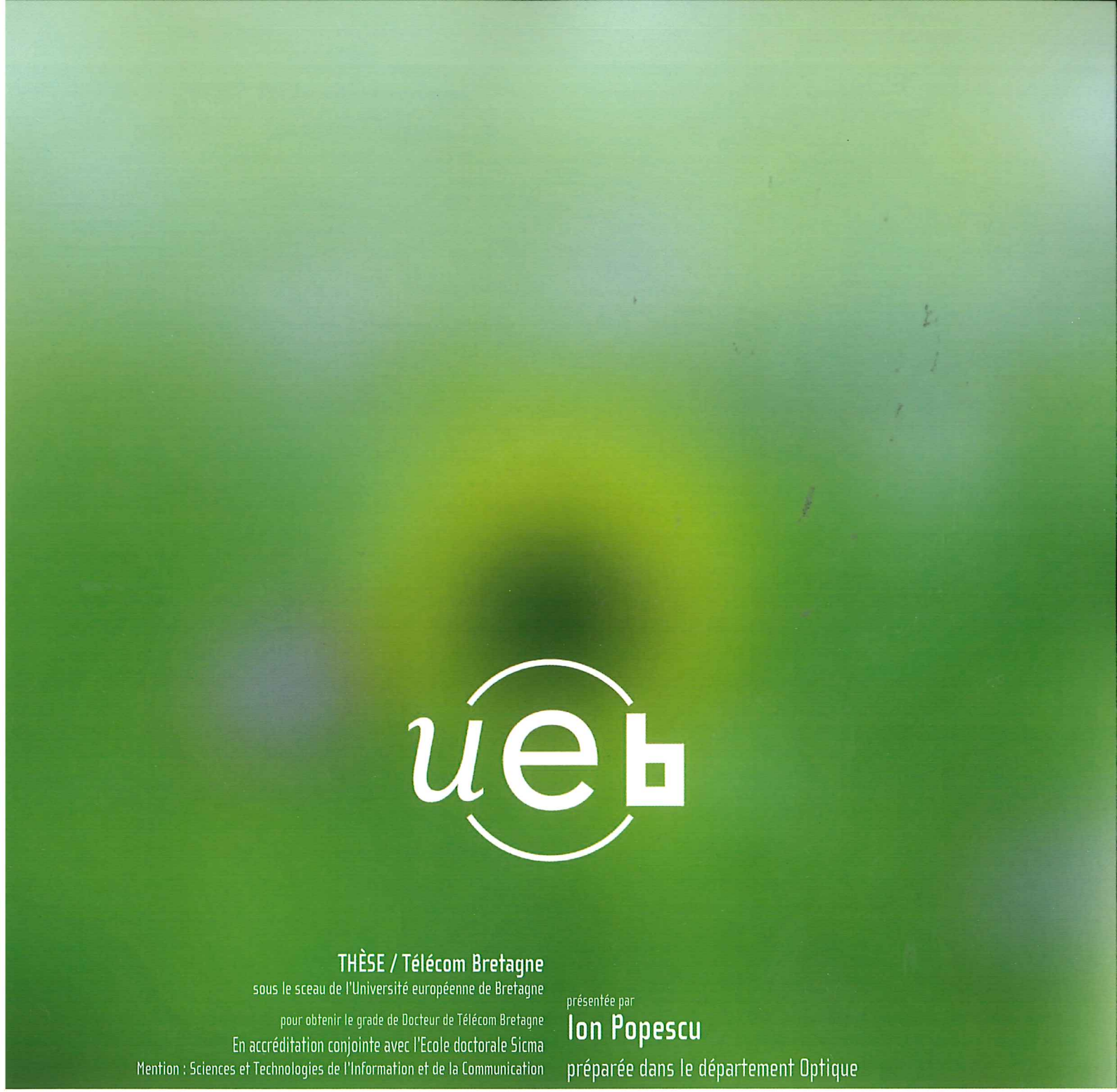
Evaluation of transparent optical multiplexing techniques in transport networks

Ion Popescu



THÈSE / Télécom Bretagne
sous le sceau de l'Université européenne de Bretagne

pour obtenir le grade de Docteur de Télécom Bretagne
En accréditation conjointe avec l'Ecole doctorale Sicma
Mention : Sciences et Technologies de l'Information et de la Communication

présentée par
Ion Popescu
préparée dans le département Optique

Evaluation of transparent optical multiplexing techniques in transport networks

Thèse soutenue le 9 septembre 2015
Devant le jury composé de :

Reinhardt Euler
Professeur, Université de Bretagne Occidentale / président

Isabella Cerutti
Maître de conférences, École Supérieure Sainte-Anne de Pise – Italie / rapporteur

Maurice Gagnaire
Professeur, Télécom ParisTech / rapporteur

Annie Gravey
Directrice d'études, Télécom Bretagne / examinatrice

Dominique Chiaroni
Ingénieur de recherche, Alcatel-Lucent Bell Labs - Nozay / examinateur

Esther Le Rouzic
Ingénieur, Orange Labs - Lannion / examinatrice

Naoaki Yamanaka
Professeur, Université de Keio – Japon / examinateur

Bruno Fracasso
Professeur, Télécom Bretagne / directeur de thèse

Philippe Gravey
Directeur d'études, Télécom Bretagne / invité

Yvan Pointurier
Ingénieur de recherche, Alcatel-Lucent Bell Labs - Nozay / invité

N° d'ordre : 2015telb0365

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En accréditation conjointe avec l'Ecole Doctorale SICMA

Evaluation of transparent optical multiplexing techniques in transport networks

---

Thèse de Doctorat

Mention : STIC (Sciences et Technologies de l'Information et de la Communication)

Présentée par **Ion Popescu**

Département : Optique

Directeur de thèse : Bruno Fracasso

Soutenue le 9 septembre 2015

Jury :

Rapporteurs

Mme Isabelle Cerutti	Maître de conférence, École Supérieure Sainte-Anne de Pise, Italie
M. Maurice Gagnaire	Professeur, Télécom ParisTech, France

Examinateurs

M. Bruno Fracasso	Professeur (Directeur de thèse), Télécom Bretagne, France
Mme Annie Gravey	Professeur (Encadrante), Télécom Bretagne, France
M. Dominique Chiaroni	Ingénieur de recherche, Alcatel-Lucent Bell Labs, France
M. Reinhardt Euler	Professeur, Université de Bretagne Occidentale, France
Mme Esther Le Rouzic	Ingénieur de recherche, Orange Labs, France
M. Naoaki Yamanaka	Professeur, Université de Keio, Japon

Invités

M. Philippe Gravey	Directeur d'études (Encadrant), Télécom Bretagne, France
M. Yvan Pointurier	Ingénieur de recherche, Alcatel-Lucent Bell Labs, France

Acknowledgements

At the beginning of this manuscript, I would like to express my gratitude to all people with whom I had collaborated, worked and inspired me during this period.

I would like to thank specially to the committee members which accepted the judge of this work and had the time to evaluate the manuscript. Especially, I would like to thank to the reporters, Dr. Isabelle Cerutti from Sant'Anna School of Advanced Studies and Prof. Maurice Gagnaire from Télécom ParisTech, Paris France, for giving detailed and valuable feedback for improvement on my work and being in touch with me for the last moths. Their constructive comments helped me to accomplish the document and realized a complex and well-organized academic manuscript.

Moreover, I would like to thank to my doctoral advisor Prof. Bruno Fracasso, which accepted me as PhD student and was always supporting me during the last tree years.

My special thanks go to my supervisor Prof. Annie Gravey which accompanied me during three years of my PhD studies. Her evaluations, analysis and remarks were extremely useful in terms of understanding the topics related to subwavelengths optical networks, as well as solving the heuristics and optimization problems, which represents the root solution in reducing the costs of today's network. Thanks to her, I managed to develop to the dimensioning tools required by the project in which I was involved. In addition, she motivated me to realize several academic papers. She helped me to achieve new milestones and develop my research and academic skills, by posing correct questions and solving them day by day.

I am extremely thankful to my supervisor Philippe Gravey who accompanied me during this thesis. His high level of professionalism and knowledge helped me with references, advices and explanations. He was always ready to give me proper guidelines, which oriented me in my research and helped to keep the deadlines. Our collaboration represented a fruitful result where I learned a lot from the technical side as well as from academic prospective.

Not least, I would like to thank Michel Morvan, he helped me and assisted me all the time when was required. I always received constructive feedback and support from him during the stay at Télécom Bretagne.

The PhD program was part for the Project Celtic-Plus - SASER-SaveNet which started in July 2012 and ends in 2015. Therefore, I would like to thank all partners and specially Alcatel-Lucent Bell Labs and Orange Lab Lannion. During the three years of the thesis, I had a positive collaboration with successful results and academic papers.

Special thank I would like to say to Dr. Bogdan Uscumlic and Dr. Ivan Pointurier from Alcatel-Lucent. We had collaborated continuously during this period realizing great results and fulfilling the project target. Definitely, it was an efficient collaboration from whom I learned, and developed new skills. Therefore, I am looking forward for future collaboration in the next research topics. Just at the beginning of my theses, I met Dr. Lida Sadeghioon which helped me to integrate into the system and obtain good results. Our collaboration represents a part of this document which is explained in chapter one.

I was a visitor student at Keio University, Tokyo in Yamanaka Lab during three months period in 2014 being part of an exceptional team and project. Therefore, I would like to thank Prof. Naoaki Yamanaka for accepting me in this team and giving me the chance to be part of the project ACTION. I had the chance to indeep my knowledge in new topics and realized good academic paper. This unique opportunity was a step-up in my career and enlarged my knowledge. In addition, I would like to thank all the Yamanaka team, especially to Prof. Satoru Okamoto and student Takahiro Miyazaki. Alos, I would like to thank Prof. Malathi Veeraraghavan from University of Virginia, U.S.A. with whom I had collaborated during that period. They gave me not only a motivation for new research topics, but also, developed my knowledge in energy efficient networks. I am looking forward for future collaboration and new achievements.

During the PhD studies, I had realized several papers submitted during the last three years. Hereby, I would like to thank all my co-authors, with whom we worked hardly to achieve the milestones and keep the deadlines. I would like to thank specially my colleague Dr. Ahmed Triki. It was a great pleasure to work day by day with them due to their professionalism, deep knowledge and hardworking attitude. Least by not last, I would like to thank all my friends and family which supported me and encourage me to set new goals and achieve them.

In Brest, May 2015.

Ion Popescu

In love and gratitude to my family.

Résumé

LE trafic est en augmentation constante dans tous les segments du réseau de transport en raison de l'accroissement du nombre de terminaux connectés (Smartphone, tablettes, téléviseurs connectés) et de la diffusion de contenus en haute définition. Cette croissance en trafic devrait se poursuivre [35], et conduit à une utilisation importante de la bande passante. L'approche conventionnelle dans les réseaux de transport opaques consiste à agréger le trafic dans le domaine temporel à chaque nœud, afin d'optimiser l'utilisation de la bande passante au détriment de la consommation énergique et de la mise en œuvre de fonctionnalités électroniques coûteuses.

D'autre part, l'agrégation transparente optique utilise le multiplexage par répartition en longueur d'onde ou "Wavelength Division Multiplexing" (WDM) et des multiplexeurs optiques d'insertion-extraction reconfigurables ou "Reconfigurable Optical Add Drop Multiplexer" (ROADM). Les besoins en matière de conversions optique à électrique (OE) et électrique à optique (EO) sont évités au niveau des nœuds intermédiaires, mais ce mode d'agrégation conduit à une utilisation non optimale des longueurs d'onde. De plus, ce mode de commutation de circuits optiques ou "Optical Circuit Switching" (OCS) est considéré comme n'étant pas suffisamment adapté au trafic aujourd'hui dominé par les paquets.

Certains moyens plus sophistiqués tentent de combiner les avantages des deux approches précédentes en combinant le multiplexage temporel ou "Time Division Multiplexing" (TDM) et le multiplexage par répartition en longueur d'onde ou "Wavelength Division Multiplexing" (WDM), afin de fournir une granularité en sous-longueurs d'onde et éviter le traitement électronique aux nœuds intermédiaires. La commutation de rafales optiques ou "Optical Burst Switching" (OBS) [96], [21] est une technique de commutation en sous-longueurs d'onde dans le domaine temporel. Depuis l'apparition du concept de commutation de rafales optiques, plusieurs solutions et mécanismes ont été proposés [6], [22], [12]. TWIN (Time-Domain Wavelength Interleaved Networking) est l'une de ces solutions [103], [76]. TWIN est considérée comme une solution plutôt attractive, car elle combine la transparence optique passive et l'utilisation efficace des longueurs d'onde.

TWIN permet ainsi de transmettre des rafales optique sur les canaux WDM, sans recourir à un traitement électronique dans nœuds de transit, les fonctions électroniques de traitement étant localisées uniquement aux nœuds source et destination.

Dans un réseau TWIN, chaque longueur d'onde est dédiée au transport des rafales de données vers un unique nœud destinataire. La transmission des rafales optiques est effectuée selon une structure d'arbre associé à chaque longueur d'onde et dont le nœud destinataire représente la racine et les nœuds source représentent les feuilles. Par

conséquent, la topologie logique de TWIN peut être considérée comme un ensemble d'arbres multipoint à point optiques superposées et construits sur un réseau maillé.

Le nœud source TWIN comprend un (ou plusieurs) transmetteur(s) laser accordable(s) lui permettant d'envoyer les rafales optiques aux différentes destinations, tandis que le nœud destinataire dispose d'un récepteur fixe recevant les rafales sur la longueur d'onde qui lui est attribuée. Les nœuds intermédiaires entre la source et la destination agrègent optiquement les rafales optiques provenant des feuilles et les aiguillent vers la racine sans aucun traitement électronique : TWIN est basé sur un routage en longueur d'onde, passif et transparent au niveau des nœuds intermédiaire. Comme les feuilles partagent la même longueur d'onde vers la destination, des collisions entre les rafales optiques peuvent survenir et doivent être évitées grâce au plan de commande du réseau. Ainsi, la simplicité photonique des nœuds TWIN impose de recourir à un plan de commande dont le rôle principal est d'éviter les collisions entre rafales optiques destinées à un même nœud destination, tout en permettant à chaque source d'utiliser son, ou ses transmetteurs laser efficacement.

Donc, en réglant son émetteur à la longueur d'onde appropriée, une source peut envoyer une rafale vers une destination. Un nœud sait quand envoyer et recevoir une rafale grâce à l'attribution de créneaux horaires dans un "échéancier" (allocation des ressources) qui est calculé par un plan de commande centralisé [84] ou distribué [92]. L'échéancier peut être envoyé périodiquement à chaque nœud au moyen d'un canal de commande. Le canal de commande peut être mis en œuvre avec un équipement de faible débit (par exemple, 1 Gb/s) afin d'atténuer son coût.

