show a 77% and 13% improvement in the portion of connections with 99.999% and 99.995% availability, respectively, compared to the baseline

DPP approach.

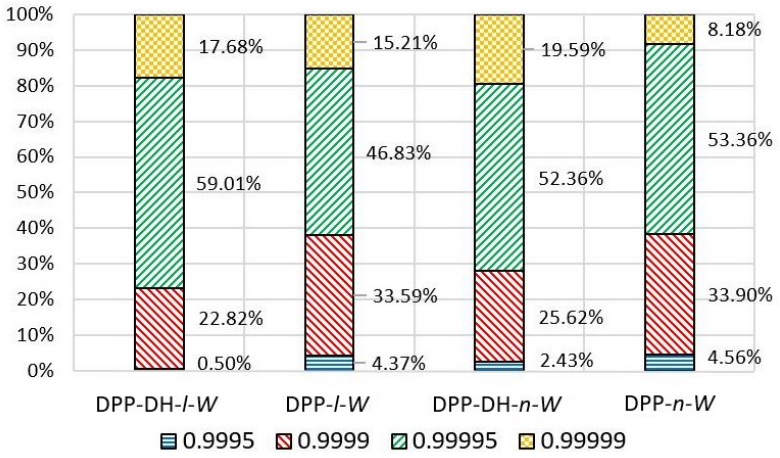


Figure 4.17: Percentage of connections in different availability categories from DPP-DH-*l*-*W* and DPP-DH-*n*-*W*.

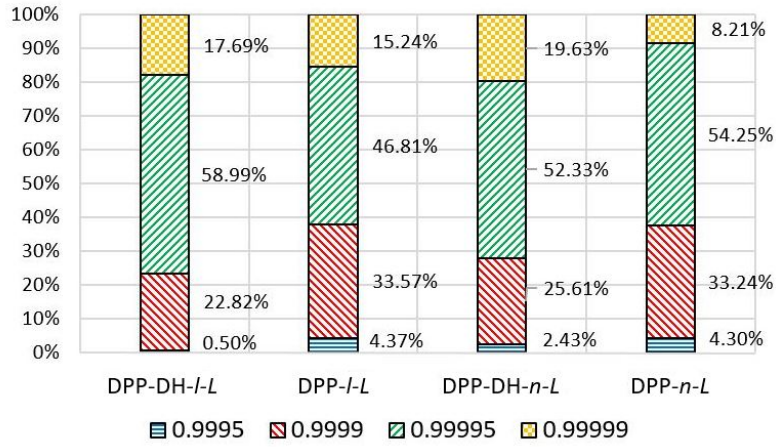


Figure 4.18: Percentage of connections in different availability categories from DPP-DH-*l*-*L* and DPP-DH-*n*-*L*.

Furthermore, comparing the portion of connections with the highest availability at 99.999%, DPP-DH-*n* variants show a 11% increment compared to the DPP-DH-*l* variants, eg. 17% in DPP-DH-*l*-*W*, compared to 19% in DPP-DH-*n*-*W* in Fig. 4.17. This improvement is due to the elimination of a single point failure in DPP-DH-*n*-*W*. It is evident that the DPP-DH approaches achieves higher connection availability while utilising fewer network resources compared to the baseline DPP cases with no dual homing. Results for

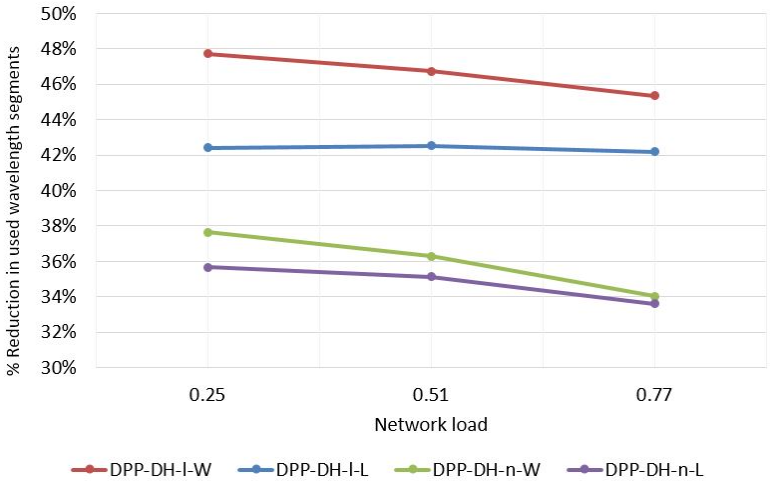


Figure 4.19: Percentage of reduction in number of used wavelength segments against the baseline approach of each variant.

other normalised total network traffic loads at 0.25 and 0.75 show similar improvements to the connection availability and are not presented here.

Power Savings

Fig. 4.19 plots the percentage of reduction in the number of active wavelength segments for each algorithm compared to their baseline DPP counterparts under three traffic matrices representing low, medium, and high traffic levels (i.e. network loads of 0.25, 0.51 and 0.77 respectively). A wavelength segment is a single wavelength on a single fibre link. Hence, a zero percent reduction represents the scenario whereby a same number of wavelength segments out of all 2650 wavelength segments (= 80 wavelengths x 33 links) are utilised in both dual-homed and baseline cases. Results show that by utilising dual homing, the reduction in the number of used wavelength segments varies from 33% (for DPP-DH-*n*-*L*) to 48% (for DPP-DH-*l*-*W*) in comparison to their baseline DPP counterparts. As network load increases, the reduction in wavelength usage decreases due to higher wavelength utilisation. However, when effectiveness of wavelength utilisation is considered, DPP-DH-*l*-*W* outperforms all other algorithm variants irrespective of the network load.

Fig. 4.20 shows the percentage of power savings calculated based on the power con-

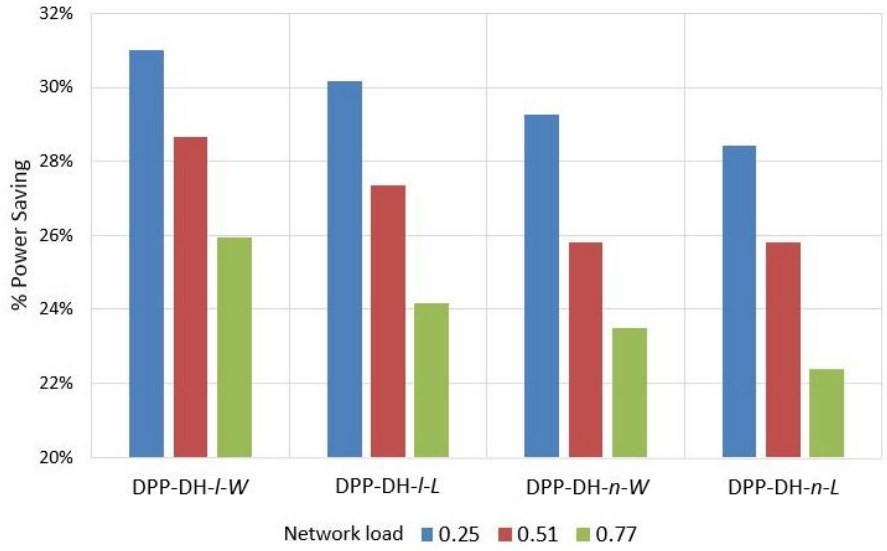


Figure 4.20: Percentage of power savings of the DPP-DH algorithm variants under different network loads.

sumption model discussed in subsection 4.5.3.2. Results correlate with those in Fig. 4.19, showing that the reduction of active wavelengths segments directly affects the power consumption of the core network. The percentage of power savings ranges from 22% (for DPP-DH-*n*-*L*) to 31% (DPP-DH-l-W). Overall, when comparing the DPP-DH-*l* approach which considers only link failures against the DPP-DH-*n* approach which considers intermediate M/C node failures as well, our results show that DPP-DH-*l* in general outperforms DPP-DH-*n*. Due to node-disjointedness constraint, the DPP-DH-*n* algorithms experience a higher congestion, and therefore uses more wavelengths. As such the percentage of power savings is lower as compared to DPP-DH-*l* algorithms. This reduction of wavelength usage can be considered as the trade-off for a higher degree of resiliency provided in the network by the DPP-DH-*n* approach, as discussed in Section 4.5.2.

4.6 Summary

This chapter studied the benefits of utilising dual homing in the optical access networks for providing resource-efficient protection against link and node failures in the optical core segment. Novel, heuristic-based RWA algorithms were proposed to provide dedi-

cated path protection in networks with dual homing (DPP-DH). Four variants of the DPP-DH algorithm were proposed which considered different failure scenarios (i.e., single link or node failures) and optimisation objectives (i.e., minimisation of wavelength usage and path length). Compared to baseline network architectures without dual homing, our approaches reduced the backup path length by up to 35% on average, used up to 25% fewer wavelengths, and increased the portion of connections with 99.999% and 99.995% availability by 77% and 13%, respectively. Furthermore, the variants of the algorithm which considered core node failures additionally enhanced core network survivability at the expense of only a small increase of the backup path length (i.e., 9%) with respect to the link failure protection approach. Therefore, the obtained results strongly indicated that core network survivability could be improved by exploiting a dual homed architecture in the access network in a cost efficient manner. A critical study on the energy-efficiency of the network is also performed. Simulation results showed up to 48% reduction in the wavelength usage compared to the baseline DPP case. The estimated power consumption of the network showed that DPP-DH can provide up to 31% power savings. Out of the four algorithm variants proposed, results showed that the DPP-DH-*l*-*W* algorithm yields the greatest power savings in the core network while providing protection against single link failures. Therefore, the obtained results clearly indicated that exploiting a dual-homed architecture in the access segment can bring benefits in terms of both improved survivability and reduced power consumption in the core network.

Chapter 5
Conclusion

This chapter summarises the key contributions of the thesis and presents some of the future research directions arising from our efforts.

5.1 Summary of Contributions

With the ever increasing demand of the Internet, it is evident that the design of future telecommunication networks should be able to provide higher bandwidth capacity with higher QoS covering a larger geographical area. Optical WDM networks have been introduced as the most feasible and adaptable network solution for the requirements of such future communication networks. However, as a result of increasing deployment of information and communications technology (ICT) infrastructure, the increase of energy consumption have increasingly contributed towards the global carbon footprint. This thesis presented our efforts towards designing energy-efficient future telecommunication networks. In chapters 2, 3, and 4, we focused on the fibre-based optical access network segment, wireless mobile access network segment, and the optical core network segment, respectively. Each chapter presented energy-efficient architectural and algorithm-based solutions to meet the requirements of energy-efficient next-generation telecommunications networks.

In Chapter 1, we presented an introduction to the thesis, which highlighted the motivation, the outline, and the original contributions of the thesis. We discussed the evolution of optical communication systems and the need for energy-efficiency in the design of future telecommunication networks. Chapter 1 also presented a list of publications arising from each of the chapters of the thesis.

In Chapter 2, we focused on fibre-based optical access networks and the need to provide increased bandwidth capacity and QoS. We chose VoD application due to the fact that VoD is identified as the main traffic generating application in the current telecommunications networks. Moreover, forecasts suggest that globally, video-related traffic over the Internet will grow substantially in the future. The chapter presented an introduction to PONs and their functionality, followed by a discussion on the prevailing VoD system architectures. The chapter also presented a comprehensive review of the related works that focus on optimising VoD delivery found in literature. Next, we proposed three dynamic bandwidth allocation algorithms to optimise VoD delivery over passive optical networks by exploiting a strategically placed LS server maintaining the set of most popular video content closer to the customer. We also proposed two receiver configurations; namely the single receiver configuration, and double receiver configuration. The three algorithms, DR-DBA, SR-DBA, and SRDC-DBA are compared against the conventional CO-based VoD delivery. Moreover, we formulated mathematical models to estimate the network operational power consumption and deployment cost of the LS-based PON. Simulation results showed that the delay and achievable throughput levels are improved when a LS is utilised. Nonetheless, jitter levels are increased with the use of the LS. However, the increment could be easily alleviated by using jitter buffers at the ONUs. Comparing the single receiver-based algorithms (SR-DBA and SRDC-DBA) against the dual receiver-based DR-DBA algorithm, results showed that both algorithms produce almost similar delay improvements and jitter variations. However, when throughput is considered, both DR-DBA and SRDC-DBA algorithms outperforms SR-DBA through the use of dual channel downstream transmission. With the introduction of the active element (the LS server) to the access network, an increase in the network power consumption was inevitable. However, the results from the power consumption models showed that when an access network of 1024 ONUs is considered (32 PONs with 32 ONUs in each PON), the single receiver-based algorithms, namely, SR-DBA and SRDC-DBA, only result in a power increment of 4% per ONU where as DR-DBA showed a high power increment of 54% per ONU compared to the no LS case. Therefore, comparatively, the SRDC-DBA algorithm delivers similar QoS and throughput improvements as the dual receiver-based

DR-DBA algorithm, but consumes a substantially less amount of power. Further, results from the cost model proved that the total deployment cost of the LS-based PON is mainly contributed by the fibre-related expenditure while the cost component arose from LS equipments is <2% of total deployment cost per ONU. For these reasons, we recommended the use of LS caching servers to improve VoD services in the next-generation telecommunications networks.

In Chapter 3, we focused on the wireless mobile access network.

We used the concept of cell zooming to propose transmit power control mechanisms for considering a heterogeneous wireless access network system. The chapter presented an introduction which discussed the most commonly recognised wireless broadband access network technologies and different architectures. The heterogeneous mobile access network architecture is discussed in detail. Next, the chapter presented a comprehensive review of the related work found in literature on the energy-efficiency of mobile access networks. We proposed two approaches to solve the transmit power control problem of such heterogeneous mobile access networks. First, we proposed a heuristic-based algorithm to estimated the optimal operational transmit power for the deployed BSs in the mobile access network. Second, we presented our efforts in solving the same problem by formulating a LP. The LP approach produced global optimal solution whereas the heuristic-based algorithm produced a near-optimal solution. However, the heuristic-based approach was computationally straightforward compared to the LP. As a result, the heuristic approach was able to model larger network segments than the LP. When the two approaches were simulated considering a single macro cell system, results showed improvement in energy-efficiency when considering traffic-based transmission power control for the small cell BSs. In the heuristic approach, results presented a near-optimal solution for transmit powers of the small cell BSs. The total transmission power of the network was reduced by 12% on average during the 24-hour time span of a day whereas the LP showed a 14% reduction of total transmission power on average. As far as the optimality of the solutions is concerned, the LP presents a highly accurate and globally-optimal solution, but with greater computational complexity. In the heuristic-based approach, even though the results are sub-optimal, due to the simplicity of the heuristic

approach, it could model complicated network architectures with more flexibility. However, considering the results, we can conclude that effective transmission power control which considers traffic demand variations of the network, can improve the energy-efficiency of the future wireless access networks.

In Chapter 4, we discussed the core segment of telecommunications networks and its fault tolerance against possible failures in the fibre links and networking hardware. Here, we exploited the DH capability in the access network to improve core network survivability. The chapter started with an introduction which discusses WDM mesh networks, their operation and fault management. Further, we discussed the DH architecture of the access network. Next, heuristic-based RWA algorithms were proposed to provide dedicated path protection in networks with dual homing (DPP-DH). Four variants of the DPP-DH algorithm were proposed which considered different failure scenarios (i.e., single link or node failures) and optimisation objectives (i.e., minimisation of wavelength usage and path length). Compared to baseline network architectures without DH, our approaches reduced the backup path length by up to 35% on average, used up to 25% fewer wavelengths, and increased the portion of connections with 99.999% and 99.995% availability by 77% and 13%, respectively. Furthermore, the variants of the algorithm which considered core node failures additionally enhanced core network survivability at the expense of only a small increase of the backup path length (i.e., 9%) with respect to the link failure protection approach. Therefore, the obtained results strongly indicated that core network survivability could be improved by exploiting a DH architecture in the access network in a cost-efficient manner. A critical study on the energy-efficiency of the network is also performed. Simulation results showed up to 48% reduction in wavelength usage compared to the baseline DPP case. The estimated power consumption of the network showed that DPP-DH can provide up to 31% power savings. Out of the four algorithm variants proposed, results showed that the DPP-DH-*l*-*W* algorithm yields the greatest power savings in the core network while providing protection against single link failures. Therefore, the obtained results clearly indicated that exploiting a DH architecture in the access segment can bring benefits in terms of both improved survivability and reduced power consumption in the core network.