Le calcul d'un échéancier pour TWIN a été largement étudié par différents groupes de recherche, qui ont en général mis l'accent sur la réduction de sa longueur, sur le plan de commande centralisé ou distribué, et en tenant compte de différentes dynamiques de trafic (voir [76], [84], [90]). La longueur de l'échéancier est comprise entre quelques dizaines et quelques centaines de slots. L'échéancier est utilisé de manière répétitive. Il pourrait être recalculé toutes les quelques millisecondes, secondes, minutes ou heures, selon les besoins, et en fonction de la variation du trafic ou des changements de configuration du réseau.

Dans [76], une méthode définissant la longueur de l'échéancier qui satisfait toutes les demandes de débit est proposé. Un plan de commande distribué est considéré à la place d'un plan centralisé proposé dans [84], pour du trafic dynamique (l'échéancier devrait être recalculé toutes les quelques millisecondes) et le trafic semi-dynamique (l'échéancier devrait être recalculé toutes les quelques minutes ou heures). Tandis que dans [92] une étude comparative entre plans de commande centralisé et distribué est proposée.

Dans [91] et [90], les auteurs prouvent la haute performance d'un échéancier statique pour TWIN, qui est basé sur le calcul d'un échéancier optimisé en utilisant un modèle de optimisation linéaire en nombres entiers ou "Integer Linear Programming" (ILP). Le problème du calcul d'un échéancier optimisé est un problème NP complet [91].

Une méthode heuristique pour calculer un échéancier en TWIN est présentée dans [76], la méthode est appelée TIIS (TWIN iterative independent set), et cet algorithme est utilisé pour une configuration de réseau fixe qui ne prend pas en compte le dimensionnement du réseau.

L'échéancier statique est utilisé pendant une longue période de temps, contrairement à un échéancier dynamique qui est changé en fonction de la variation du trafic. Dans [90], les auteurs démontrent la possibilité de déployer un tel échéancier statique dans un réseau réel et la capacité d'un tel échéancier à répondre aux exigences de qualité de service.

Les ILP utilisés dans [91] et [90] ne considèrent que la question du calcul de l'échéancier et ne prennent pas en compte le problème de RWA (Routing and Wavelength Assignment). Cependant, l'échéancier est dépendant de temps de propagation entre les nœuds des couples source-destination. Par conséquent, il semble intéressant d'intégrer la résolution du problème classique de RWA à celui du calcul de l'échéancier dans TWIN.

Résoudre de manière optimale les problèmes de planification dans TWIN ne peut être réalisé en temps polynomial, donc les outils de programmation linéaires de dimensionnement proposés dans les travaux précédents sont limités à de petits exemples et ne peuvent pas être utilisés pour les réseaux de taille opérationnelle. Par conséquent, les outils de dimensionnement qui fournissent une solution flexible pour adresser simultanément RWA et optimisation de l'échéancier sont nécessaires.

De plus, la protection d'un réseau TWIN est fondamentale, et doit être prise en compte lors de la planification d'un tel réseau. En cas de faute comme une coupure de liaison, une partie du trafic sera déviée sur d'autres chemins pré-calculés, ce qui implique une réservation de bande passante supplémentaire. Cette bande passante doit être disponible à tout moment, mais sa quantité dépendra du type de protection utilisé : protection partagée (la même ressource peut être utilisée pour la protection de plusieurs arbres TWIN) ou protection dédiée (les ressources sont attribuées à chaque connexion et sont pas partagées).

La protection en TWIN n'a été que partiellement abordée jusqu'ici. Dans [103], la possibilité d'utiliser plusieurs arbres pour fournir la protection est considérée. Dans le présent document, le processus de calcul de l'échéancier est basé sur un algorithme heuristique : le problème de calcul des arbres pour TWIN est résolu préalablement et le coût de la protection n'est pas considéré. Dans [8], les auteurs proposent des méthodes de protection possibles pour une variation de TWIN qui utilise la technologie cohérente, mais les coûts de protection ne sont pas numériquement évalués.

Donc, une formulation de dimensionnement combiné pour TWIN est nécessaire, pouvant traiter simultanément le problème RWA, et calculer un échéancier optimal ainsi que les chemins pour la protection dédiée et partagée (les collisions de slots doivent être évitées entre chemins primaires et chemins de protection). Cet outil de dimensionnement peut être utile pour les planificateurs de réseau, si les fibres du réseau sont déjà installés, ou dans le cas de ce qui est appelé "green field" en anglais lorsqu'il est nécessaire d'optimiser les chemins dans le réseau.

Cette thèse est consacrée au dimensionnement des réseaux TWIN et à la comparaison avec la technologie POADM ou “Packet Optical Add Drop Multiplexer” dans différents scénarios. Ces technologies offrent toutes deux beaucoup plus de souplesse que les présentes technologies basées sur OCS.

La contribution de cette thèse fait partie du projet européen Saser-SaveNet. Ce projet est consacré à l'évaluation de différentes technologies et architectures à base de paquets optiques. Le projet Saser-SaveNet a débuté le 1er Juillet 2012 et son leader en France est Alcatel-Lucent Bell Labs. Les autres participants sont Télécom Bretagne, France Télécom - Orange Labs, Télécom SudParis, Univ. Rennes 1 et autres.

Le travail dans cette thèse est présenté dans les chapitres suivants :

- Le Chapitre 1 fournit un aperçu détaillé sur les réseaux de commutation optique, en mettant l'accent sur les technologies de commutation sans perte de paquet en sous-longueur d'onde optiques. La transmission sans perte signifie qu'il n'y a pas la perte de paquets, ce qui est rendu possible par l'utilisation d'un contrôle des émetteurs qui évitent la collision de données, et impose une synchronisation précise du réseau.

  Le chapitre explique brièvement les problèmes de dimensionnement dans les réseaux optiques, et les travaux existants sur TWIN. Ce chapitre résume également les travaux antérieurs sur le sujet réalisés à Télécom Bretagne. Le chapitre conclue par la motivation de thèse.

- Le Chapitre 2 propose un plan de contrôle pour le réseau TWIN basé sur un anneau virtuel qui joint toutes les sources et les destinations, afin de supporter un canal de commande hors bande. La configuration du réseau est obtenue en utilisant le canal de contrôle, en particulier les messages liés à la découverte du réseau, la synchronisation et les problèmes de protection sont porter sur le anneau virtuel.

- Le Chapitre 3 traite les problèmes de dimensionnement en TWIN. Dans ce chapitre, deux moyens d'améliorer l'efficacité du délai d'insertion des paquets sont proposés : en minimisant la longueur de l'échéancier et en appliquant une contrainte d'espacement entre slots attribués à un couple (source, destination). Les résultats détaillés des simulations sont présentés, montrant comment la topologie du réseau, le type de trafic et la contrainte d'espacement impactent la performance du réseau.

  Pour la première fois, un outil de dimensionnement ILP est proposé, qui traite simultanément les problèmes de RWA et le calcul de l'échéancier dans TWIN. En plus, un algorithme est développé pour le même problème et qui considère également a virtualisation de réseau. L'algorithme permet la construction de domaines de réseaux virtuels superposée qui partagent les transpondeurs et les liens de réseau ce qui permet de minimiser le coût de réseau, tout en autorisant l'utilisation de différentes politiques d'échéancier. La performance de l'algorithme est comparée avec la solution ILP de dimensionnement pour un seul domaine virtuel, et dans différents scénarios. En outre, l'étude porte sur un réseau de taille réelle.

- Dans le Chapitre 4, les coûts des différentes solutions à base de paquets optiques sont comparés. Deux scénarios sont considérés : les réseaux métropolitains protégés et les réseaux LTE (Long Term Evolution) backhauling, avec l'accent mis sur les technologies TWIN et POADM. La commutation dans les deux technologies est en sous-longueur d'onde : dans TWIN sur une topologie maillée et dans POADM sur une topologie en anneau. Le modèle tient en compte les coûts pour les composants de nœuds, le coût de transpondeur et le coût de location de longueur d'onde par km de fibre.

  Un modèle de dimensionnement de réseau unique, lesquels supporte la protection et le trafic de type multicast est proposé pour la première fois pour ces technologies. La solution de planification du réseau est basée sur la formulation ILP, et aborde la RWA et le calcul de l'échéancier optimal.

- Une conclusion récapitule les différentes contributions de la thèse et présente des perspectives pour la suite des travaux.

Abstract

This work addresses the problem of dimensioning in Time-Domain Wavelength Interleaved Network (TWIN). The study is done in the frame of the CELTIC-Plus project SASER-SaveNet, that is still evolving until the end of this year.

TWIN is a sub-wavelength switching technique that transmits bursts or slots of data on Wavelength Division Multiplexing (WDM) channels without resorting to electronic processing in transit nodes, the electronic processing functions being pushed to its edge source and destination nodes.

Through this thesis, two ways of improving the slot allocation efficiency in the TWIN network are proposed: by minimizing the schedule length and by enforcing a cyclical Quality of Service (QoS) constraint on the allocated slots. We report detailed simulation results on how network topology, traffic pattern and the proposed QoS improving method impact the network performance.

A novel 0-1 Integer Linear Programming (0-1 ILP) dimensioning tool is developed, which simultaneously addresses the Routing and Wavelength Assignment (RWA) and scheduling (resource allocation) problem in TWIN.

Furthermore, a comprehensive scalable algorithm is proposed in order to assess the same problem of cost dimensioning and, in addition, to consider network virtualization. The algorithm enables the construction of overlaid virtual networking domains at minimum cost, that share the transponders and network links, and can employ different scheduling policies.

Lastly, we compare the cost of the optical transport solutions with sub-wavelength switching granularity. Two scenarios are considered: survivable metro networks and Long Term Evolution (LTE) mobile backhauling, with the focus on TWIN and Packet Optical Add Drop Multiplexer (POADM) technologies. Both technologies apply optical sub-wavelength switching: TWIN on a physical mesh topology, and POADM on a physical ring topology. The cost model accounts for the node components, the transponder cost and the wavelength cost. Thus, a single network dimensioning model, supporting protection and multicast traffic, is proposed for the first time for these technologies.

Contents

Acronyms

0-1 ILP 0-1 Integer Linear Programming

ALR Adaptive Link Rate

BER Bit Error Rate

BFS Breadth-First Search

CAPEX Capital Expenditures

CoMP Coordinated MultiPoint

DAVID Data And Voice Integration on Dense WDM

EO Electrical to Optical

HMVTD Heuristic Method for Virtual TWIN Dimensioning

ILP Integer Linear Programming

ITU-T International Telecommunication Union - Telecommunication

LTE Long Term Evolution

MMF Multi Modulation Formats

MMF Multi Line Rates

MSL Minimum Schedule Length

MST Minimum Spanning Tree

OAM Operations, Administration, and Maintenance

OBS Optical Burst Switching

OCS Optical Circuit Switching

OE Optical to Electrical

OEO Optical-Electrical-Optical

OPEX Operational Expenditures

OPS Optical Packet Switching

POADM Packet Optical Add Drop Multiplexer

PON Passive Optical Network

QoS Quality of Service

QoT Quality of Transmission

RINGO Ring Optical Network

ROADM Reconfigurable Optical Add Drop Multiplexer

RPR Resilient Packet Ring

RWA Routing and Wavelength Assignment

SLPSN Sub-Lambda Photonically Switched Networks

SOA Semiconductor Optical Amplifier

TDM Time Division Multiplexing

TSON Time Shared Optical Network

TWIN Time-Domain Wavelength Interleaved Network

VTR Virtual TWIN Ring

WDM Wavelength Division Multiplexing

WSS Wavelength Selective Switch

List of Figures

List of Tables

General Introduction

TRAFFIC in all segments of the transport network has been constantly growing in the past years as a result of increasing number of connected end user terminals (smartphones, tablets, connected TVs) and the race towards higher definition contents. This growth is expected to continue [35], and it leads to significant bandwidth utilization. The conventional approach in opaque transport networks consists in aggregating/grooming traffic in the time domain at every node, in order to optimize bandwidth use at the expense of costly traffic grooming electronics and high power consumption. Thus, this method offers the maximum flexibility and control. On the other side, optical (i.e., transparent) aggregation uses Wavelength Division Multiplexing (WDM) and Reconfigurable Optical Add Drop Multiplexer (ROADM). In this case, many Optical to Electrical (OE) and Electrical to Optical (EO) conversions are avoided at intermediate nodes, but it induces a non-optimum channel use. Furthermore, this Optical Circuit Switching (OCS) method is considered not flexible enough for packet oriented traffic.

There are more sophisticated ways that try to combine the advantages of both previous approaches with a smart use of Time Division Multiplexing (TDM) and WDM methods to provide sub-wavelength granularity with reduced electronic processing. These methods are referred to as Optical Packet Switching (OPS)/Optical Burst Switching (OBS) [96], [21] or Time Shared Optical Network (TSON) [109] depending on the use of a time slotted transmission, aggregation type and on the typical packet/burst duration. The time domain sub-wavelength switching requires a heavy change in minds: shifting to burst mode receivers, ultra-fast tunable lasers and new switching paradigms. Many declinations and variants, derived from the original OPS/OBS concepts, have been proposed in the literature. The Sub-Lambda Photonically Switched Networks (SLPSN) [40] is a recently proposed term in some International Telecommunication Union - Telecommunication (ITU-T) contributions for this OPS/OBS concept.

The new transport networks should support very high bitrate interfaces, dynamic bandwidth provisioning such that they meet customer's demands and use the existing fibers. In this context, a SLPSN paradigm may be more appropriate than the currently deployed OCS, in order to bring the required flexibility in the transport layer.

Besides, SLPSN techniques have proved to be very efficient in terms of bandwidth occupancy for simple topologies and constant packet duration, e.g., a time-slotted metropolitan ring based on Packet Optical Add Drop Multiplexer (POADM) technology [96]. For mesh networks, other interesting SLPSN approaches, such as Time-Domain Wavelength Interleaved Network (TWIN) architecture [103] and its variant, rely on concentrating all the complexity at network edge and using passive optical nodes.

This work is devoted to the dimensioning of TWIN and comparison with POADM technology in different scenarios, these technologies provide much more flexibility than the present ones based on OCS and fixed capacity channels. The main objectives of this document concern:

- the cost dimensioning, resource allocation and Routing and Wavelength Assignment (RWA) in TWIN;
- the definition of a generic framework for the comparison of these solutions, and the identification of the main cost elements for Capital Expenditures (CAPEX) and Operational Expenditures (OPEX), including the impact of the optical node structure;
- the definition of different scenarios for using these techniques in metropolitan or core networks, taking into account protection issues;
- the dimensioning and the evaluation of these architectures according to the most relevant scenarios.