5.2 Future Directions

The direction of research on energy-efficient telecommunications networks is a continuously changing subject due to the availability of new concepts, architectures and technologies. In this section, we discuss the potential future directions of our work proposed in this thesis.

5.2.1 Energy- and Cost-Efficient VoD Systems with Cloud Infrastructure

Cloud computing allows VoD service providers to rent resources such as networks, servers, storage, applications, and services from cloud service providers, without investing in their own infrastructure. Today, there are many commercial cloud platforms such as Amazon AWS [75] and Microsoft Azure [76], that allows users to rent virtual computers and storage resources to run their own applications. Deploying VoD services using cloud resources is an attractive solution to VoD service providers. Due to high power consuming cloud resources being shared amongst many clients, cloud computing offers opportunity for VoD service providers to operate with minimal self-owned resources and directly reduce the deployment costs as well as the operational power consumption. Such an approach allows them to focus more on service level considerations while leaving the infrastructure level issues such as reliability and scalability concerns to the cloud service providers. However, most of the VoD service providers have already deployed their own data centres and therefore have their own infrastructure. Therefore, a hybrid architecture that uses both the cloud-based resource and the existing self-owned resource will be an interesting and a energy-efficient solution for those VoD service providers.

In Chapter 2, we proposed a energy-efficient VoD system that exploited storing the videos in strategically placed, self-owned local storage servers. A hybrid VoD solution could be considered as a future direction to continue this work. We believe that there are many research questions that will need to be addressed in order to design an energy-efficient and cost optimal hybrid VoD system that uses both cloud and self-owned resources. As the cloud resources can be rented on the fly when needed, a rent schedule that minimises the cost and/or maximise the efficiency, is required. For example, one

option is to rent and release the cloud resources at a pre-defined times during the busy hours whereas another option is to do that based on a real time load measurements. It is an interesting research question to investigate the energy-efficiency of the VoD system when different rent schedule options are adapted.

There are also content placement and load balancing aspects that need to be addressed in a hybrid VoD solution. Most probably, replicating all the videos stored in the self-owned data centres in the rented cloud storage might not be a cost-efficient option. In such instances, we need to carefully select the set of videos to be stored in the cloud storage and the set of videos to be stored only in the self-owned servers. Moreover, when a request arrives, we may have the option to use either cloud infrastructure or self-owned infrastructure to serve the request. In such cases, strategically selecting the video server while minimising the operational costs such as power consumption is another interesting research topic for the future VoD systems.

Many recent work tried to analyse and design data centres, network equipment, and Internet services, from an energy conservation standpoint [237, 238]. We hope to explore the possibilities of improving the energy-efficiency of VoD systems through scheduling and load balancing. In particular, we hope to consider a general deployment scenario of a hybrid VoD system where some of the data centres are powered by renewable power sources. As discussed before, due to the constraints on storage space and server bandwidth capacities, storing all the videos in a single data centre or serving all the users from a single server may not be possible. Therefore, we hope to study methods to improve the energy-efficiency of such VoD deployments by addressing the concerns such as placement of content across different serves, scheduling of requests, balancing the load amongst servers, etc.

5.2.2 Energy Harvesting Small Cell Networks

With the dense deployment of small cell BSs that is needed to meet the capacity and coverage requirement of next-generation wireless networks, it becomes increasingly difficult to provide grid power supply to all the small cell BSs in a cost-effective manner. Energy harvesting small cell networks (SCNs) that consists of off-grid and green energy

powered small cell BSs have be considered as a viable solution . Energy harvesting technology is a promising solution, which can harvest ambient renewable energy (e.g., solar and wind energy) to power small cell BSs [239]. Interestingly, it is estimated that applying energy harvesting techniques to SCNs can achieve a 20% reduction in CO_2 emissions in the ICT industry. In Chapter 3, we proposed a transmit power control mechanism towards a design of a energy-efficient heterogeneous network. We believe the initial phase of the future heterogeneous network will consist of a combination of grid-powered small cell BSs and off-grid renewable-energy-powered small cell BSs. It is a interesting research problem to consider the transmit power control for such hybrid heterogeneous networks. For example, a more energy-efficient transmit power control method could exploit the off-gird powered small cells to maintain coverage over a larger area during the off-peak hours, while the grid-powered BSs are transitioned to a sleep mode or power saving mode. Further, strategic decision of the deployment location such renewable-energy-powered small cells is also a interesting future research direction of our work.

5.2.3 Considering the Sleep-Mode Features for Energy-Efficient Core Network Survivability

The nominal power consumption of network devices and equipment considered for installation must be assessed during the network design phase. Emerging equipment features, such as sleep-mode and dynamic reconfiguration of modulation format and transmission reach, could provide new opportunities to reduce the power consumption of the backup links, by which the primary links are protected against possible fibre and/or hardware failure. In Chapter 4, we proposed a protection scheme which exploits the DH features of the access network to provide more energy-efficient survivability against probable failures. However, the novel sleep-mode feature emerging in the network equipment could help improve the energy-efficiency. For example, a more energy-efficient protection scheme could be devised to exploit sleep-modes at the expense of slower recovery. A new protection scheme could also allow multi-rate transponders to fall back to lower transmission rates depending on the service level agreement and reduced energy consumption.

Another interesting future direction of our work is investigating the trade-offs in dynamic adaptation of the network equipment depending on the temporary variations of traffic. Traffic in the core network segment is usually higher during the day and lower during the night. Putting idle devices into sleep or energy saving mode (e.g. adaptation of transmission rates) can effectively reduce the energy consumption of the backup resources. Another strategy relies on concatenating backup paths into separate fibres in order to be able to put the devices along these links into sleep-mode without being contained by the presence of working paths. Moreover, it will be important to determine the best trade-off between resilience, energy-efficiency, and low reconfiguration costs.

5.3 Summary of Thesis

In this thesis, our efforts towards a design of energy-efficient next-generation telecommunications networks are presented. The thesis consists of five chapters, in Chapter 1 (Introduction) of the thesis, we discussed the evolution of the optical communication systems, research motivation, and a brief outline of the thesis. The three contribution chapters (Chapters 2, 3, and 4), focus on different segments of the telecommunications network individually. In Chapter 2, we focus on the optical fibre access networks, where an energy-efficient approach to enhance VoD services over PONs is presented. In Chapter 3, we focus on the wireless access network systems, where a power control mechanism for heterogeneous mobile access networks is presented. In Chapter 4, we focus on the core network segment, where an energy-efficient approach to enhance survivability of the core network is presented. Finally, in this chapter, (Chapter 5) we conclude the thesis with a summary of contributions and a discussion on potential future research directions arising from our work.

Bibliography

[1] G. Kramer, B. Mukherjee, S. Dixit, Y. Ye, and R. Hirth, "Supporting differentiated classes of service in Ethernet passive optical networks," *J. Opt. Netw.*, vol. 1, pp. 280–298, aug 2002.

[2] S. Aleksic, G. Franzl, T. Bogner, and O. Tinkhof, "Energy-Efficient Internet Access," *2nd RESPONDER Workshop on Green ICT for Sustainable Consumption, Vienna*, 2013.

[3] C. Jayasundara, N. Nadarajah, A. Nirmalathas, and E. Wong, "Novel Bandwidth Efficient VoD Distribution Architecture for Broadband Access Networks," *Opto-Electronics and Communications Conference*, no. July, pp. 430–431, 2010.

[4] C. Jayasundara, A. Nirmalathas, E. Wong, and N. Nadarajah, "Popularity-Aware Caching Algorithm for Video-on-Demand Delivery over Broadband Access Networks," in *2010 IEEE Global Telecommunications Conference GLOBECOM*, dec 2010.

[5] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. R. Nakhai, J. Zhang, and L. Hanzo, "Green radio: radio techniques to enable energy-efficient wireless networks," *Communications Magazine, IEEE*, vol. 49, pp. 46–54, jun 2011.

[6] M. Deruyck, E. Tanghe, W. Joseph, and L. Martens, "Characterization and optimization of the power consumption in wireless access networks by taking daily traffic variations into account," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, p. 248, aug 2012.

[7] ITU Recommendation G.694.1, "Spectral grids for WDM applications," 2002.

[8] B. Mukherjee, *Optical WDM Networks*. Springer US, 1 ed., 2006.

[9] K. M. Sivalingam and S. Subramaniam, *Emerging Optical Network Technologies*. Springer US, 1 ed., 2005.

[10] DISCUS Consortium, "Deliverable D6.1: First Report on the Specification of the Metro / Core Node Architecture," 2013.

[11] R. J. Sanferrare, "Terrestrial lightwave systems," *AT T Technical Journal*, vol. 66, pp. 95–107, jan 1987.

[12] D. Gloge, A. Albanese, C. A. Burrus, E. L. Chinnock, J. A. Copeland, A. G. Dentai, T. P. Lee, T. Li, and K. Ogawa, "High-speed digital lightwave communication using LEDs and PIN photodiodes at 1.3 um," *The Bell System Technical Journal*, vol. 59, pp. 1365–1382, oct 1980.

[13] J. I. Yamada, S. Machida, and T. Kimura, "2 Gbit/s optical transmission experiments at 1.3 m with 44 km single-mode fibre," *Electronics Letters*, vol. 17, pp. 479–480, jun 1981.

[14] T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate low-loss single-mode fibre at 1.55 um," *Electronics Letters*, vol. 15, pp. 106–108, feb 1979.

[15] A. H. Gnauck, B. Kasper, R. A. Linke, R. Dawson, T. L. Koch, T. J. Bridges, E. Burkhardt, R. Yen, D. P. Wilt, J. C. Campbell, K. Nelson, and L. G. Cohen, "4-Gbit/s transmission over 103 km of optical fiber using a novel electronic multiplexer/demultiplexer," *Journal of Lightwave Technology*, vol. 3, pp. 1032–1035, oct 1985.

[16] Y. Kobayashi, Y. Akatsu, K. Nakagawa, H. Kikuchi, and Y. Imai, "Compact 10-Gbit/s optical transmitter and receiver circuit packs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, pp. 1916–1922, aug 1995.

[17] N. S. Bergano, J. Aspell, C. R. Davidson, P. R. Trischitta, B. M. Nyman, and F. W. Kerfoot, "Bit error rate measurements of 14000 km 5 Gbit/s fibre-amplifier trans-

mission system using circulating loop," *Electronics Letters*, vol. 27, pp. 1889–1890, oct 1991.

[18] T. Otani, K. Goto, H. Abe, M. Tanaka, H. Yamamoto, and H. Wakabayashi, "5.3 Gbit/s 11300 km data transmission using actual submarine cables and repeaters," *Electronics Letters*, vol. 31, pp. 380–381, mar 1995.

[19] T. Welsh, R. Smith, H. Azami, and R. Chrisner, "The FLAG cable system," *Communications Magazine, IEEE*, vol. 34, pp. 30–35, feb 1996.

[20] W. C. Marra and J. Schesser, "Africa ONE: the Africa Optical Network," *Communications Magazine, IEEE*, vol. 34, pp. 50–57, feb 1996.

[21] K. Fukuchi, T. Kasamatsu, M. Morie, R. Ohhira, T. Ito, K. Sekiya, D. Ogasahara, and T. Ono, "10.92-Tb/s (273/spl times/40-Gb/s) triple-band/ultra-dense WDM optical-repeatered transmission experiment," in *Optical Fiber Communication Conference and Exhibit, 2001. OFC 2001*, vol. 4, pp. PD24–PD24, mar 2001.

[22] G. A. Thomas, B. I. Shraiman, P. F. Glodis, and M. J. Stephen, "Towards the clarity limit in optical fibre," *Nature*, vol. 404, no. 6775, p. 262, 2000.

[23] P. J. Winzer and R. Essiambre, "Advanced Optical Modulation Formats," *Proceedings of the IEEE*, vol. 94, pp. 952–985, may 2006.

[24] X. Zhou, J. Yu, M.-F. Huang, Y. Shao, T. Wang, L. Nelson, P. Magill, M. Birk, P. I. Borel, D. W. Peckham, R. Lingle, and B. Zhu, "64-Tb/s, 8 b/s/Hz, PDM-36QAM Transmission Over 320 km Using Both Pre- and Post-Transmission Digital Signal Processing," *Journal of Lightwave Technology*, vol. 29, pp. 571–577, feb 2011.

[25] D. Qian, M.-F. Huang, E. Ip, Y.-K. Huang, Y. Shao, J. Hu, and T. Wang, "High Capacity/Spectral Efficiency 101.7-Tb/s WDM Transmission Using PDM-128QAM-OFDM Over 165-km SSMF Within C- and L-Bands," *Journal of Lightwave Technology*, vol. 30, pp. 1540–1548, may 2012.

[26] U.S. Energy Information Administration, "Yearly Energy Review - March 2015," 2015.

[27] "Climate Change Information Kit," *UNOE and UNFCCC*, 2002.

[28] T. climate group, "SMART 2020: Enabling the low carbon economy in the information age," 2008.

[29] Y. Lan and H. Thomas, "A review of research on the environmental impact of e-business and ICT," *Environment International*, vol. 33, pp. 841–849, 2007.

[30] S. Aleksic, "Green ICT for sustainability: A holistic approach," in *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014*, pp. 426–431, may 2014.

[31] Cisco Visual Networking Index, "Global Mobile Data Traffic Forecast Update 20142019," 2015.

[32] The Central Intelligence Agency, "The World Factbook,"

[33] T. climate group, "SMART 2020: The Role of ICT in Driving a Sustainable Future," 2012.

[34] Cisco Visual Networking Index, "Forecasts and Methodology, 2013-2018," 2014.

[35] Cisco Visual Networking Index, "The Zettabyte Era, Trends and Analysis," 2015.

[36] ITU-T Recommendation G.1010, "End-user multimedia QoS categories," 2001.

[37] ITU-T Recommendation J.241, "J.241 (04/2005)," 2005.

[38] P. Dymarski, S. Kula, and T. N. Huy, "QoS Conditions for VoIP and VoD," *Journal of Telecommunications and Information Technology*, pp. 29–37, 2011.

[39] D. De Vleeschauwer and K. Laevens, "Performance of Caching Algorithms for IPTV On-Demand Services," *IEEE Transactions on Broadcasting*, vol. 55, pp. 491–501, jun 2009.

[40] J. Baliga, R. Ayre, K. Hinton, and R. S. Tucker, "Architectures for Energy-Efficient IPTV Networks," in *IEEE/OSA Optical Fibre Communications Conference*, 2009.

[41] C. Jayasundara, A. Nirmalathas, E. Wong, and C. A. Chan, "Energy Efficient Content Distribution for VoD Services," in *IEEE/OSA Optical Fibre Communications Conference*, 2011.

[42] C. Jayasundara, A. Nirmalathas, E. Wong, and C. A. Chan, "Improving Energy Efficiency of Video on Demand Services," *Journal of Optical Communications and Networking*, vol. 3, p. 870, oct 2011.

[43] T. Koonen, "Fiber to the Home/Fiber to the Premises: What, Where, and When?," *Proceedings of the IEEE*, vol. 94, pp. 911–934, may 2006.

[44] C. H. Lee, W. V. Sorin, and B. Y. Kim, "Fiber to the Home Using a PON Infrastructure," *Lightwave Technology, Journal of*, vol. 24, pp. 4568–4583, dec 2006.

[45] E. Harstead, "Future bandwidth demand favors TDM PON, not WDM PON," in *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, pp. 1–3, mar 2011.

[46] IEEE, "IEEE 802.3ah: Ethernet in the first mile task force," 2004.

[47] ITU-T Recommendation G.984.1, "Gigabit-capable passive optical networks (GPON): General characteristics," 2008.

[48] IEEE, "IEEE 802.3av: 10G-EPON task force," 2009.

[49] ITU-T Recommendation G.987, "10Gigabit-capable passive optical networks (XG-PON)."

[50] M. P. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms," *Communications Magazine, IEEE*, vol. 42, pp. S8–15, aug 2004.

[51] M. P. McGarry, M. Reisslein, and M. Maier, "Ethernet passive optical network architectures and dynamic bandwidth allocation algorithms," *Communications Surveys Tutorials, IEEE*, vol. 10, no. 3, pp. 46–60, 2008.

[52] B. Skubic, J. Chen, J. Ahmed, L. Wosinska, and B. Mukherjee, "A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON," *Communications Magazine, IEEE*, vol. 47, pp. S40–S48, mar 2009.

[53] J. Zheng and H. T. Mouftah, "A survey of dynamic bandwidth allocation algorithms for Ethernet Passive Optical Networks," *Optical Switching and Networking*, vol. 6, pp. 151–162, jul 2009.