The contribution of this thesis is a part of European SASER-SaveNet project. The project is devoted to the evaluation of different optical packet-based technologies and architectures. SASER-SaveNet project started on July 1st, 2012 and its leader in France is Alcatel-Lucent Bell Labs. The other participants are Télécom Bretagne, France Télécom - Orange Labs, Télécom SudParis, Univ. Rennes 1.

The outline of the document is given in the following:

Chapter 1 provides a detailed overview on optical switching networks, with the focus on lossless optical sub-wavelength switching technologies. Lossless transmission meaning no burst or packet loss, it is achieved thanks to the use of a control of the emitters that avoid collision of data based on accurate synchronization of sources. The chapter briefly explains the dimensioning issues in optical networks, and the existing work on TWIN dimensioning. This chapter also summarizes the previous work done at Télécom Bretagne. The chapter is concluded with the thesis motivation.

Chapter 2 proposes a control plane for TWIN network based on virtual ring joining all sources and destinations, in order to support an out-of-band control channel. In addition, the network configuration is achieved by using the control channel, in particular to carry messages related to network discovery, synchronization and protection issues.

Chapter 3 deals with TWIN network design. In this chapter, two ways of improving the slot allocation efficiency in such networks, are proposed: by minimizing the schedule length and by enforcing a cyclical Quality of Service (QoS) constraint on the allocated slots. The detailed simulation results are reported, showing how network topology, traffic pattern and the proposed QoS improving method impact the network performance. For the first time, a 0-1 Integer Linear Programming (0-1 ILP) dimensioning tool is proposed, which simultaneously addresses the RWA and scheduling (resource allocation) problem in TWIN. In addition, a comprehensive scalable algorithm, that assesses the same problem and considers network virtualization, is developed. The algorithm enables the construction of overlayed virtual networking domains at minimum

cost sharing the transponders and network links, and can employ different scheduling policies. The performance of the algorithm is compared with the 0-1 ILP solution for single virtual domain, and for different scenarios. Furthermore, the study is extended to a real size network.

In Chapter 4, the cost of the optical transport solutions with sub-wavelength switching granularity, is compared. Hence, two scenarios are considered: survivable metro networks and Long Term Evolution (LTE) mobile backhauling, with the focus on TWIN and POADM technologies. The cost model accounts for the node components, the transponder cost and the wavelength leasing cost per km of fiber. A single network dimensioning model supporting protection and multicast traffic is proposed for these technologies. The network planning solution is based on 0-1 ILP formulation, and addresses the RWA and the optimal schedule calculation.

The document is concluded by providing the general conclusion and possible perspectives.

Appendixes A and B describe the work that was carried out, during my three months visit at Keio University in Yamanaka Lab, Japan, in a context of a Japan-USA Project named ACTION. Appendix A presents the research dedicated to application-centric energy efficient Ethernet, which meets the QoS requirement for different Classes of Service (CoS). In Appendix B, the adaptive link rates for energy efficient 100 Gb/s Ethernet (100GE) is proposed, with the focus on the rate switching time reduction using coherent lightware transmission technologies. The studies presented in the Appendixes are outside of the scope of this thesis.

In addition, Appendix C provides the list of personal publications done during the last three years.

CHAPTER 1

State of the Art

This chapter represents an overview on existing work related to optical switching network, and to optical sub-wavelength switching technologies, focusing especially on existing work on Time-Domain Wavelength Interleaved Network (TWIN) dimensioning. Finally, in the last section of this chapter, a list of problems that occur when designing a TWIN network and which are addressed in this thesis, is given.

1.1 Optical networks

With the advance in technology, the cost of optical technologies has become very low, and a huge amount of information can be transmitted over a single fiber. The optical networks are the preferred medium for the fixed data transport network since fiber offers a large amount of bandwidth with minimum attenuation and longer reach. Nowadays, the transport networks rely on Wavelength Division Multiplexing (WDM) optical networks. The WDM partitions the optical bandwidth into a large number of channels, that allows multiple data channels to be transmitted along the same fiber [53], which reduces network cost.

In general, there are two major types of channel switching in optical networks:

Opaque switching

In opaque switching networks, all the point-to-point channels are transmitted over fibers, however Optical-Electrical-Optical (OEO) conversion is performed at each node. The traffic is then processed in order to drop or forward some flows. In the case of forwarding, the electrical signal is converted back into optical domain and sent into fibers towards its destination.

However, the OEO conversion is costly, in addition, this process is very energy consuming and the large increase of the traffic volume has made the energy consuming even larger. Thus, it would be useful to find solutions to have all optical transmission at intermediate nodes (i.e., without conversion to the electrical domain).

Transparent switching

In case of transparent switching networks, the optical channels (i.e., wavelengths) between two network nodes are optically switched at the intermediate nodes and the signal is converted back to the electrical domain only at the destination. Thus, from source to destination signal stays in optical domain.

1.2 Switching in transparent optical networks

Since the appearance of WDM technique, different switching solutions have been studied during the years. In the following sections are introduced the Optical Circuit Switching (OCS) in fixed and flexible grid networks, as well as optical sub-wavelength switching solutions.

1.2.1 OCS in fixed grid network

In OCS, optical switching is done at the wavelength level [53]. The optical circuits are being established statically or dynamically between source and destination. In case of

wavelength routed networks, it is possible to change the assigned wavelength within the same communication, if the nodes in the network have wavelength converters. Otherwise, the same wavelength is used on all the links along which the communication is routed.

One of the advantages of the OCS is that usually the duration of the established circuits are long and also the Quality of Service (QoS) provided by this method is high thanks to the stability of the multiplexing and switching methods [25]. The main inconvenience of transparent OCS is the poor aggregation capability (i.e., lack of bandwidth flexibility), which leads to the insufficient use of the available bandwidth. These problems make the OCS networks more sensitive to bursty traffic.

1.2.2 OCS in flexible grid network

Over the past few years, the advances in optical transmission and switching technologies have paved the way for flexible optical networks [42], [19], [83] and [7]. Differently from fixed grid (e.g., with a fixed channel spacing of 50 GHz), flexible grid enables optical channels with diverse bandwidth and a fine channel spacing, and thus, higher spectrum efficiency.

Flexible grid network represents a relevant innovation for improving the spectrum efficiency, as the bandwidth of the *lightpath* (wavelengths in case of fixed grid network) can be optimized to offer the requested bitrate for the adopted modulation format. As an example of spectrum optimization, a lightpath at 100 Gb/s using polarization multiplexing 16 quadrature amplitude modulation (PM-16QAM) requires half of the theoretical Nyquist bandwidth of 100 Gb/s polarization multiplexing quadrature phase shift keying (PM-QPSK), i.e., 14 GHz for PM-16QAM and 28 GHz for PM-QPSK including the overhead.

In such networks, the portion of the spectrum occupied by a lightpath is identified by a frequency slot that is defined by the central frequency (i.e., the carrier) and the slot width (i.e., the occupied bandwidth). According to International Telecommunication Union - Telecommunication (ITU-T) recommendations on a flexible grid [39], the central frequency may vary at a step of 6.25 GHz, while the slot width is defined as a multiple of a minimum bandwidth granularity of 12.5 GHz. Hereafter, the minimum granularity of 12.5 GHz is called a frequency slice. As an example, for a lightpath requiring 35 GHz of bandwidth, the corresponding slot width would be 37.5 GHz, which consists of three spectrally adjacent slices of 12.5 GHz.

1.2.3 Optical sub-wavelength switching

The increase of bandwidth demand in the metropolitan networks [1] fosters the need for the new optical transport layer that shall ensure a highly efficient bandwidth utilization of the optical resources, in the same time, improving the network energy efficiency.

The sub-wavelength switching optical technologies, like Optical Packet Switching (OPS) and Optical Burst Switching (OBS) are proposed in order to get around the

lack of flexibility of OCS solutions and to benefit from the whole available bandwidth by efficiently filling wavelengths [51]. This is possible thanks to their fine switching granularity and benefits of statistical multiplexing. In the same time, as result of the reduction in the amount of electronic processing of the transit traffic (optical transparency), these networks achieve high energy efficiency [3].

Optical Packet Switching offers the finest sub-wavelength granularity by converting directly the electrical packets (i.e., IP/Ethernet client packets) to the optical domain. The main advantage of OPS is the statistical multiplexing that leads to higher throughput and maximizing the network resource utilization [25]. There are intermediate solutions such as OBS [107], for implementation purposes.

Optical Burst Switching is a sub-wavelength switching paradigm that can have a variable or fixed size duration to carry more than one IP/Ethernet packet aggregated in a burst. The burst header information is carried in a separate channel that is transmitted in band being processed in electronically.

In the ITU-T standards, the sub-wavelength switching networks are referred to as Sub-Lambda Photonically Switched Networks (SLPSN) [40].

In following subsections, a short review of several technologies for lossless optical sub-wavelength switching solution is considered.

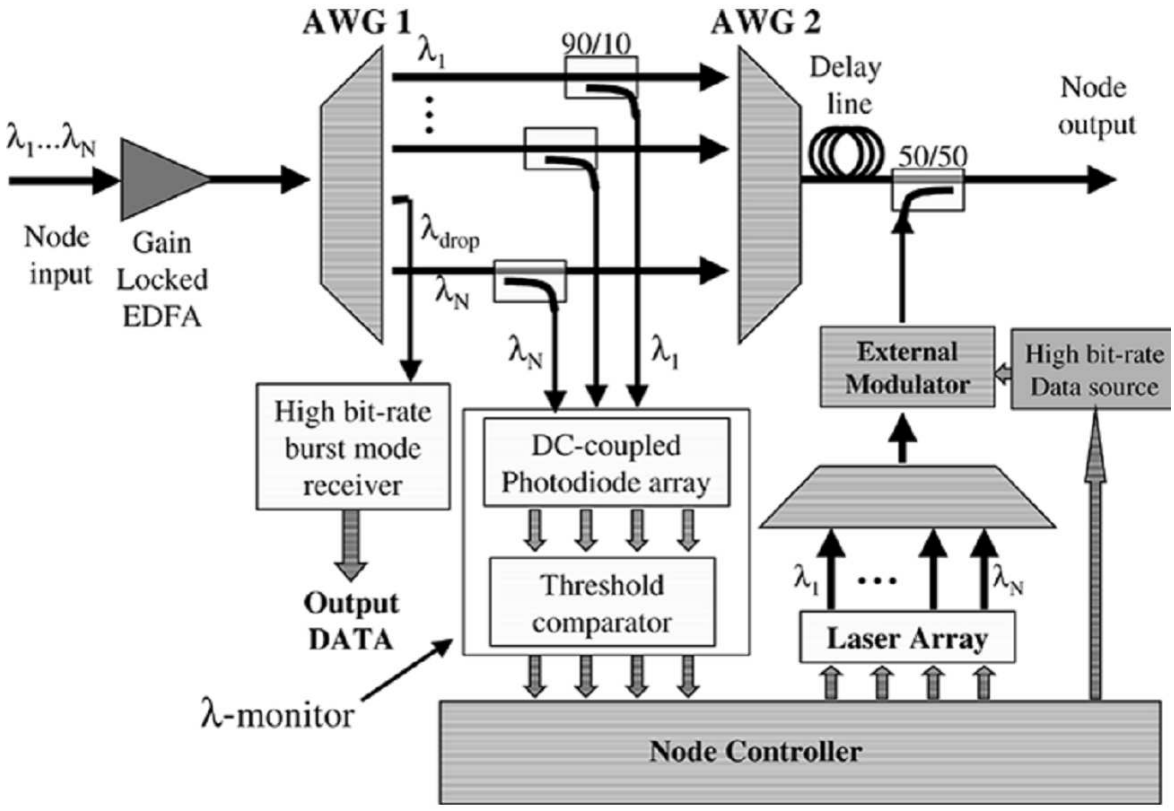


Figure 1.1 — Architecture of the first version of RINGO node (according to [6]).

Ring Optical Network (RINGO)

Ring Optical Network (RINGO) is the outcome of Italian project focused on optical time-slotted packet switched WDM network. It originates from the same work as the HORNET project. Early version of RINGO was defined based on unidirectional ring [6]. This version consists of N nodes and per each node one wavelength is used for packet reception. The wavelength for packet emission is variable. A packet is sent per frame, by using a wavelength on which no incoming packet has been detected, as presented in Fig. 1.1. To avoid collision, the free time slot is sensed by the monitoring method per each time slot. To access the time slots, the priority always has upstream transit packet and the packets to be sent are electronically stored in virtual queues per destination. The virtual queues are used to avoid the problem of head of line blocking. For this purpose, specific algorithms have also been studied to improve fairness.

The latest version of RINGO node is based on a bidirectional ring. It allows the sharing of a reception wavelength by several nodes. Insertion and extraction are performed on two different fibers (Fig. 1.2). At some point, the two directions of rings are folded at a node called Master node in the ring. Thus, the actual topology forms a double bus.

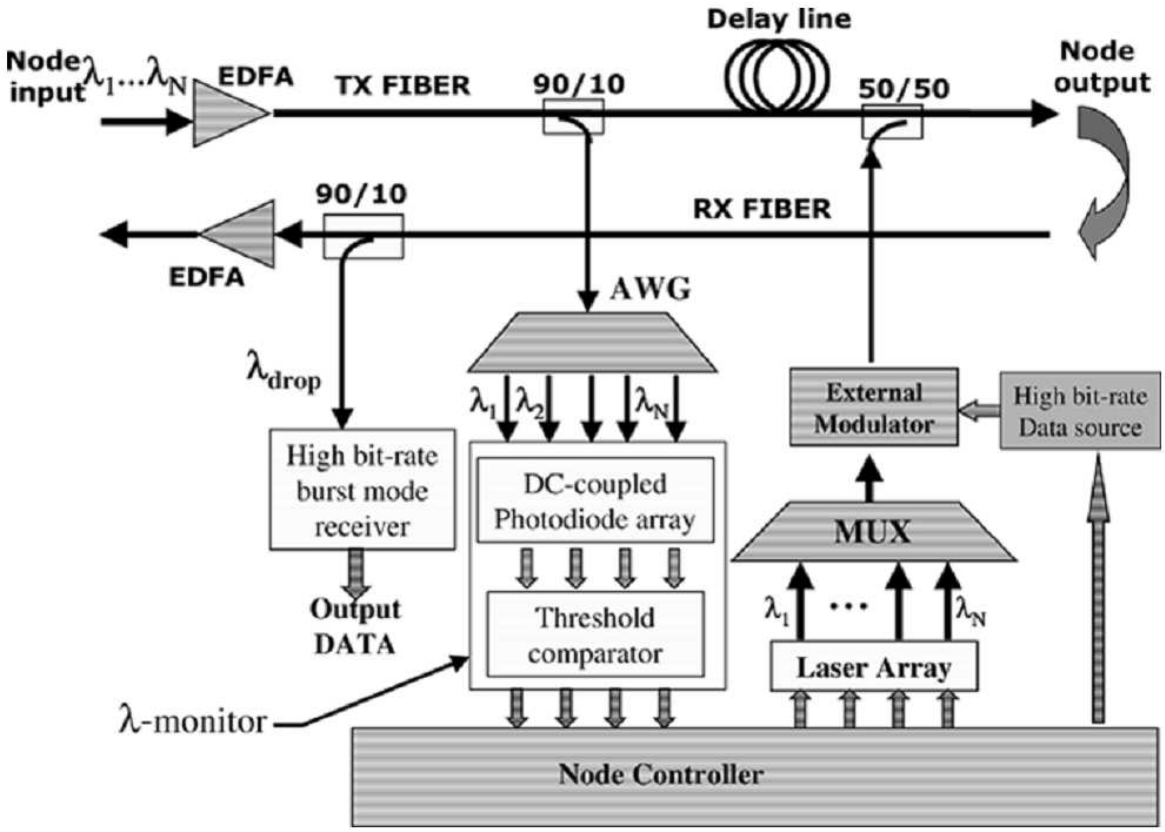


Figure 1.2 — Architecture of the latest version of RINGO node (according to [6]).