[54] G. Kramer, B. Mukherjee, and G. Pesavento, "Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network," *Photonic Network Communications*, vol. 4, no. 1, pp. 89–107, 2002.

[55] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT a dynamic protocol for an Ethernet PON (EPON)," *Communications Magazine, IEEE*, vol. 40, pp. 74–80, feb 2002.

[56] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for quality-of-service over Ethernet PONs," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 1467–1477, nov 2003.

[57] H.-J. Byun, J.-M. Nho, and J.-T. Lim, "Dynamic bandwidth allocation algorithm in Ethernet passive optical networks," *Electronics Letters*, vol. 39, pp. 1001–1002, jun 2003.

[58] Y. Luo and N. Ansari, "Limited sharing with traffic prediction for dynamic bandwidth allocation and QoS provisioning over Ethernet passive optical networks," *J. Opt. Netw.*, vol. 4, pp. 561–572, sep 2005.

[59] I.-S. Hwang, Z.-D. Shyu, L.-Y. Ke, and C.-C. Chang, "A novel early DBA mechanism with prediction-based fair excessive bandwidth allocation scheme in EPON," *Computer Communications*, vol. 31, pp. 1814–1823, jun 2008.

[60] M. Ma, Y. Zhu, and T. H. Cheng, "A bandwidth guaranteed polling MAC protocol for Ethernet passive optical networks," in *INFOCOM 2003. Twenty-Second An-*

nual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 1, pp. 22–31 vol.1, mar 2003.

[61] Y. Luo and N. Ansari, "Bandwidth allocation for multiservice access on EPONs," *Communications Magazine, IEEE*, vol. 43, pp. S16–S21, feb 2005.

[62] Y. Wang, Z.-L. Zhang, D. H. C. Du, and D. Su, "A network-conscious approach to end-to-end video delivery over wide area networks using proxy servers," in *INFOCOM '98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 660–667 vol.2, mar 1998.

[63] F. Thouin and M. Coates, "Video-on-Demand Networks: Design Approaches and Future Challenges," *Network, IEEE*, vol. 21, pp. 42–48, mar 2007.

[64] J. Nussbaumer, B. V. Patel, F. Schaffa, and J. P. G. Sterbenz, "Networking requirements for interactive video on demand," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 779–787, jun 1995.

[65] J. Segarra and V. Cholvi, "Placement of storage capacity in distributed video servers," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 4, pp. 2537–2541 vol.4, 2002.

[66] L. B. Sofman and B. Krogfoss, "Analytical Model for Hierarchical Cache Optimization in IPTV Network," *Broadcasting, IEEE Transactions on*, vol. 55, pp. 62–70, mar 2009.

[67] B. Krogfoss, L. Sofman, and A. Agrawal, "Caching architectures and optimization strategies for IPTV networks," *Bell Labs Technical Journal*, vol. 13, no. 3, pp. 13–28, 2008.

[68] B. Krogfoss, L. B. Sofman, and A. Agrawal, "Hierarchical cache optimization in IPTV networks," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2009*, pp. 1–10, may 2009.

[69] S. Sen, J. Rexford, and D. Towsley, "Proxy prefix caching for multimedia streams," in *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1310–1319 vol.3, mar 1999.

[70] K. L. Wu, P. S. Yu, and J. L. Wolf, "Segment-based Proxy Caching of Multimedia Streams," in *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, (New York, NY, USA), pp. 36–44, ACM, 2001.

[71] S. Chen, S. Wee, B. Shen, and X. Zhang, "Adaptive and Lazy Segmentation Based Proxy Caching for Streaming Media Delivery," in *IN PROCEEDINGS OF ACM NOSSDAV*, pp. 22–31, 2003.

[72] K. Suh, C. Diot, J. Kurose, L. Massoulie, C. Neumann, D. Towsley, and M. Varvello, "Push-to-Peer Video-on-Demand System: Design and Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 1706–1716, dec 2007.

[73] Y. Huang, T. Z. J. Fu, D.-M. Chiu, J. C. S. Lui, and C. Huang, "Challenges, Design and Analysis of a Large-scale P2P-vod System," *SIGCOMM Comput. Commun. Rev.*, vol. 38, pp. 375–388, aug 2008.

[74] S. Annapureddy, S. Guha, C. Gkantsidis, D. Gunawardena, and P. R. Rodriguez, "Is High-quality Vod Feasible Using P2P Swarming?," in *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, (New York, NY, USA), pp. 903–912, ACM, 2007.

[75] Amazon Web Services, "aws.amazon.com."

[76] Windows Azure: Microsoft's Could Platform, "www.windowsazure.com."

[77] Techblog.netflix.com, "Four reasons we choose amazon's cloud as our computing platform."

[78] H. Li, L. Zhong, J. Liu, B. Li, and K. Xu, "Cost-Effective Partial Migration of VoD Services to Content Clouds," in *IEEE International Conference on Cloud Computing (CLOUD), 2011*, pp. 203–210, jul 2011.

[79] Y. Wu, C. Wu, B. Li, X. Qiu, and F. C. M. Lau, "CloudMedia: When Cloud on Demand Meets Video on Demand," in *31st International Conference on Distributed Computing Systems (ICDCS)*, pp. 268–277, jun 2011.

[80] V. Aggarwal, X. Chen, V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, and V. A. Vaishampayan, "Exploiting virtualization for delivering cloud-based IPTV services," in *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 637–641, apr 2011.

[81] V. Aggarwal, V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, and V. A. Vaishampayan, "Optimizing Cloud Resources for Delivering IPTV Services Through Virtualization," *IEEE Transactions on Multimedia*, vol. 15, pp. 789–801, jun 2013.

[82] H. Ma and K. G. Shin, "Multicast Video-on-Demand Services," *SIGCOMM Comput. Commun. Rev.*, vol. 32, pp. 31–43, jan 2002.

[83] J. B. Kwon and H. Y. Yeom, "Providing VCR functionality in staggered video broadcasting," *IEEE Transactions on Consumer Electronics*, vol. 48, pp. 41–48, feb 2002.

[84] Y. Guo, S. Sen, and D. Towsley, "Prefix caching assisted periodic broadcast for streaming popular videos," in *IEEE International Conference on Communications*, vol. 4, pp. 2607–2612 vol.4, 2002.

[85] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the Internet with Nano Data Centers," in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, (New York, NY, USA), pp. 37–48, ACM, 2009.

[86] F. Liu, B. Li, B. Li, and H. Jin, "Peer-Assisted On-Demand Streaming: Characterizing Demands and Optimizing Supplies," *IEEE Transactions on Computers*, vol. 62, pp. 351–361, feb 2013.

[87] D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, and K. K. Ramakrishnan, "Optimal Content Placement for a Large-scale VoD System," in *Proceedings of the 6th*

International COnference, Co-NEXT '10, (New York, NY, USA), pp. 4:1—-4:12, ACM, 2010.

[88] R. Rejaie, M. H, H. Yu, and D. Estrin, "Proxy Caching Mechanism for Multimedia Playback Streams in the Internet," in *In Proceedings of the 4th International Web Caching Workshop*, 1999.

[89] S. Chen, B. Shen, S. Wee, and X. Zhang, "Designs of high quality streaming proxy systems," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1512–1521 vol.3, mar 2004.

[90] Z. L. Zhang, Y. Wang, D. H. C. Du, and D. Su, "Video staging: a proxy-server-based approach to end-to-end video delivery over wide-area networks," *IEEE/ACM Transactions on Networking*, vol. 8, pp. 429–442, aug 2000.

[91] W.-H. Ma and D. H. C. Du, "Reducing bandwidth requirement for delivering video over wide area networks with proxy server," in *IEEE International Conference on Multimedia and Expo, 2000.*, vol. 2, pp. 991–994 vol.2, 2000.

[92] Z. Miao and A. Ortega, "Scalable proxy caching of video under storage constraints," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1315–1327, sep 2002.

[93] S. Acharya and B. Smith, "Middleman: A video caching proxy server," in *Proceedings of NOSSDAV*, 2000.

[94] S. Borst, V. Gupta, and A. Walid, "Distributed Caching Algorithms for Content Distribution Networks," in *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9, mar 2010.

[95] J. Dai, Z. Hu, B. Li, J. Liu, and B. Li, "Collaborative hierarchical caching with dynamic request routing for massive content distribution," in *INFOCOM, 2012 Proceedings IEEE*, pp. 2444–2452, mar 2012.