Data And Voice Integration on Dense WDM (DAVID)

The Data And Voice Integration on Dense WDM (DAVID) project [22] proposes a practical approach for OPS in Metropolitan area network. The project took place between 1999 and 2003. The metro network comprises one or more unidirectional optical physical rings interconnected by a Hub, as presented in Fig. 1.3. A ring within the MAN consists of one or more fibers, each operate in dense WDM regime, where each wavelength is used to transport optical packets having a fixed duration in time (i.e., time slot).

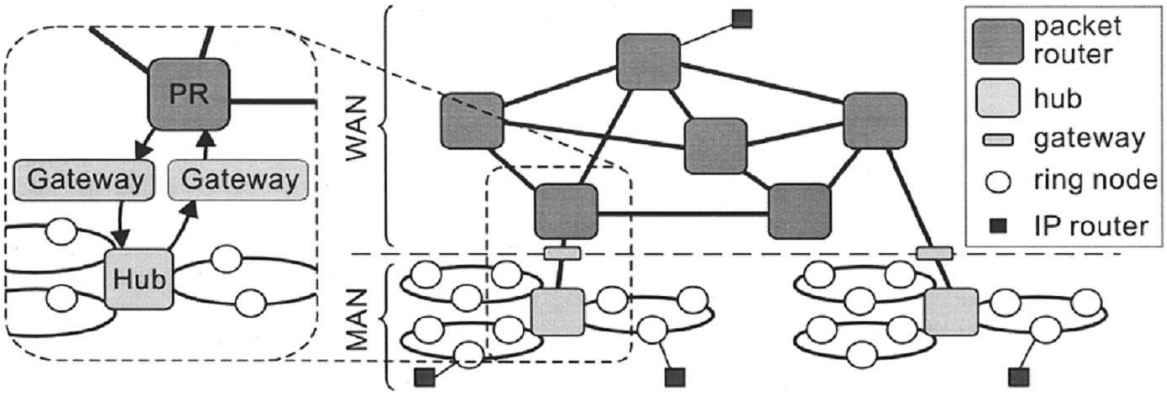


Figure 1.3 — DAVID network architecture (according to [22]).

At each node it is only possible to receive or transmit on one channel at a time. The hub is buffer-less and is the interconnection point of multiple rings and provides access toward the wide-area network (WAN) through a gateway. Each ring uses a wavelength channel dedicated to control information.

Packet Optical Add Drop Multiplexer (POADM)

In Packet Optical Add/Drop Multiplexer (POADM), the physical topology of the network is based on bidirectional ring [12], as in Fig. 1.4, where each network node can use any fiber and any wavelength for the insertion or the extraction of traffic [13]. The wavelengths for the insertion/extraction process might be shared by different network destinations. The only condition is that the destinations must have the transponders able to receive on these wavelengths.

For network control purpose, the nodes are connected with a single out-of-band control channel. The data channel is switched all optically. However, the control channel packets are OEO converted at each node in order to be processed, since control information relative to each data channel and to the global ring (e.g. Operations, Administration, and Maintenance (OAM) messages) is carried within the control packets. The data and control channel are synchronized and have fixed size slots.

The simplified node architecture of POADM [12] is illustrated in Fig. 1.5. The incoming WDM line is demultiplexed passing through the optical DEMUX.

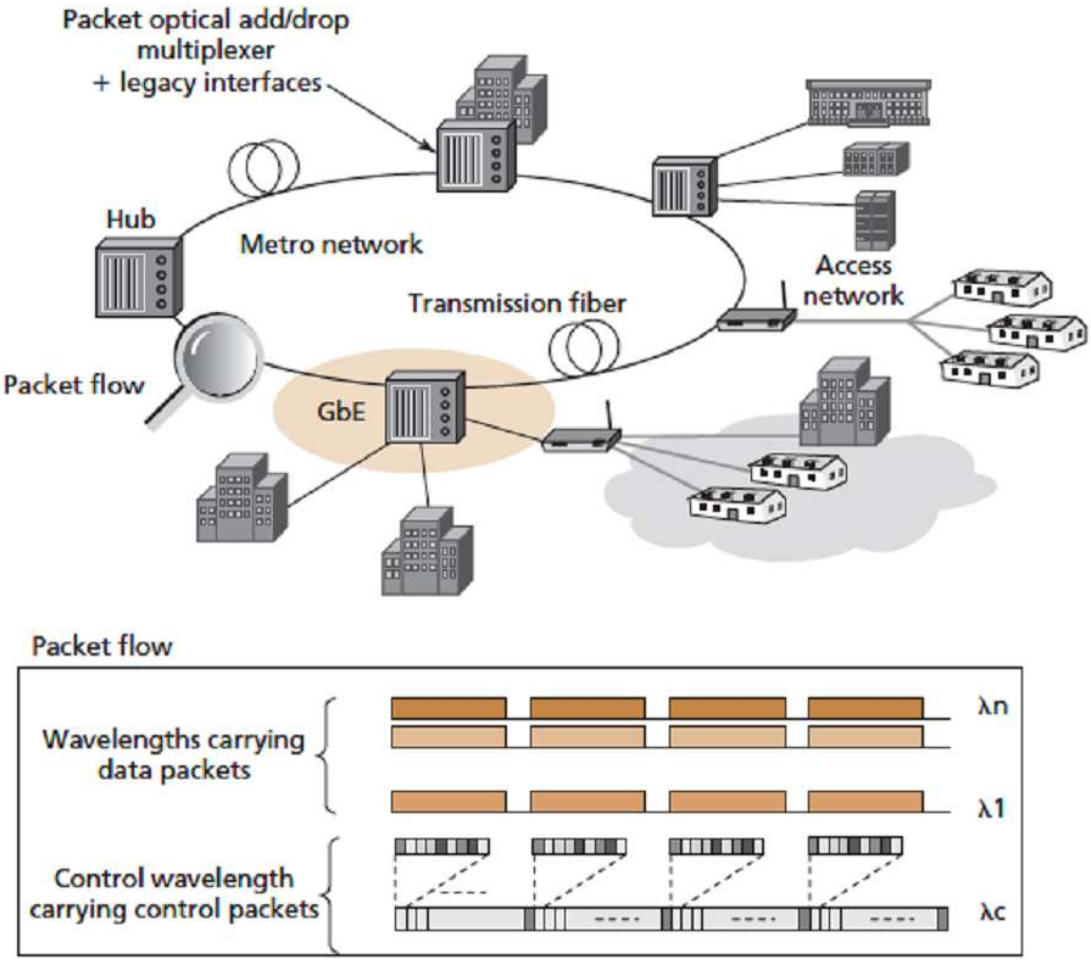


Figure 1.4 — Metro network based on POADM (according to [13]).

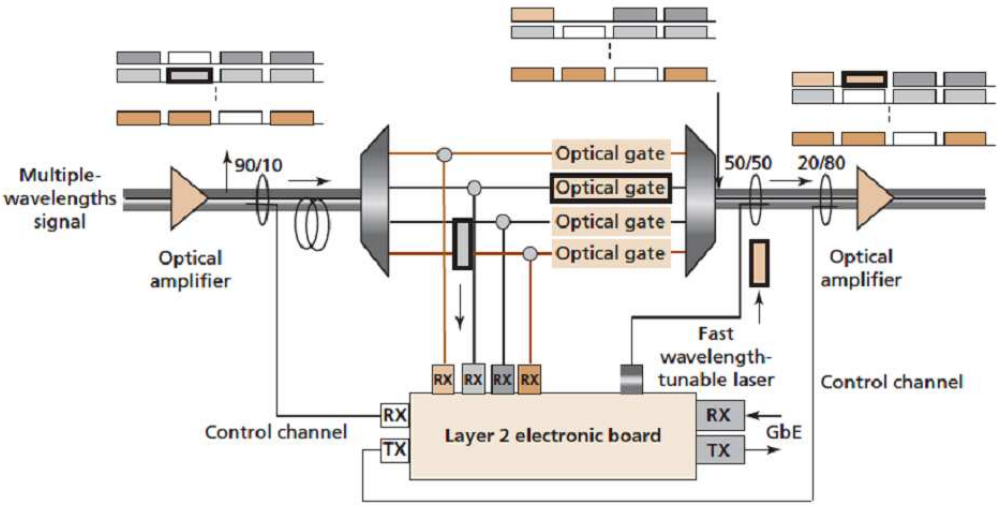


Figure 1.5 — POADM node structure (according to [13]).

The wavelength blocking is achieved with fast optical gates that can selectively erase any slot, based on the content of control packet. These gates can be based on Semiconductor Optical Amplifiers (SOAs), and their number is equal to the number of wavelengths used in the network (per ring direction).

Time-Domain Wavelength Interleaved Network (TWIN)

Among many approaches which have been proposed for OPS/OBS, the Time-Domain Wavelength Interleaved Network (TWIN) [103], [76] is considered as a rather attractive solution, as it aims combining passive optical transparency and efficient use of wavelength resources.

TWIN is a sub-wavelength switching technique that transmits bursts or slots of data on WDM channels without resorting to electronic processing in transit nodes, the electronic processing functions being pushed to its edge source and destination nodes (Fig. 1.6).

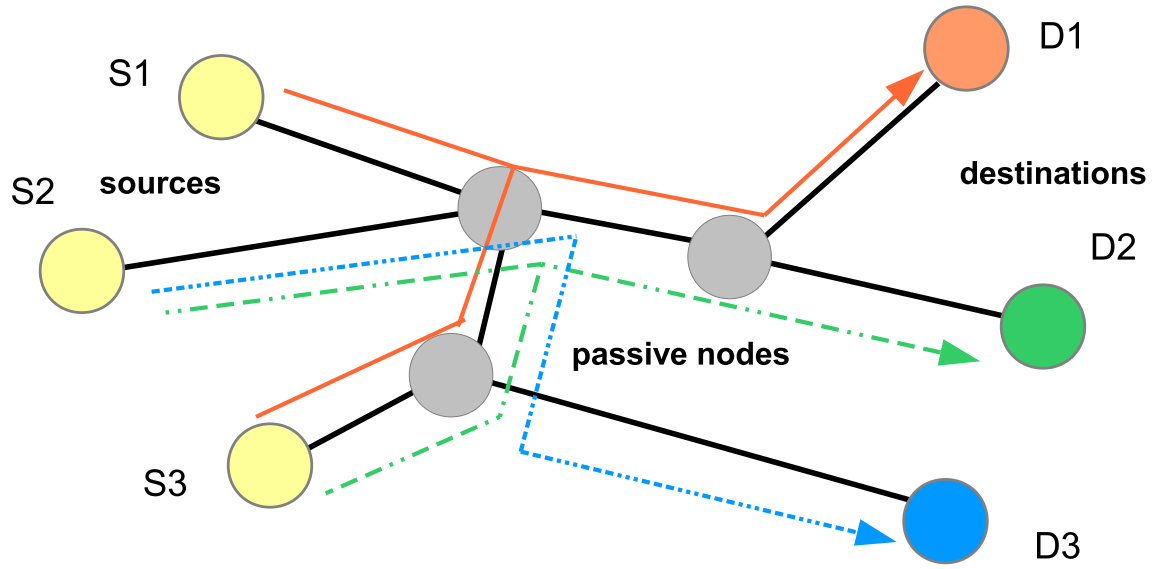


Figure 1.6 — TWIN architecture.

The switching in the network is based only on the burst wavelength, thus, is purely passive. Therefore, the logical topology of TWIN can be viewed as a set of optical multipoint-to-point trees overlaid on top of a mesh network, one for each destination. Each tree is rooted at a fixed-wavelength (colored) receiver on a (destination) node. The number of required wavelengths in TWIN is higher or equal to the number of nodes (i.e., destinations), because of the specific wavelength distribution to the nodes,

Fig. 1.6 illustrates several trees, with nodes D1, D2, D3 being destinations that receive traffic from nodes S1, S2, and S3, by using a dedicated wavelength. In this configuration, the intermediate nodes are “core” passive nodes, connecting two “edge” nodes (e.g., S1 and D2).

Furthermore, by tuning its emitter to the appropriate wavelength, a source can send a burst to any destination. A node knows when to send and receive slots thanks to a schedule that is computed in a centralized [84] or distributed [92] fashion. The schedule can be periodically sent to each node via a control channel that is processed at every node. The control channel can be implemented with low data-rate equipment (e.g., 1 Gb/s) to mitigate its cost.

In TWIN, each TRX is equipped with a fast wavelength-tunable transmitter and a burst-mode (e.g., access-grade) wavelength-colored receiver. Relatively inexpensive non-coherent (e.g., 10 Gb/s) technology can be leveraged to implement those edge nodes, but the coherent solution can be also envisaged. The TWIN is expected to be realized over Reconfigurable Optical Add Drop Multiplexer (ROADM) architecture.

1.3 Dimensioning of transparent optical networks

The purpose of dimensioning of an optical network is to ensure that the incoming traffic will be served in an economical way, respecting the QoS and operator requirements.

In the following subsections, the dimensioning principles for OCS in fixed and flexible grid network and OPS/OBS networks (with the focus on TWIN network) are described.

1.3.1 Dimensioning of OCS in fixed grid network

The main dimensioning problem of OCS in fixed grid network is the Routing and Wavelength Assignment (RWA) problem. Even further, its main constraint in RWA problem is devoted to OCS networks operating under the wavelength-continuity constraint, in which lightpaths are set up for connection requests between node pairs, and a single lightpath must occupy the same wavelength on all of the links that it spans.

It is well known that the WA can be mapped to the coloring problem [108], which aims at coloring the nodes of a graph such that adjacent nodes are assigned different colors. The WA problem can thus be solved by the following steps [108]:

- Construct an auxiliary graph $G(V, E)$ such that each channel in the system is represented by a node V in graph G. There is an undirected edge E between two nodes in graph G, in case when the corresponding channels pass through a common physical fiber link. This graph is named the path conflict graph.
- Solve the coloring problem by coloring the nodes of the path conflict graph G such that no two adjacent nodes have the same color.

The number of colors obtained by the coloring problem is equal to the total number of wavelength required by the network, which are sufficient to satisfy all the traffic demands. This is one of approaches for solving the RWA problem, however, in [108], various routing and wavelength assignment approaches (based on Integer Linear Programmes or heuristics) proposed in the literature are reviewed.

1.3.2 Dimensioning of OCS in flexible grid network

The shift of paradigm from fixed to flexible grid network poses additional challenges not only at the physical layer, but also at the network layer. Whereas in the fixed grid network, the provisioning of lightpaths is addressed by solving the RWA, while in the flexible grid network, the problem becomes a Routing and Spectrum Assignment (RSA) one. The main difference is that the problem of assigning bandwidth to the lightpaths has changed from an assignment of one single color (representing a fixed amount of bandwidth) to an assignment of a variable number of colored frequency slices. The former problem, i.e., the wavelength assignment (WA), requires the assignment of a single wavelength or color (or multiple wavelengths when wavelength converters are available) on each link of the lightpath route, so that, the same wavelength is not used by any other lightpaths passing on the same link(s). The latter problem, i.e., the spectrum assignment (SA), consists of assigning a central frequency and a slot width (i.e., a set of spectrally adjacent frequency slices) to the lightpaths from source to destination, while ensuring that the frequency slices are not reserved by any other lightpaths passing on the same link(s).

In the SA problem, the adjacency constraint of the frequency slices is introduced and must be enforced. Moreover, the number of slices to be assigned depends on the bit rate of the lightpath and the modulation format selected for the transmission. Thus, the SA problem is further complicated by the fact that flexible grid network can be enabled for supporting Multi Modulation Formats (MMF) and Multi Line Rates (MMF). The use of MMF and MMF leads to efficient bandwidth utilization.

The (R)SA problem in a flexible grid network is addressed by various works and can be optimally modeled by Integer Linear Programming (ILP) formulations [16], [59], [100] and [48]. The objectives of such works span from maximization of the spectrum efficiency [101], [77] to minimization of the cost due to regenerators and transponders [24], [77] to minimization of the overall cost including the fiber cost [60]. Similarly, an ILP formulation is proposed in [16] for an optical network without grid (i.e., the slot width can take any value and is not necessarily a multiple of 12.5 GHz) supporting orthogonal frequency division multiplexing signals with different modulation formats.

In all these related works, the ILP formulations impose the adjacency of the frequency slices [83], [7] assigned to each given lightpath (i.e., the slices composing the frequency slot must be contiguous in the spectrum without any interruption). The adjacency constraint is also enforced in the ILP formulation presented in [29], which aims at rearranging already established connections to reduce the spectrum fragmentation. Thus, it is correct and necessary, that the enforcement of the slice adjacency adds a new constraint to the formulation, leading to a more complex problem formulation with respect to the ILP formulations for (R)WA in a fixed grid network. Besides, an additional degree of complexity arises when MMF [16] [60], [59], and MMF [16], [77] are supported in flexible grid network (or in fixed grid network [54]), as the line rates and the modulation formats need to be optimally selected for each lightpath. Indeed, the selected modulation format and line rate determine the number of frequency slices to be allocated to the lightpath, thus, affecting the SA problem.

In the paper [64], authors focus only on the SA problem (i.e., lightpath routes are

fixed and precomputed) in a static flexible grid network supporting MMF and MMF. The main relevant contribution of the paper is the proof that the SA problem needs not enforce the adjacency constraint. Indeed, after solving the weighted coloring problem [33], [2], [18] on the frequency slices (without adjacency constraint), the frequency slices of each slot can be allocated in a spectrally adjacent manner in polynomial time, having the same minimum number of slices. This findings drastically simplify the problem of (R)SA in flexible grid networks.

1.3.3 Dimensioning of OBS/OPS networks

Optical slot switching is a networking technology that can provide statistical multiplexing energy-efficiently, at the expense of complex dimensioning tools as a result of small granularity. In this thesis, we focus on dimensioning of such sub-wavelength granularity optical transport technologies. In particular, we focus on TWIN, which exploits OPS/OBS in mesh networks. In the following subsections, the literature review on scheduling and RWA in TWIN is presented. In addition, the protection problem in TWIN is also considered.

Scheduling and RWA in the TWIN network

The scheduling in TWIN has been extensively studied by different research groups, with accent on minimizing the schedule length, on centralized and distributed scheduling, and considering different traffic dynamicities (e.g., see [76], [84], [90]). Schedule length is a few dozens to a few hundreds of slots, which repeats in time, and could be recomputed every few milliseconds, seconds, minutes, or hours, as needed, depending on the traffic variation or network configuration changes.

In [76], a method defining the minimum schedule length that satisfies all flow demands is proposed. The distributed scheduling with random allocation is considered instead of the centralized one, for both dynamic (schedule should be recomputed every few milliseconds) and semi-dynamic traffic (schedule should be recomputed every few minutes or hours) in [84], while in [92] a comparative study between centralized and distributed burst scheduling is proposed. Furthermore, in [91] and [90], the authors prove the high performance of static scheduling for TWIN, which is based on computing an optimized schedule using an ILP model. The problem of computing an optimized schedule is a NP complete problem [91]. A heuristic method for scheduling the optical bursts in TWIN, for a fixed network configuration, called TIIS (TWIN iterative independent set) algorithm, is presented in [76], the proposed heuristic method schedules bursts without taking into account the dimensioning issues.

The static schedule is kept unchanged during a long period of time, contrary to the dynamic schedule that is changed according to the traffic variation. In [90], authors demonstrate the possibility to deploy the static schedule in a real network and its ability to meet QoS requirements.

The ILP used in [91] and [90] considers only the scheduling issue and does not take into account RWA problem. However, the schedule additionally should define what

link and wavelength (in addition to which slot) is used for a source to send data to a destination. Hence, the classical problem of RWA is a sub-problem of schedule computation in TWIN.

Optimally solving the scheduling sub-problems in TWIN cannot be done in polynomial time [91], so the Linear Programming dimensioning tools proposed in previous works are limited to small examples and cannot be used for real-scale networks. Therefore, the dimensioning tools which provide a scalable solution for the RWA and scheduling optimization are needed.

Protection in the TWIN network

The scheduling issues due to propagation delays is the central topic in TWIN, and has been studied for instance in [75], [70], [105]. When scheduling the slots in TWIN, the conflicts between different sources and destinations may result in slot collisions, and hence, they must be avoided in the entire network.

Since enabling protection in TWIN is a very important task, it needs to be taken into account when planning a network. In case of a failure such as a link cut, a part of the traffic shall be rerouted on other precomputed paths, which yields for a need of additional bandwidth reservation. This bandwidth shall be available at every moment, but its quantity will depend on the protection scheme — shared (the same resource can be used for the protection of several paths, or, in the context of TWIN, trees) or dedicated (the separate backup resources are allocated to each working connection) — that is used.

The protection in TWIN has been only partially addressed so far. In [103], the possibility of using multiple trees to provide dedicated protection is considered. In that paper, the process of allocating the slots is based on a heuristic algorithm, while the TWIN-tree calculation problem has been resolved separately from the slot allocation. The cost of shared protection was not considered, and the cost of allocating the backup resources is considered in terms of the schedule size increase, and not in terms of the Capital Expenditures (CAPEX) cost. In [8], the authors discuss the possible protection methods for a variation of TWIN that uses coherent technology, but protection is not numerically assessed.

It is very important to simultaneously address the RWA and the slot allocation problem in TWIN, since the schedule depends on the propagation delays in the network. Indeed, in TWIN, the scheduling should be calculated to avoid any slot collisions between the concurrent sources and destinations. The configuration of the overlapping slots depends on the propagation distances between source-destination pairs in the network, which are the output of the RWA problem. Thus, a combined dimensioning formulation for TWIN is required, which is able to simultaneously address the routing and wavelength assignment problem, and to calculate the optimal schedule and the backup paths for dedicated and shared protection (i.e., the slot collisions should be avoided at both working and backup paths). Such dimensioning tool can be of use to network planners, whether the network fibers are already installed, or in the case of so-called “green field” investments, when the network paths should be optimized as well.

1.4 Previous work done at Télécom Bretagne

During the last years the Télécom Bretagne, and especially the Optics and Computer Science departments, were involved in several projects on sub-wavelength switching technologies.

The work done by Dr. Bogdan Uscumlic, as a part of his thesis, addressed the problem of dimensioning [94], performance evaluation [97] and selection of the scheduling algorithm in the metropolitan optical packet switched rings (i.e., Packet Optical Add Drop Multiplexer (POADM)). This work studied a new slotted optical packet switching technology for metropolitan network in form of ring, which was developed within the French ECOFRAME project. ECOFRAME project took place between 2007 and 2010, its leader was Alcatel-Lucent Bell Labs. The other participants in the project were Télécom Bretagne, Télécom SudParis, and others.

As an extension of the previous work on POADM, the thesis done by Dr. Lida Sadeghioon, provided different features such as QoS, Multicast, and protection by defining identifier fields in the control packet content per each time slot [80]. Moreover, the author adopted the labelling concept to identify flows and service features via the control channel packet, such as unicast or multicast traffic, guaranteed or best effort traffic, protection type (including un-protected). As a result, the author defined a label based MAC offering multiple services relying on separate control packet messaging system. In addition, during the last year of her thesis and first year of my thesis, we addressed together the control issues of TWIN. We suggested a use of dedicated control channel on a virtual control ring topology, then we extended the idea of virtual control topology to introduce a protection path and studied the impacts on network configuration and failure recovery. This work is partly described in the first chapter of this document.

Lastly, Dr. Ahmed Triki, during his thesis, revised the original proposal of the TWIN architecture and mechanisms, and proposed several algorithms to realize the management/control plane [92]. He considered dynamic control planes that realize a closed loop control avoiding burst contention on the basis of a dynamic evaluation of requested resources, and a static control plane that operates in open loop, under the assumption that requested resources are known (e.g. thanks to management plane information). The dynamic control planes are based on a heuristic approach for resource allocation, which changes according to the traffic variation observed during a short period. On the other hand, the static scheme is based on an optimized resource allocation implementing a 0-1 Integer Linear Programming (0-1 ILP) formulation. Lastly, he proved the feasibility of TWIN by designing an experimental node operating with a static control plane. On a small test-bed [78], he has succeeded in ensuring synchronization and in obtaining correct time accuracy with good quality optical signals. However, the RWA, protection and multicast were not addressed in his thesis.

1.5 Thesis problem formulation

In this thesis, the problems related to the Time-Domain Wavelength Interleaved Network (TWIN) design are addressed. We focus on TWIN paradigm since it is a lossless solution and it is designed to be deployed on a mesh topology. TWIN is a promising Optical Packet Switching (OPS)/Optical Burst Switching (OBS) technology because of its node structure simplicity and its bandwidth flexibility. In addition, only edge nodes perform electronic buffering, while the intermediate nodes consist of passive optical components and operate at full optical capacity without any electronic processing.

The contribution of this thesis is divided into two main parts. First contribution concerns the development of dimensioning tool which simultaneously considers the Routing and Wavelength Assignment (RWA) and scheduling (resource allocation) problem in TWIN. The problem of scheduling in TWIN is important and it is a sub-problem RWA. This is why, in this thesis, these problems are studied together, in order to try to understand the mutual impact and to find the optimal cost of the network design. The problem of how network topology, traffic pattern and the proposed QoS improving method impact the scheduling, are also studied. The problem of Quality of Transmission (QoT) in reference to Bit Error Rate (BER) are not considered in the RWA algorithms proposed in this thesis.

The second contribution is dedicated to the comparison of sub-wavelength optical switching networks in case of different scenarios. Two scenarios are considered, the survivable metro networks and Long Term Evolution (LTE) mobile backhauling with the focus on TWIN and Packet Optical Add Drop Multiplexer (POADM) technologies. A single network dimensioning model (based on 0-1 Integer Linear Programming (0-1 ILP)) supporting protection and multicast traffic is proposed, for the first time, for these technologies.

CHAPTER 2

Control Plane of Time-Domain Wavelength Interleaved Network

In current chapter, we propose using a virtual ring joining all sources and destinations in Time-Domain Wavelength Interleaved Network (TWIN), in order to support an out-of-band control channel, which is used in particular to carry messages related to network discovery, synchronization (i.e., clock reference management) and protection issues. Synchronization and failure notifications (generated by standard optical layer monitoring procedures) are processed by the central element in charge of resource allocation. The use of a single virtual control ring is well suited to networks with a limited extension such as metro networks. However, this approach can be extended to larger scale networks by using several control rings. We describe an algorithm to cover a large network operated by TWIN with minimum number of interconnected rings, taking into account the Quality of Service (QoS) requirements and/or the paths maximum transparent reach.

2.1 Introduction

Time-Domain Wavelength Interleaved Network (TWIN) technology relies on sending and receiving grants and reports based on the demands at the sources and scheduling performed at destination or at central entity, to be able to access the medium and provide collision and blocking free access. Both centralized [103], [75] and distributed control schemes [70] have been proposed for TWIN.

In case of distributed scenario, each destination assigns to each source time slots, to assure proper scheduling, based on report information generated by the sources. By issuing the reports, each source notifies the destinations about its capacity needs. A source may receive grants from different destinations for the same time slot, which can lead to source blocking.

On the other hand, in the centralized scenario, all the sources send the reports to a Control Entity (CE), which is in charge of assigning the grants to all sources. It has been recently shown [92] that the scheduling efficiency is higher in centralized schemes, in respect to the distributed one.