[96] Y. D. Lin, H. Z. Lai, and Y. C. Lai, "A hierarchical network storage architecture for video-on-demand services," in *21st IEEE Conference on Local Computer Networks*, pp. 355–364, oct 1996.

[97] C. Vassilakis, M. Paterakis, and P. Triantafillou, "Video placement and configuration of distributed video servers on cable TV networks," *Multimedia Systems*, vol. 8, no. 2, pp. 92–104, 2000.

[98] R. H. Hwang and Y. C. Sun, "Optimal video placement for hierarchical video-on-demand systems," in *7th International Conference on Computer Communications and Networks*, pp. 454–461, oct 1998.

[99] R. H. Hwang and P. H. Chi, "Fast optimal video placement algorithms for hierarchical video-on-demand systems," *IEEE Transactions on Broadcasting*, vol. 47, pp. 357–366, dec 2001.

[100] N. Laoutaris, P. Rodriguez, and L. Massoulie, "ECHOS: Edge Capacity Hosting Overlays of Nano Data Centers," *SIGCOMM Comput. Commun. Rev.*, vol. 38, pp. 51–54, jan 2008.

[101] S. Podlipnig and L. Boszormenyi, "A Survey of Web Cache Replacement Strategies ," *ACM Computing Surveys*, vol. 35, no. 4, pp. 374–398, 2003.

[102] IEEE 802.3ah, "IEEE Standard for Information Technology: Telecommunications and Information exchange between systems: Local and metropolitan area networks: Specific Requirements. Part 3: Carrier sense Multiple Access with Collision Detection (CSMA/CD)," 2002.

[103] A. Sneh and K. M. Johnson, "High-speed Continuously Tunable Liquid Crystal Filter for WDM Networks," *Journal of Lightwave Technology*, vol. 14, no. 6, pp. 1067–1080, 1996.

[104] E. Wong, M. Mueller, P. I. Dias, C. A. Chan, and M. C. Amann, "Energy-Efficient Optical Network Units with Vertical-Cavity Surface-Emitting Lasers," *Optics express*, vol. 20, no. 14, pp. 14960 – 14970, 2012.

[105] E. Wong and K.-L. Lee, "Characterization of highly-sensitive and fast-responding monitoring module for extended-reach passive optical networks.," *Optics express*, vol. 20, pp. 9019–9030, apr 2012.

[106] Cisco Visual Networking Index, "Global Mobile Data Traffic Forecast Update 2015-2020," 2016.

[107] Verizon, "LTE: The Future of Mobile Broadband Technology. http://innovation.verizon.com."

[108] M. S. Sesia and I. Toufik, *LTE The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Ltd., 2009.

[109] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *Wireless Communications, IEEE*, vol. 17, pp. 10–22, jun 2010.

[110] 3GPP, "3GPP TS 23.107 Quality of Service (QoS) concept and architecture,"

[111] H. Ekstrom, "QoS control in the 3GPP evolved packet system," *Communications Magazine, IEEE*, vol. 47, pp. 76–83, feb 2009.

[112] K. Etemad, "Overview of mobile WiMAX technology and evolution," *Communications Magazine, IEEE*, vol. 46, pp. 31–40, oct 2008.

[113] S. Ahmadi, "An overview of next-generation mobile WiMAX technology," *Communications Magazine, IEEE*, vol. 47, pp. 84–98, jun 2009.

[114] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 26–38, jan 2007.

[115] B. Li, Y. Qin, C. P. Low, and C. L. Gwee, "A Survey on Mobile WiMAX [Wireless Broadband Access]," *Communications Magazine, IEEE*, vol. 45, pp. 70–75, dec 2007.

[116] C. So In, R. Jain, and A. K. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 156–171, feb 2009.

[117] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 2012.

[118] P. Henry and H. Luo, "WiFi: what's next?," *Communications Magazine, IEEE*, vol. 40, pp. 66–72, dec 2002.

[119] D. Niyato and E. Hossain, "WIRELESS BROADBAND ACCESS: WIMAX AND BEYOND - Integration of WiMAX and WiFi: Optimal Pricing for Bandwidth Sharing," *Communications Magazine, IEEE*, vol. 45, pp. 140–146, may 2007.

[120] N. Ghazisaidi, H. Kassaei, and M. Bohlooli, "Integration of WiFi and WiMAX mesh networks," in *IEEE Advances in Mesh Networks*, pp. 1–6, IEEE, 2009.

[121] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G Using WiFi," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, (New York, NY, USA), pp. 209–222, ACM, 2010.

[122] L. Korowajczuk, "LTE WiMAX and WLAN Network Design Optimization and Performance Analysis, Chapter 12," John Wiley & Sons Ltd., 2011.

[123] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter, "Wireless communications: past events and a future perspective," *Communications Magazine, IEEE*, vol. 40, pp. 148–161, may 2002.

[124] J. Wells, "Faster than fiber: The future of multi-G/s wireless," *Microwave Magazine, IEEE*, vol. 10, pp. 104–112, may 2009.

[125] T. Janevski, "5G Mobile Phone Concept," in *Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE*, pp. 1–2, jan 2009.

[126] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE transactions on Communications*, vol. 47, pp. 199–207, feb 1999.

[127] J. Chuang and N. Sollenberger, "Beyond 3G: wideband wireless data access based on OFDM and dynamic packet assignment," *Communications Magazine, IEEE*, vol. 38, pp. 78–87, jul 2000.

[128] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and A. Viterbi, "CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users," *Communications Magazine, IEEE*, vol. 38, pp. 70–77, jul 2000.

[129] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, pp. 2127–2159, sep 2006.

[130] J. Mitola and J. Maguire G.Q., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, pp. 13–18, aug 1999.

[131] J. Hoydis and M. Debbah, "Green, Cost-effective, Flexible, Small Cell Networks," *IEEE Communications Society MMTC*, vol. 5, pp. 1–3, 2010.

[132] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo, H. S. Dhillon, and T. D. Novlan, "Heterogeneous cellular networks: From theory to practice," *Communications Magazine, IEEE*, vol. 50, pp. 54–64, jun 2012.

[133] B. A. Bjerke, "LTE-advanced and the evolution of LTE deployments," *Wireless Communications, IEEE*, vol. 18, pp. 4–5, oct 2011.

[134] S.-p. Yeh, S. Talwar, G. Wu, N. Himayat, and K. Johnsson, "Capacity and coverage enhancement in heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, pp. 32–38, jun 2011.

[135] 3GPP, "New work item proposal: Enhanced ICIC for non-CA based deployments of heterogeneous networks for LTE," 2010.

[136] Z. Bharucha, E. Calvanese, J. Chen, X. Chu, A. Feki, A. D. Domenico, A. Galindo-Serrano, W. Guo, R. Kwan, J. Liu, D. López-Pérez, M. Maqbool, Y. Peng, S. Perlaza, G. de la Roche, S. Uygungelen, A. Valcarce, and J. Zhang, "Small Cell Deployments: Recent Advances and Research Challenges," *Cornell University Library: Networking and Internet Architecture*, pp. 1–19, 2012.

[137] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *Vehicular Technology Magazine, IEEE*, vol. 6, pp. 37–43, mar 2011.

[138] BellAir Networks, "Cell Site Backhaul over Unlicensed Bands," 2012.

[139] J. Hoadley and P. Maveddat, "Enabling small cell deployment with HetNet," *Wireless Communications, IEEE*, vol. 19, pp. 4–5, apr 2012.

[140] J. Robson, "Small cell deployment strategies and best practical backhaul," *Cambridge Broadband Networks, White Paper*, 2012.

[141] Small Cell Forum, "Backhaul Technologies for Small Cells, Use Cases, Requirements and Solutions," *White Paper*, 2013.

[142] S. Fili, "Wireless backhaul can ease transition to fibre," *White Paper*, 2012.

[143] Alcatel Lucent, "Wireline Mobile Backhaul for Metro Cells - Leveraging GPON and VDSL2 Fixed Broadband Access for Metro Cell Backhaul," *Application Note*, 2012.

[144] DISCUS Consortium, "GreenTouch."

[145] Int. Telecommun. Union, "ITU Green Standards Week," 2011.

[146] J. Wu, Y. Zhang, M. Zukerman, and E.-N. Yung, "Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 2, pp. 803–826, 2015.

[147] G. Fettweis and E. Zimmermann, "ICT energy consumption- Trends and challenges," *International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 2006–2009, 2008.

[148] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 4, pp. 524–540, 2011.

[149] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy Efficiency in Telecom Optical Networks," *Communications Surveys Tutorials, IEEE*, vol. 12, no. 4, pp. 441–458, 2010.

[150] A. P. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, "A Survey of Green Networking Research," *Communications Surveys Tutorials, IEEE*, vol. 14, no. 1, pp. 3–20, 2012.

[151] H. Claussen, L. T. W. Ho, and F. Pivit, "Effects of joint macrocell and residential picocell deployment on the network energy efficiency," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1–6, sep 2008.

[152] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Godor, G. Auer, and L. der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *Communications Magazine, IEEE*, vol. 48, pp. 66–72, nov 2010.

[153] J. Jeong, D. F. Kimball, M. Kwak, P. Draxler, C. Hsia, C. Steinbeiser, T. Landon, O. Krutko, L. E. Larson, and P. M. Asbeck, "High-Efficiency WCDMA Envelope Tracking Base-Station Amplifier Implemented With GaAs HVHBTs," *IEEE Journal of Solid-State Circuits*, vol. 44, pp. 2629–2639, oct 2009.

[154] K. J. Cho, J. H. Kim, and S. P. Stapleton, "A highly efficient Doherty feedforward linear power amplifier for W-CDMA base-station applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 292–300, jan 2005.

[155] Y. Wei, J. Staudinger, and M. Miller, "High efficiency linear GaAs MMIC amplifier for wireless base station and Femto cell applications," in *2012 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*, pp. 49–52, jan 2012.

[156] B. Badic, T. O'Farrrell, P. Loskot, and J. He, "Energy Efficient Radio Access Architectures for Green Radio: Large versus Small Cell Size Deployment," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, pp. 1–5, sep 2009.

[157] J. Wu, S. Zhou, and Z. Niu, "Traffic-Aware Base Station Sleeping Control and Power Matching for Energy-Delay Tradeoffs in Green Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 4196–4209, aug 2013.

[158] J. T. Louhi, "Energy efficiency of modern cellular base stations," in *Telecommunications Energy Conference, 2007. INTELEC 2007. 29th International*, pp. 475–476, sep 2007.

[159] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing Energy Consumption in LTE with Cell DTX," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp. 1–5, may 2011.

[160] I. Ashraf, F. Boccardi, and L. Ho, "SLEEP mode techniques for small cell deployments," *Communications Magazine, IEEE*, vol. 49, pp. 72–79, aug 2011.

[161] M. Ajmone Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "On the effectiveness of single and multiple base station sleep modes in cellular networks," *Computer Networks*, vol. 57, pp. 3276–3290, dec 2013.

[162] G. Micallef, P. Mogensen, and H.-O. Scheck, "Cell size breathing and possibilities to introduce cell sleep mode," in *Wireless Conference (EW), 2010 European*, pp. 111–115, apr 2010.

[163] M. Li, P. Li, X. Huang, Y. Fang, and S. Glisic, "Energy Consumption Optimization for Multihop Cognitive Cellular Networks," *Mobile Computing, IEEE Transactions on*, vol. 14, pp. 358–372, feb 2015.

[164] H. Elsawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 3, pp. 996–1019, 2013.

[165] A. Attar, H. Li, and V. C. M. Leung, "Green last mile: how fiber-connected massively distributed antenna systems can save energy," *Wireless Communications, IEEE*, vol. 18, pp. 66–74, oct 2011.

[166] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-Efficient Resource Allocation in OFDMA Networks," *Communications, IEEE Transactions on*, vol. 60, pp. 3767–3778, dec 2012.

[167] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1089–1098, aug 2004.

[168] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *Communications Magazine, IEEE*, vol. 49, pp. 30–37, jun 2011.

[169] G. He, S. Zhang, Y. Chen, and S. Xu, "Fundamental tradeoffs and evaluation methodology for future green wireless networks," in *1st IEEE International Conference on Communications in China Workshops (ICCC)*, pp. 74–78, aug 2012.

[170] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, pp. 59–67, sep 2008.

[171] W. Guo and T. O'Farrell, "Green cellular network: Deployment solutions, sensitivity and tradeoffs," in *Wireless Advanced (WiAd), 2011*, pp. 42–47, jun 2011.

[172] H. Claussen, I. Ashraf, and L. T. W. Ho, "Dynamic idle mode procedures for femtocells," *Bell Labs Technical Journal*, vol. 15, pp. 95–116, sep 2010.

[173] W. Li, W. Zheng, Y. Xie, and X. Wen, "Clustering based power saving algorithm for self-organized sleep mode in femtocell networks," in *15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 379–383, sep 2012.

[174] Y. K. Chia, S. Sun, and R. Zhang, "Energy Cooperation in Cellular Networks with Renewable Powered Base Stations," *IEEE Transactions onWireless Communications*, vol. 13, pp. 6996–7010, dec 2014.

[175] D. U. Ike, A. U. Adoghe, and A. Abdulkareem, "Analysis Of Telecom Base Stations Powered By Solar Energy," *International Journal of Scientific & Technology Research*, vol. 3, no. 4, pp. 369–374, 2014.

[176] 3GPP, "3GPP TS 32.521 Telecommunication management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); requirements," 2012.

[177] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "Toward self-organizing sectorization of LTE eNBs for energy efficient network operation under QoS constraints," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 1279–1284, apr 2013.

[178] R. Balasubramaniam, S. Nagaraj, M. Sarkar, C. Paolini, and P. Khaitan, "Cell Zooming for Power Efficient Base Station Operation," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pp. 556–560, jul 2013.

[179] S. Cai, L. Xiao, H. Yang, J. Wang, and S. Zhou, "A cross-layer optimization of the joint macro- and picocell deployment with sleep mode for green communications," in *Wireless and Optical Communication Conference (WOCC), 2013 22nd*, pp. 225–230, may 2013.

[180] P. Dini, M. Miozzo, N. Bui, and N. Baldo, "A Model to Analyze the Energy Savings of Base Station Sleep Mode in LTE HetNets," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, pp. 1375–1380, aug 2013.

[181] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy Efficient Heterogeneous Cellular Networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, pp. 840–850, may 2013.

[182] M. Wildemeersch, T. Q. S. Quek, C. H. Slump, and A. Rabbachin, "Cognitive Small Cell Networks: Energy Efficiency and Trade-Offs," *IEEE Transactions on Communications*, vol. 61, pp. 4016–4029, sep 2013.

[183] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, "Optimal Control of Wake Up Mechanisms of Femtocells in Heterogeneous Networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 664–672, apr 2012.

[184] D. Cao, S. Zhou, and Z. Niu, "Optimal Combination of Base Station Densities for Energy-Efficient Two-Tier Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 4350–4362, sep 2013.

[185] K. Huang and V. K. N. Lau, "Enabling Wireless Power Transfer in Cellular Networks: Architecture, Modeling and Deployment," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 902–912, feb 2014.

[186] T. Nylander, J. Rune, and J. Vikberg, "Energy Efficient Base Station Entering Sleep Mode," 2012.

[187] IEEE, "IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access," 2002.

[188] S. S. Prasad, C. K. Shukla, and R. F. Chisab, "Performance analysis of OFDMA in LTE," in *Third International Conference on Computing Communication Networking Technologies (ICCCNT)*, pp. 1–7, jul 2012.

[189] F. J. Gallego, "Mapping rural/ urban areas from population density grids," *White Paper*.

[190] G. Auer, O. Blume, V. Giannini, I. Godor, M. A. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, and W. Wajda, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *EARTH Project: Deliverable D2.3*, 2012.

[191] C. Shannon, "Communication in the Presence of Noise," *Classic paper, Proceedings of IEEE*, vol. 86, pp. 447–457, 1998.

[192] S. E. Nai, T. Q. S. Quek, M. Debbah, and A. Huang, "Slow admission and power control for small cell networks via distributed optimization," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 2261–2265, apr 2013.

[193] K. H. Liu, *IP over WDM*. New York: John Wiley and Sons, 2002.

[194] H. Anpeng, X. Linzhen, L. Zhengbin, and X. Ansbi, "Time-Space Label Switching Protocol (TSL-SP)A New Paradigm of Network Resource Assignment," *Photonic Network Communications*, vol. 6, no. 2, pp. 169–178, 2003.