In TWIN network the edge nodes perform the main functions, including all those related to data, management and control planes. Inside of the network are the passive optical components that are preconfigured and passively perform the switching of the arriving optical bursts according to their destination wavelength. Such network configuration is possible thanks to the wavelength based destination addressing, where a single wavelength is allocated to each edge node in a TWIN network. The operation of the network slotted (as considered here), the slots are of fixed. TWIN uses transporders composed of fixed-wavelength receiving part and fast-wavelength tunable transmitting part. A detailed description of the TWIN architecture and functionality is given in the Section 1.2.3 of the previous chapter.

In this chapter, the control and synchronization issues are addressed, which are the key elements for TWIN operation as a transport network. The proposed approach involves the use of a set of virtual rings covering all sources and destinations (i.e., edge nodes), in order to support an out-of-band control channel. To note that, self-healing rings have been studied previously, (e.g., see [28]), to simplify the control plane operation as well as to provide the alternative paths for protection. However, the previous works usually addressed optical circuit switched networks, except the work done in [82], where the authors present a model to cover the network with rings in order to provide a fully transparent design of a hybrid optical packet/circuit metropolitan area network.

The work described in this chapter was developed in collaboration with previous PhD student, Dr. Lida Sadeghioon, during the time when she was at Télécom Bretagne. This study was a part of her last thesis chapter.

The proposed concept is detailed in Section 2.2, which describes the introduction of a virtual control ring in TWIN and its application for synchronization purpose (Section 2.3). Section 2.6 presents a planning method for extending this approach to a global network using multiple virtual rings. The application of the method is illustrated in Section 2.7, by considering Deutsche Telekom's topology. Finally, the concluding remarks are given in the last section.

2.2 Virtual TWIN rings for control

In this section, we present a dedicated control channel on a virtual control topology for transmitting scheduling, control, synchronization and Operations, Administration, and Maintenance (OAM) messages based on centralized control schemes.

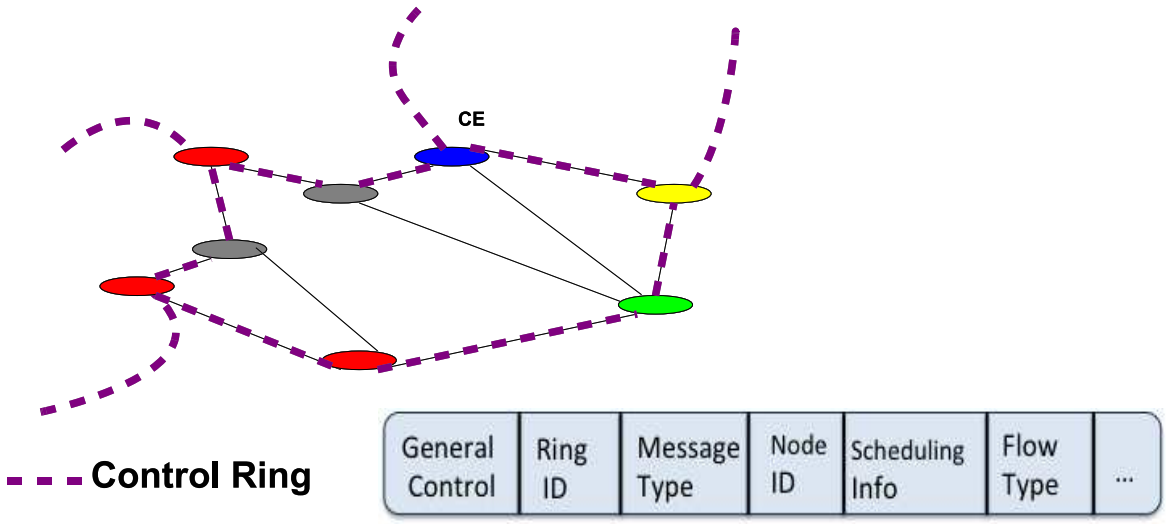


Figure 2.1 — Virtual control ring on a mesh network.

Fig. 2.1 shows a TWIN-based network with eight nodes on a physical mesh topology. The dashed line shows the control path that resembles a ring. This ring covers all the nodes through a dedicated wavelength that carries the control information. Thus, all the nodes within the network can exchange the reports/grants information between the CE and all the other nodes, in addition to other control and management messages via this dedicated path that we call Virtual TWIN Ring (VTR). The physical VTR links on which the information is carried are the preexisting TWIN links. However, these physical links do not mandatory form a ring topology, thus the name "virtual".

Note that using a mesh fiber infrastructure to support a ring architecture may lead to a rather large ring circumference, which may preclude the control messages propagation along the ring. For that reason, we extend the same control concept to multiple rings, considering a limit on the ring circumference. Furthermore, this approach allows us to limit the propagation delay within each VTR, which yields small propagation time for control packets having rapid reaction to failures (as these packets allow to report events and update node configuration). In Section 2.6, we propose the criteria to define the VTR and the algorithm that extends this concept to multiple VTR [81].

2.3 Clock synchronization

In order to be able to provide a non-blocking scheduling in TWIN, the network nodes should be synchronized, and CE needs to know about the distance, that can be ex-

pressed in the propagation delay between each source-destination pair. The propagation delay depends essentially on the path that is chosen on the network topology, therefore, if the path changes under any circumstances such as link failure, the distance needs to be recalculated. However, a mandatory step before propagation delay calculation is the clock synchronization [66].

Fig. 2.2 shows the clock synchronization implementing the master-slave mechanism. We assume bidirectional links where fibers per direction are in the same trunk, therefore, there is a negligible difference between the fibers length on each direction per link (i.e., the fibers of different directions on the same link are calibrated), with a similar environment effect. The CE exchanges the control messages with time-stamps with network nodes, by performing the following procedure.

Let $\mathcal{N} = N_1, N_2, ..., N_k$ be the set of nodes in a TWIN network. If the communication with CE is done via the control channel on the wavelength λ_C, and a node $N_m \in \mathcal{N}$ receives data on the wavelength λ_D, the propagation times t^{λ_C} and $t_m^{\lambda_D}$ corresponding to these wavelengths, respectively, might be different. This difference needs to be considered when calculating the propagation time between the node N_m and the other nodes in the network (Section 2.4).

Fig. 2.2 shows the time diagram of the control messages exchange when synchronizing the clock and calculating the propagation delay between the CE and the node N_m, marked with $t_{PD}(CE, N_m)$.

Here, N_n, $N_m \in \mathcal{N}$. Note that $t_{PD}(N_m, N_n) \neq t_{PD}(N_n, N_m)$, because of different light propagation speed at wavelengths received at N_m and N_n. Since the messaging between CE and N_m is always performed on the same wavelength (the control one), $t_{PD}(CE, N_m)$ is equal to $t_{PD}(N_m, CE)$, being valid for all nodes N_m.

In the beginning, only CE has the correct clock. In the first part of the procedure, the station N_m sets up its clock value to the CE clock value. This value will depend on the distance between N_m and CE. This part of the procedure is as follows:

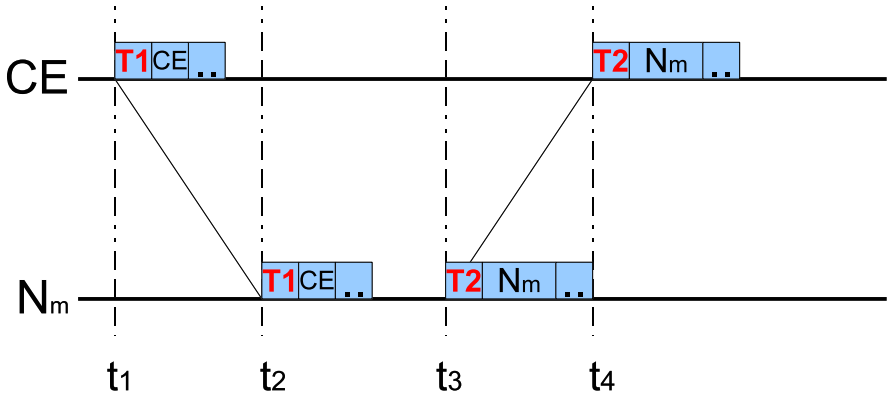


Figure 2.2 — Flow diagram for clock synchronization procedure.

- The CE sends on the control ring towards all the nodes N_n the OAM message that contains the stamp with the local time $T1 = t_1$.

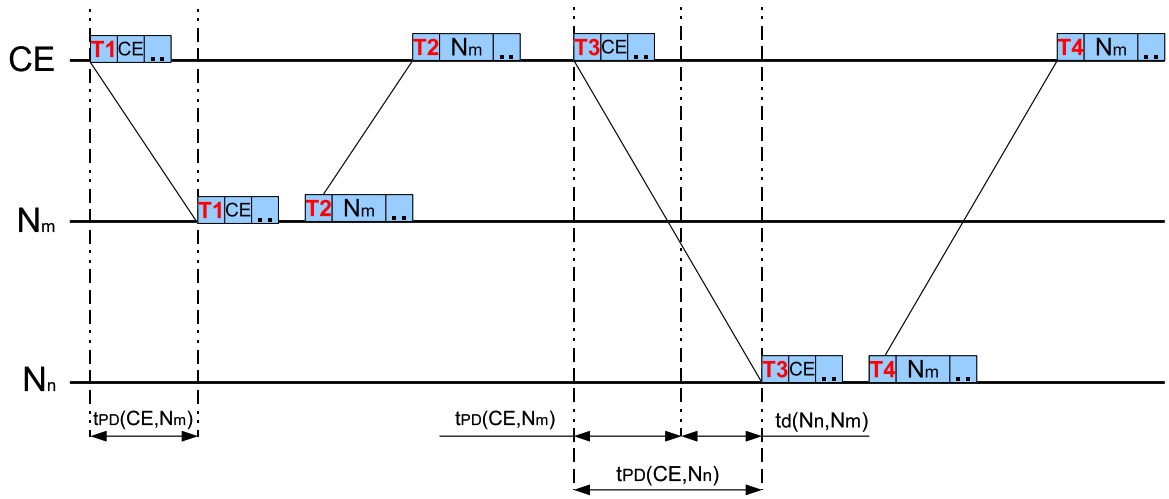


Figure 2.3 — Flow diagram for time difference calculation between nodes local clock.

- The nodes receive the time-stamps at the absolute time t_2 and set the clock to the CE local time $T1$.
- Each node within the network N_n sends back on the same fiber its local time $T2$ within an OAM message to CE on the control ring.
- The CE computes the propagation delay t_{PD} towards each node on the control ring based on the stamp received from N_n:

$$t_{PD}(CE, N_n) = (t_2 - t_1 + t_4 - t_3)/2; \tag{2.1}$$

$$T2 - T1 = t_3 - t_2; \tag{2.2}$$

$$t_{PD}(CE, N_n) = (t_4 - T2)/2; \tag{2.3}$$

- The CE computes the time difference t_d of the nodes local clock, as shown on Fig. 2.3, that is equal to the difference between propagation time from CE towards the nodes:

$$\Delta t_d(N_n, N_m) = t_{PD}(CE, N_n) - t_{PD}(CE, N_m); \tag{2.4}$$

To conclude, the CE is aware about the time difference between nodes local clock, that can be positive or negative, and it will be considered in the propagation delay calculation between all nodes in the network (Section 2.4).

The procedure shown in Figs. 2.2 and 2.3 is repeated for all the nodes.

2.4 Topology discovery

Fig. 2.4 shows the time diagram of the propagation delay calculation between all nodes in the network, assuming that the nodes are synchronized as described in the previous section:

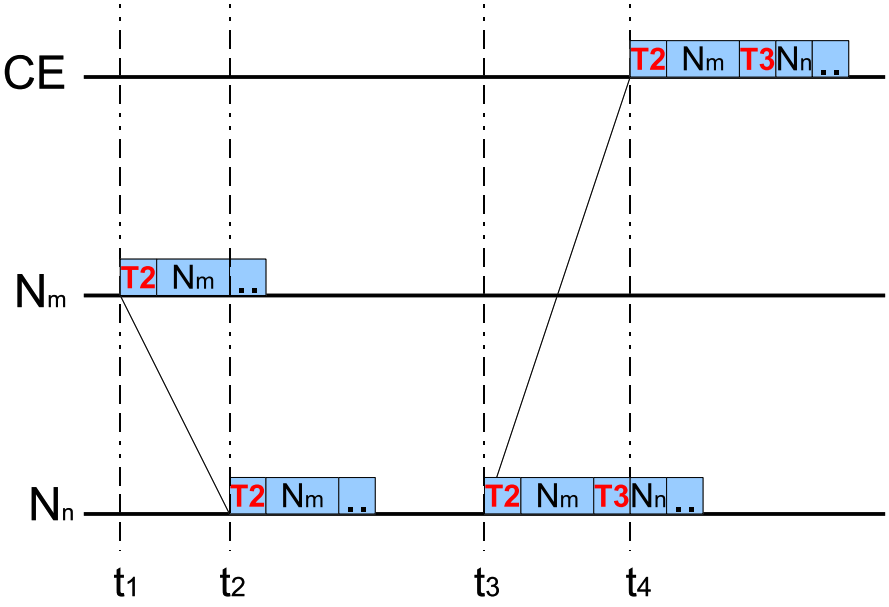


Figure 2.4 — Flow diagram for propagation delay calculation.

- The node N_m broadcasts the time-stamp $T2$ to all the nodes on the TWIN data path (to note that time-stamp is not broadcasted on the control ring!) within the in-band dedicated control channel.
- Each node within the network N_n sends its local time $T3$ at reception of the message from N_m, in addition to the time-stamp T2 received from N_m, within an OAM message to CE on the control ring.
- The CE calculates the propagation delay time between the node N_m and all the other nodes on the TWIN data path:

$$t_{PD}(N_n, N_m) = T3 - T2 + \Delta t_d(N_n, N_m) + (t_n^{\lambda_D} - t^{\lambda_C}); \tag{2.5}$$

In addition to the propagation delay and time difference between nodes local clock, CE takes into consideration the Chromatic Dispersion effect. The CE calculates the difference in the propagation time between the control channel $t^{\lambda C}$ and the data channel $t_n^{\lambda D}$ for each node N_n. The maximum value of this difference is equal to 600 ps/km, considering the wavelengths in C band (1530-1560 nm) and the typical fiber dispersion of 20 ps/nm/km [106]. Hence, having a synchronized network, the knowledge of the propagation delay and of chromatic dispersion map, the schedule is such that,

the arrival times of the packets have no conflicts. Accordingly, there are no conflicts anywhere in the network.

The procedure shown in Fig. 2.4 is repeated for all the node pairs in the network, in order to determine the propagation distances in the entire network.