[195] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath Communications: An Approach to High Bandwidth Optical WANs," *IEEE transactions on Communications*, vol. 40, no. 7, pp. 1171–1182, 1992.

[196] W. Goralski, *SONET*. New York, NY: McGraw-Hill, 2 ed., 2000.

[197] I. Chlamtac, A. Farago, and T. Zhang, "Lightpath (wavelength) routing in large WDM networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, pp. 909–913, jun 1996.

[198] R. Ramaswami and K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 489–500, oct 1995.

[199] C. Chen and S. Banerjee, "A new model for optimal routing and wavelength assignment in wavelength division multiplexed optical networks," in *INFOCOM '96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE*, vol. 1, pp. 164–171 vol.1, mar 1996.

[200] H. Zang and J. P. Jue, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks," *Optical Networks Magazine*, vol. 1, pp. 47–60, 2000.

[201] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM mesh networks," *Lightwave Technology, Journal of*, vol. 21, pp. 870–883, apr 2003.

[202] B. T. Doshi, S. Dravida, P. Harshavardhana, O. Hauser, and Y. Wang, "Optical network design and restoration," *Bell Labs Technical Journal*, vol. 4, pp. 58–84, jan 1999.

[203] G. Mohan, C. Siva Ram Murthy, and A. K. Somani, "Efficient algorithms for routing dependable connections in WDM optical networks," *IEEE/ACM Transactions on Networking*, vol. 9, pp. 553–566, oct 2001.

[204] G. Ellinas, E. Bouillet, R. Ramamurthy, J.-F. Labourdette, S. Chaudhuri, and K. Bala, "Routing and restoration architectures in mesh optical networks," *Optical Networks Magazine*, vol. 4, no. 1, pp. 91–106, 2003.

[205] A. Fumagalli, I. Cerutti, and M. Tacca, "Optimal design of survivable mesh networks based on line switched WDM self-healing rings," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 501–512, jun 2003.

[206] M. Goyal, G. Li, and J. Yates, "Shared mesh restoration: a simulation study," in *Optical Fiber Communication Conference and Exhibit, 2002. OFC 2002*, pp. 489–490, mar 2002.

[207] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. II. Restoration," in *IEEE International Conference on Communications*, vol. 3, pp. 2023–2030 vol.3, 1999.

[208] J. Wang, L. Sahasrabuddhe, and B. Mukherjee, "Fault monitoring and restoration in optical WDM networks," in *National Fiber Optic Engineers Conference*, 2002.

[209] J. M. Simmons, *Optical Network Design and Planning*. Springer Publishing Company, Incorporated, 1 ed., 2008.

[210] R. Ramamurthy, J. Labourdette, A. Akyamac, and S. Chaudhuri, "Limited sharing on protection channels in mesh optical networks," in *Optical Fiber Communications (OFC '03)*, 2003.

[211] C. Y. Lee and S. J. Koh, "A Disign of the Minimum Cost Ring-Chain Network with Dual-Homing Survivability: A Tabu Search Approch," *Computer Operations Research*, vol. 24, no. 9, pp. 883–897, 1997.

[212] M. Ruffini, L. Wosinska, M. Achouche, J. Chen, N. Doran, F. Farjady, J. Montalvo, P. Ossieur, B. O'Sullivan, N. Parsons, T. Pfeiffer, X. Qiu, C. Raack, H. Rohde, M. Schiano, P. Townsend, R. Wessaly, X. Yin, and D. B. Payne, "DISCUS: An End-to-End Solution for Ubiquitous Broadband Optical Access," *IEEE Communications Magazine*, vol. 52, no. February, pp. 24–32, 2014.

[213] A. Proestaki and M. C. Sinclair, "Design and dimensioning of dual-homing hierarchical multi-ring networks," *Communications, IEE Proceedings-*, vol. 147, pp. 96–104, apr 2000.

[214] J. J. Shi and J. P. Fonseka, "Analysis and design of survivable telecommunications networks," *Communications, IEE Proceedings-*, vol. 144, pp. 322–330, oct 1997.

[215] A. Akella, B. Maggs, S. Seshan, A. Shaikh, and R. Sitaraman, "A measurement-based analysis of multihoming," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 353–364, ACM, 2003.

[216] D.-R. Din and S. Tseng, "A genetic algorithm for solving dual-homing cell assignment problem of the two-level wireless ATM network," *Computer Communications*, vol. 25, pp. 1536–1547, nov 2002.

[217] DISCUS Consortium, "Deliverable D2.1: Report on the initial DISCUS End to End Architecture," 2013.

[218] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. Part I-Protection," in *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 744–751 vol.2, mar 1999.

[219] G. Li, D. Wang, C. Kalmanek, and R. Doverspike, "Efficient distributed path selection for shared restoration connections," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1, pp. 140–149, IEEE, 2002.

[220] G. Li, D. Wang, K. Charles, and R. Doverspike, "Efficient distributed restoration path selection for shared mesh restoration," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 761–771, oct 2003.

[221] P.-H. Ho and H. T. Mouftah, "A framework for service-guaranteed shared protection in WDM mesh networks," *Communications Magazine, IEEE*, vol. 40, no. 2, pp. 97–103, 2002.

[222] P.-h. Ho and H. T. Mouftah, "Shared protection in mesh WDM networks," *Communications Magazine, IEEE*, vol. 42, pp. 70–76, jan 2004.

[223] O. Gerstel and R. Ramaswami, "Optical layer survivability-an implementation perspective," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1885–1899, oct 2000.

[224] D. Zhou and S. S. Subramaniam, "Survivability in optical networks," *Network, IEEE*, vol. 14, pp. 16–23, nov 2000.

[225] O. Crochat, J.-Y. Le Boudec, and O. Gerstel, "Protection interoperability for WDM optical networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 8, no. 3, pp. 384–395, 2000.

[226] E. Limal, S. L. Danielsen, and K. E. Stubkjaer, "Capacity utilization in resilient wavelength-routed optical networks using link restoration," in *Optical Fiber Communication Conference and Exhibit, 1998. OFC '98., Technical Digest*, pp. 297–298, feb 1998.

[227] A. Fumagalli and L. Valcarenghi, "IP restoration vs. WDM protection: is there an optimal choice?," *Network, IEEE*, vol. 14, pp. 34–41, nov 2000.

[228] J. Wang, V. M. Vokkarane, X. Qi, and J. P. Jue, "Dual-Homing Protection in WDM Mesh Networks," in *IEEE/OSA Optical Fibre Communications Conference*, 2004.

[229] V. M. Vokkarane, J. Wang, X. Qi, R. Jothi, B. Raghavachari, and J. P. Jue, "Dual-Homing Protection in IP-Over-WDM Networks," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3111–3124, 2005.

[230] V. M. Vokkarane, J. Wang, and J. P. Jue, "Coordinated Survivability in IP-over-Optical Networks with IP-Layer Dual-Homing and Optical-Layer Protection," in *Broadband Networks*, pp. 259–268, 2005.

[231] J. Wang, M. Yang, B. Yang, and S. Q. Zheng, "Dual-Homing Based Scalable Partial Multicast Protection," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1130–1141, 2006.

[232] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic Matrix Estimation: Existing Techniques and New Directions," *SIGCOMM*, no. August, pp. 19–23, 2002.

[233] G. Shen and R. Tucker, "Energy-Minimized Design for IP Over WDM Networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, pp. 176–186, jun 2009.

[234] DISCUS Consortium, "Deliverable D7.1: Power Consumption Studies," pp. 1–52, 2014.

[235] CISCO, "Cisco Catalyst 4500 Series Switch Data Sheet,"

[236] J. Chugh, "Resilience, survivability and availability in WDM optical mesh network," in *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 222–227, mar 2015.

[237] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems," p. 49, jul 2010.

[238] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures," *IEEE Communications Surveys Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.

[239] Y. Mao, Y. Luo, J. Zhang, and K. B. Letaief, "Energy harvesting small cell networks: feasibility, deployment, and operation," *Communications Magazine, IEEE*, vol. 53, pp. 94–101, jun 2015.

Author/s:
Abeywickrama, Sandu

Title:
Design of energy-efficient next-generation telecommunications networks

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/115189

File Description:
Design of Energy-Efficient Next-Generation Telecommunications Networks