However, if the time at the nodes is synchronized by Global Positioning System (GPS), then the above method should be slightly modified. In this case, the ranging procedure contains the following steps:

- The node N_m sends its local time T2 towards N_n.
- The node N_n sends its local time T3 to the CE, as soon as it receives the message from N_m. In addition, the node N_n will include the time-stamp T2 received from N_m, in its message.
- The CE calculates the propagation delay time between node N_m and N_n on the working topology, by using the following formula:

$$t_{PD}(N_n, N_m) = T3 - T2 + (t_n^{\lambda_D} - t^{\lambda_C}); \tag{2.6}$$

The above procedure is repeated for all node pairs in the network.

2.5 VTR for failure recovery

Fig. 2.5(b) shows the TWIN architecture working in the nominal operation mode with all nodes connected with multipoint-to-point TWIN trees. The same nodes are connected by dashed line control channel via a single bidirectional ring in Fig. 2.5(c). In case of failure, the affected sources and destinations will stop the flow on the nominal working path and start sending traffic on the backup path. Thus, the affected nodes uses different slot scheduling, computed statically (see Section 4.2.3) or dynamically.

However, in the case of failure, the following events take place between CE and affected nodes, in order to restore the network:

1. Failure Detection - optical signal loss is detected by fault manager entity of the two adjacent nodes, near both ends of the link failure in the Wavelength Division Multiplexing (WDM) layer. The failure alarms generated in WDM layer is processed at control/management layer to localize the link failure [104].
2. At CE, the propagation delays for the affected nodes are updated according to the above described procedure (Section 2.4).
3. The CE informs the affected nodes about routes that are no longer available and announces that the previous grants (schedules) are no longer valid.
4. At CE, the new schedules and grants are generated.

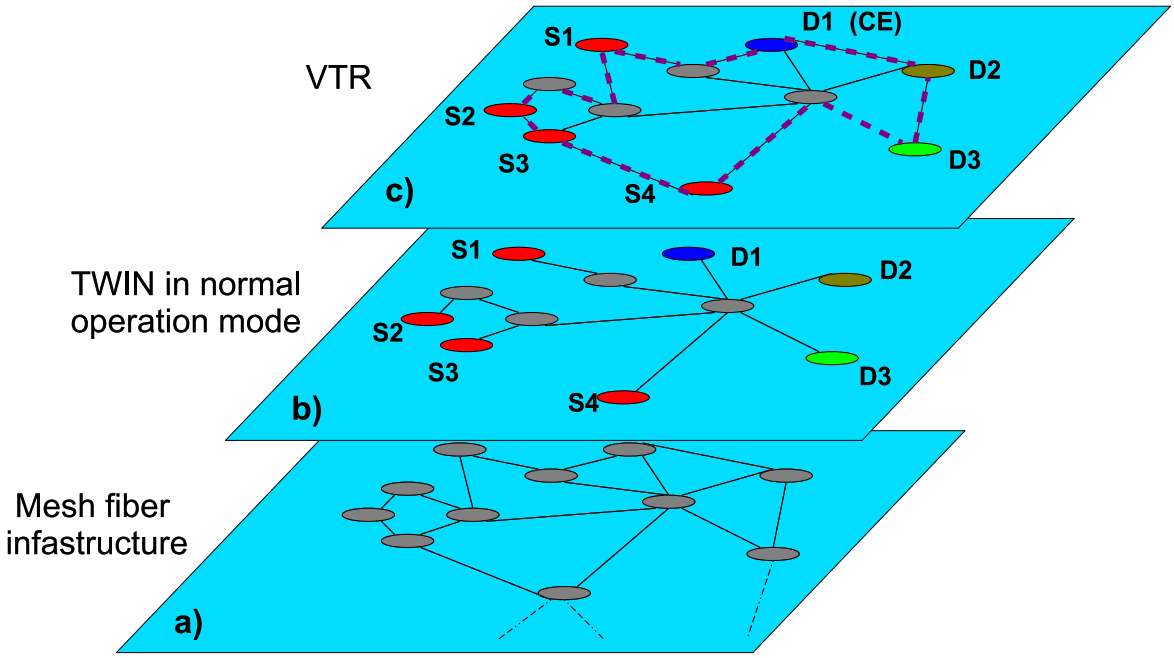


Figure 2.5 — Ring control method for TWIN architecture.

5. Source switches on the backup path, and uses the new grants (schedule).
6. Path Restoration - when the failure recovery signal is detected by the two nearest active/passive nodes, they inform the CE.
7. The nodes are notified by the CE that the previous grants for the backup paths are no longer valid.
8. The CE generates the new grants for the new reports and the regular working paths.

The simplification of the procedure is obtained at the cost of potentially increased propagation distance between some source destination pairs in the network. This effect is quantified in Section 2.6.

Control plane protection

The out-of-band control channel is carried over a single bidirectional ring or over multiple interconnected bidirectional rings in case of a larger network. Therefore, in an event of a link failure on the control path, the self-healing property of the ring topology and the protection procedure detailed in [79] can be used for the control channel recovery.

2.6 VTR algorithm

Here, we present the VTR algorithm (Alg. 1) that has no limitation in terms of number of nodes, and allows the extension of the proposed concept to a general transport

Input: $G(V, E)$, set $\mathcal{R}$ of candidate rings R_i with a circumference limited to C, network covering parameter α;
for *(MIN=1;MIN<= $|\mathcal{R}|$;MIN=MIN+1;)* **do**
  for *each subset $\theta_{ref} \subset \mathcal{R}$, such that $|\theta_{ref}|$ ==MIN* **do**
    if *each two rings in θ_{ref} are different in at least α nodes and all network nodes are covered by the rings in θ_{ref}* **then**
      θ_{ref} is found for $|\theta_{ref}|$ ==MIN;
      break the loops;
    end
  end
end
initialize the set of candidate solutions, $\Theta = \theta_{ref}$;
for *each subset $\theta_k \subset \mathcal{R}$, such that $|\theta_k| == |\theta_{ref}|$* **do**
  if *each two rings in θ_k are different in at least α nodes and all network nodes are covered by the rings in θ_k* **then**
    $\Theta = \Theta \cup \theta_k$;
  end
end
for *each set $\theta_k \in \Theta$* **do**
  for *each ring $R_i \in \theta_k$* **do**
    for *each ring node S_i ($S_i \in R_i$) communicating with nodes $\mathcal{D}(S_i) = \{D_1, D_2, ..., D_k\}$ outside and nodes $\mathcal{I}(S_i) = \{I_1, I_2, ..., I_l\}$ inside the ring R_i* **do**
      find the shortest path routes from ring R_i to each of nodes in $\mathcal{D}(S_i)$, if sent over the other rings in θ_k;
      calculate the propagation distance on the ring, $\Delta(S_i, D_j)$, between nodes S_i and D_j, for each $D_j \in \mathcal{D}(S_i)$, by using the shortest path routes calculated in the step above, and by accounting for the maximum path length, over both ring directions ;
      find the maximum distance route inside the ring R_i, $\Delta(S_i, I_j)$, between nodes S_i and I_j, for each $I_j \in \mathcal{I}(S_i)$;
    end
    calculate the sum of propagation distance for control in ring R_i:
    $\delta(R_i) = \sum_{S_i \in R_i} (\sum_{D_j \in \mathcal{D}(S_i)} \Delta(S_i, D_j) + \sum_{I_j \in \mathcal{I}(S_i)} \Delta(S_i, I_j))$;
  end
  calculate the average propagation distance for the solution θ_k:
  $\overline{\delta}(\theta_k) = \sum_{R_i \in \theta_k} \delta(R_i)/V(V-1)$;
end
for the final ring covering solution choose the set θ_k,

$$k = \arg\min_k \overline{\delta}(\theta_k);$$

Algorithm 1: VTR: Virtual TWIN Ring.

network. Depending on the size of the network the result may offer one or multiple number of interconnected rings to cover the whole network.

The input data for VTR algorithm are the graph of network topology $G(V, E)$ with set of vertices V and bidirectional edges E, and the so-called "network covering parameter" α, specifying the number of nodes that virtual rings are allowed to share. Instead of considering a particular traffic matrix, the solution considers the propagation distances between all node pairs and CE. Therefore, the optical signal reach constraint is accounted by limiting the maximum size C of rings used for network covering. C depends on many parameters (e.g., Quality of Service (QoS) requirements, maximum propagation delay), in particular the modulation format and baud rate used at transmitter side [4], [63]. Assuming a standard single mode fiber and a 50 GHz channel width, the conservative values could be e.g., $C = 500$ km at 200 Gb/s, and $C = 2500$ km at 100 Gb/s.

The VTR algorithm is based on a given and predefined set of candidate rings $\mathcal{R}$, constructed over a given network, considering the limit C for the ring circumference. The set $\mathcal{R}$ can contain any number of candidate rings, such that the rings of $\mathcal{R}$ cover all network nodes. The set $\mathcal{R}$ can be populated by resorting to the cycle finding algorithms, e.g., the one given in [44].

In the first step, the algorithm finds a reference "network covering" solution θ_{ref}. The network covering solution consists in a set of candidate rings that fully cover the transport network. For θ_{ref}, the algorithm finds the solution with the minimum number of rings.

In the next step, the VTR algorithm builds a set of all candidate solutions for network covering, i.e., the set Θ of all subsets of $\mathcal{R}$, which consist of exactly $|\theta_{ref}|$ rings, that provide full covering of the network. The set Θ is, in this way, restricted only to solutions with the minimum possible number of rings. The advantages of such approach are not only in reducing the algorithm complexity, but also in finding the network covering that is easier for administration, because of the minimized number of rings.

Then, for each candidate solution θ_k, a procedure is applied for calculating the average propagation distance of this solution, over all source destination pairs and CE in the network. To do so, for each ring $R_i \in \theta_k$, each node S_i is separately examined for the traffic this node exchanges with all the nodes $\{D_1, D_2, ..., D_k\}$ outside and inside $\{I_1, I_2, ..., I_l\}$ the ring. For all the exterior nodes, the distance to R_i is calculated according to the shortest path rule, the corresponding propagation distances between the node S_i and the outside nodes are calculated.

At the end of this step of the VTR algorithm, the overall sum $\delta(R_i)$ of the propagation distance is calculated, for each ring R_i. For all the interior nodes, the propagation distance is taken to be the maximum possible, to account for the failure. Finally, by summing the values of $\delta(R_i)$ over all rings R_i and by calculating the mean value, an average of propagation distance for θ_k, $\overline{\delta}(\theta_k)$ is found.

For the final solution, the VTR algorithm chooses $\theta_k \in \Theta$ with the minimum value of $\overline{\delta}(\theta_k)$. In the case where several candidate solutions θ_k have very close average propagation distances (e.g., with relative difference $< 1 - 2\%$), the ties are broken in favor of the solution with minimum number of source-destination paths exceeding C km.

Obviously, the VTR algorithm always finds a solution, if the size of the predefined candidate ring set $\mathcal{R}$ is large enough. By choosing $\mathcal{R}$ and α, we can adjust the accuracy of the algorithm solution, but also the speed of calculating the solution.

2.7 Application of VTR in a real network

To illustrate the operation of the VTR algorithm, in this section we consider the Deutsche Telekom network in Fig. 2.6, composed of 17 nodes and 26 links. In the figure all the link lengths are expressed in km.

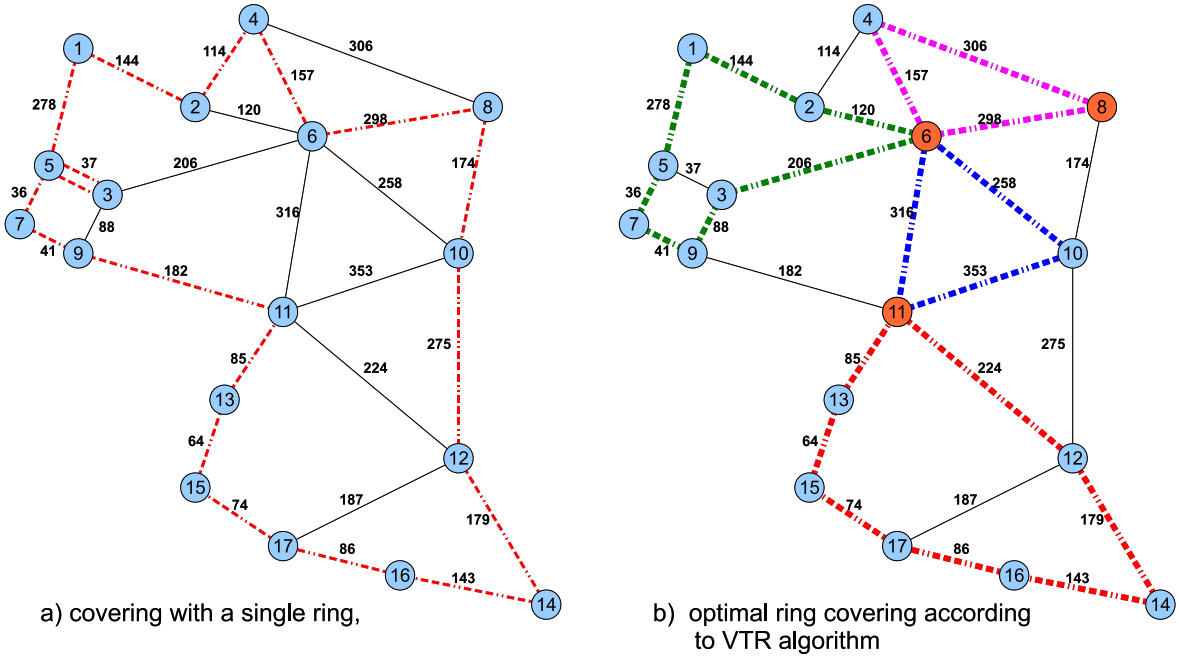


Figure 2.6 — Deutsche Telekom network.

According to the control scheme, we proposed that, all the nodes should be covered by at least one bidirectional ring, and should meet the QoS requirements of maximum 5 μs of propagation delay on the ring [90]. The minimum single ring which covers the entire network is shown in Fig. 2.6(a). The size of this ring is 2404 km, which greatly surpasses the QoS requirements and/or optical reach of signal e.g., at 100 Gb/s and 200 Gb/s, and thus, is not an acceptable solution [4], [63].

To find the network covering with multiple virtual rings, the upper bound for the control and protection recovery ring circumference is set to $C = 1000$ km, while the network covering parameter $\alpha = 1$.

The set of candidate rings, found by the VTR algorithm is given in Tab. 2.1. There are 13 potential candidate rings, each of them have a circumference limited to C. Tab. 2.1 shows the nodes contained by each ring, and their circumference in km.

In the initialization phase of the VTR algorithm, for the referent ring covering we set: $\theta_{ref} = \{R_1, R_2, R_3, R_4\}$. It is easy to check that 4 is the minimum number of

Table 2.1 — Set of candidate rings.

Candidate ring	Contained Nodes	Circumference [km]
R_1	7 , 9 , 3 , 6 , 2 , 1 , 5	913
R_2	4 , 6 , 8	716
R_3	6 , 10 , 11	927
R_4	11 , 13 , 15 , 17 , 16 , 14 , 12	855
R_5	4 , 2 , 6 , 8	838
R_6	4 , 6 , 10 , 8	895
R_7	2 , 6 , 10 , 8 , 4	972
R_8	10 , 11 , 12	852
R_9	1 , 2 , 6 , 3 , 5	785
R_{10}	7 , 9 , 11 , 6 , 3 , 5	818
R_{11}	6 , 8 , 10	730
R_{12}	1 , 2 , 4 , 6 , 3 , 5	936
R_{13}	3 , 9 , 11 , 6	792

Table 2.2 — Set of candidate solutions.

Candidate solution	Contained rings	Number of paths with length >1000 km	Average Propagation distance $\overline{\delta}(\theta_k)$ [km]
$\theta_{ref} = \theta_1$	R_1,R_2,R_3,R_4	8	496
θ_2	R_1,R_3,R_4,R_5	8	500
θ_3	R_1,R_4,R_6,R_8	40	563
θ_4	R_1,R_4,R_7,R_8	42	565
θ_5	R_1,R_3,R_4,R_7	6	498
θ_6	R_1,R_3,R_4,R_6	6	495
θ_7	R_4,R_6,R_9,R_{10}	4	435
θ_8	R_4,R_7,R_9,R_{10}	4	437
θ_9	R_4,R_{10},R_{11},R_{12}	2	436
θ_{10}	R_4,R_6,R_{10},R_{12}	4	444
θ_{11}	R_4,R_7,R_{10},R_{12}	4	431
θ_{12}	R_1,R_4,R_6,R_{13}	4	440
θ_{13}	R_1,R_4,R_7,R_{13}	4	443

rings needed to cover the studied topology. Tab. 2.2 summarizes the set of candidate solutions (with 4 rings) that the algorithm has found. Therefore, the number of the candidate solutions is 13.

Control with VTR simplifies the control plane for synchronization and failure recovery. On the other hand, the control messages might be penalized with an increased path, in comparison to a mesh network control scheme. The number of paths exceeding 1000 km, for each of the candidate ring coverings is given in Tab. 2.2.

These results show that choosing θ_3 or θ_4 would have bad impact on the control performance. We can see that the minimum number of paths exceeding 1000 km is provided by solution θ_9, so this solution might be convenient.

The VTR algorithm has found that θ_7, θ_8 θ_9, θ_{11} have very close average propagation distances (with the relative difference less than 2%). However, the solution θ_9 (presented in Fig. 2.6(b)) has the minimum number of paths exceeding $C = 1000$ km (as shown in Tab. 2.2), so, it is selected as the final result of the algorithm.

2.8 Conclusion

In this chapter we have proposed a new method for the control and the synchronization in a Time-Domain Wavelength Interleaved Network (TWIN) optical burst switching network. This method facilitates the TWIN's control and operation with an out-of-band control channel, carried on one or multiple Virtual TWIN Rings (VTRs) that cover all the nodes in the network. The control scheme is centralized and the control channel accommodates the operational messages regarding the resource requests and allocations, in addition to the functional information related to the synchronization, the network configuration and failures, all of them exchanged between the nodes and the central Control Entity (CE).

To sum up, the proposed control scheme for TWIN relies on the virtual rings, while the data paths are the regular TWIN trees, within the mesh topology.

Since TWIN architecture relies on a precise knowledge of propagation delays between source-destination pairs, the main benefit of the proposed solution is to provide combined means of control, protection recovery and synchronization.

We have proposed and implemented a heuristic algorithm to choose the control rings, with a maximum ring circumference as an input parameter, with the purpose of keeping the total number of virtual rings as low as possible. Indeed, limiting the number of rings makes the passage of the administrative and control messages less complex and limits the overhead traffic on the control channel between the rings.

By using the algorithm, the mentioned features are ensured with the shortest propagation distance on the rings.

Finally, we have applied the heuristics on Deutsche Telekom's mesh topology with realistic distances. For that particular network with 17 nodes and 26 links, a minimum of 4 interconnected rings with less than 1000 km circumference is needed, in order cover all the nodes and ensure the control with limited overhead traffic.

The work presented in the following chapters (i.e., Chapter 3 and 4) has the focus on the development of the optimization solutions taking into account the realistic traffic matrices and capable of dimensioning the TWIN network, and optimizing the protection capacity on the mesh networks, respectively.

CHAPTER 3

Design of Time-Domain Wavelength Interleaved Network

In this chapter, we tackle the design problem of Time-Domain Wavelength Interleaved Network (TWIN). We propose two ways of improving the slot allocation efficiency in such networks: by minimizing the schedule length and by enforcing a cyclical Quality of Service (QoS) constraint on the allocated slots. We report the detailed simulation results on how network topology, traffic pattern and the proposed QoS improving method impact the network performance. In addition, for the first time, we develop a 0-1 Integer Linear Programming (0-1 ILP) dimensioning tool which simultaneously addresses the Routing and Wavelength Assignment (RWA) and scheduling (i.e., resource allocation) problem in TWIN. We propose a comprehensive scalable algorithm which assesses the same problem and, in addition, considers network virtualization. The algorithm enables the construction of overlayed virtual networking domains at minimum cost, which share the transponders and network links, and can employ different scheduling policies. The performance of the algorithm is compared with the 0-1 ILP solution for single virtual domain. Furthermore, the study is extended to a real size network.

3.1 Introduction

This chapter considers different aspects of Time-Domain Wavelength Interleaved Network (TWIN) design. In Section 3.2, we consider the problem of static slot allocation, as in [75] or [91], while also considering Quality of Service (QoS) (latency) aspects, as in [98]. Hence, contrary to other works, Section 3.2 investigates, at the same time, the problem of finding a minimum schedule length while considering latency constraints [30].

Section 3.3 considers the problem of TWIN cost dimensioning. The dimensioning tool is based on 0-1 Integer Linear Programming (0-1 ILP) and simultaneously solves two sub-problems: the static slot allocation and the Routing and Wavelength Assignment (RWA).

In Section 3.4, we propose an algorithm that as primary concern provides a scalable solution for the RWA and scheduling (resource allocation, term interchangeable with anterior) optimization. In addition, for the first time, the proposed algorithm allows the creation of the arbitrary number of virtual domains over the same physical TWIN network.

Finally, Section 3.5 concludes the chapter and highlights the significant contribution of this work.

3.2 Optimal scheduling

The central issue in a TWIN network is the scheduling of the fixed-duration bursts (slots). The slots are classified according to their destination, prior to the insertion on the fiber medium, in order to avoid the Head-Of-Line (HOL) blocking effect. Scheduling is important as it determines not only the network efficiency in terms of the capacity utilization, but also, the QoS (latency) performance of each traffic flow routed in the network. Thus, scheduling consists in allocating the time slots to source-destination flows, in a way that each traffic demand is supported by a sufficient number of slots, according to its capacity.

One of the goals of the optimization method given in this section is to allocate slots in a way that the maximum waiting time of a HOL packet is limited to a wanted number of time slots, in order to improve the latency of each traffic flow.

The scheduling in TWIN has been extensively studied by different research groups. Paper [76] studied the static allocation of the slots with a minimum schedule length, while in [75] the authors have shown how the slots can be dynamically distributed. Scheduling was also studied in [105], with a focus on multimedia traffic, and congestion management in [5]. More recently, in [84], the authors consider distributed scheduling with static allocation, for both dynamic and semi-dynamic traffic. A detailed description of the TWIN architecture and scheduling issues is given in the Sections 1.2.3 and 1.3.3 of the state of the art chapter.

In [91], a static slot allocation is proposed for a centralized control plane, and its well performance in comparison to distributed, non-static allocation is demonstrated.

Medium Access Control algorithm for TWIN, based on Passive Optical Network (PON) principles is proposed in [74]. In [98], the authors proposed a 0-1 ILP-based planning and dimensioning algorithm for a TWIN network with an arbitrary number of transmitters and receivers per node, including QoS (latency) constraints.

In this section, we consider the problem of static slot allocation, as in [75] or [91], while also considering QoS (latency) aspects, as in [98]. Hence, contrary to other works, this section investigates, at the same time, the problem of finding a minimum schedule length while considering latency constraints [30].

Contrary to Section 3.3.2 and 4.2.3, in this section, we are not focusing on the dimensioning (transponder allocation, RWA) problem, but rather, we consider a plain, low-cost TWIN network with a single transponder per node and focus on scheduling issues. (More complex nodes can easily be built using several single transponder nodes.) We also prove, for the first time, the tightness of a known bound on the minimum schedule length under certain conditions.

The remainder of the section is organized as follows. Subsection 3.2.1 is devoted to our theoretical contributions. Subsection 3.2.3 describes an algorithm that finds a minimal schedule under latency constraints. The following subsections are devoted to detailed numerical results and the last subsection presents a brief conclusion.

3.2.1 Theoretical contribution on minimum schedule length

We define the slot allocation efficiency as the ratio between the number of slots used to carry traffic to the total number of slots present in the network, at a given instant.[1] Slot allocation efficiency depends on the input traffic matrix and on the node distances between adjacent nodes.

In the current section we consider that the node is equipped with a single tunable transmitter and a fixed wavelength receiver. The traffic matrix is static and the schedule is semi-static and cyclic, i.e., the same schedule is repeatedly used by all nodes until it is updated; the schedule can be computed by a central node and updated when some of the problem inputs (e.g., the demand between two nodes) change. In practice, a slot is a few to a few dozens of microseconds long, the schedule length is a few dozens to a few hundreds of slots, and the schedule could be recomputed every few milliseconds, seconds, minutes, or hours, as needed. It is assumed that each node is able to receive the schedule from the central controller, for instance, using a dedicated out-of-band control channel, as described in Chapter 2.

We measure link lengths in terms propagation time; indeed the distances can easily be converted into the number of slots. We assume further that link lengths are multiples of the slot duration, such that the allocated slots belonging to different connections (i.e., source-destination pair) do not overlap at the source or at the destination. Such

[1] The problem of aggregating incoming traffic frames into optical slots is out of the scope of this work. In particular, we are not accounting for the average slot filling efficiency, i.e., the ratio between the capacity used in a slot to the slot size in bits, which depends on the aggregation process; the network efficiency, i.e., the ratio of the number of utilized capacity to the available capacity in the network, is simply the product of the slot allocation efficiency with the slot filling efficiency.

constraint can be enforced through proper engineering of the transmission lines, i.e., by using suitable fiber delay lines on each network link, or using two TRX at each source. The effect of real propagation delay is considered in Section 3.3.3.

Let T be the traffic matrix ($T_{s,d}$ is the demand from s to d in Gb/s; $s, d \in V$), and N the number of time slots, on which the scheduling is calculated (by using one of the algorithms from the next subsection). Thus, N is the "schedule length". As in [75], let $n_{s,d}$ be the number of slots, arriving during an interval of duration N slots, and corresponding to traffic between source s and destination d:

$$K^s(N) = \max_s \sum_{d=1}^{|V|} n_{s,d}, \qquad \forall s \in V; \tag{3.1}$$

$$K^d(N) = \max_d \sum_{s=1}^{|V|} n_{s,d}, \qquad \forall d \in V;$$

$$K(N) = \max(K^s(N), K^d(N)); \tag{3.2}$$

$K^s(N)$ is the maximum over all nodes of the number of slots during which the transmitter of node s should be active whereas $K^d(N)$ is the maximum over all nodes of the number of slots during which the receiver of node d actually receives a packet. Definitely, if $K(N) > N$, it is impossible to schedule a TWIN system with single receiver/single transmitter. $K(N)$ is thus a lower bound for N.

We first show that if all propagation delays are null, it is possible to schedule all traffic in $K(N)$, i.e., the bound is tight (Theorem 1). Then, we relax the zero propagation delay assumption to an equal propagation delay (Corollary). We also show that the bound is tight if the load is large (Theorem 2).

Theorem 1. *Let us suppose that the maximum number of bursts to send/receive by a source or destination in such network is K. Let us suppose that there is zero propagation delay between the nodes. Then, the schedule can be done in K time slots.*

Proof. In order to prove this, we will suppose, without loss of generality, that the maximum traffic sent by a source is greater or equal than/to the maximum traffic received by a destination. (If this is not the case, the same proof holds in case if the destinations becomes the sources and vice versa.) We propose a proof by induction.

Let us suppose that $K = 1$, i.e., each source can have at most one burst to send to some of the destinations. Since the propagation between the nodes are all zero, these bursts could be sent in exactly 1 time slot, since the maximum number of slots received by destinations is not greater than 1. Thus, the theorem holds for $K = 1$.

Let us suppose that the theorem holds for K, i.e., when the maximum number of bursts that need to be sent is K, then the scheduling of these bursts is possible in K time slots. Let us observe the same system of sources and destinations, with $K + 1$ bursts to send by a most loaded source.

Let us suppose that there are S sources with $K+1$ bursts to send. The maximum number of bursts a destination receives is not greater than $K+1$. Let us suppose that there are D destinations with $K+1$ bursts to receive. Let us suppose that at each source S or destination D, traffic load is reduced to K bursts per source/destination, with an additional 1 burst left aside. The "reduced traffic load" can be sent/received in K time slots, according to the mathematical induction hypothesis. During each of K time slots, there is a matching, including all the S sources and D destinations, in bipartite graph composed of set of sources and set of destinations. Indeed, if no matching would exist, then it would not be possible to schedule K bursts of each source S and destination D, in exactly K time slots.

Let us choose one such matching M. (The matching M allows to each of S sources and D destinations to send/receive 1 burst, with no source or destination sending/receiving more than a single burst.)

If we come back to a "complete traffic load", with maximum of $K+1$ slots to be send by S sources and D destinations, the scheduling could be done as follows: firstly, we schedule matching M in first time slot. The maximum number of bursts to be sent/received by S sources/D destinations falls to K. According to the mathematical induction hypothesis, the remaining traffic can be scheduled in K time slots, meaning that the "complete traffic load" can be scheduled in $K+1$ slots.

Corollary. *Let us suppose that the maximum number of bursts to send/receive by a source or destination in such network is K. Let us suppose that the propagation distances between the nodes are equal. Then, the schedule can be done in K time slots.*

This is a direct consequence of the previous theorem; actually, all packets shall arrive with the same delay to the destinations, but the schedule can be computed as if all delays were null.

Theorem 2. *If the overall traffic that is received or sent by a node is so large that it becomes almost equal to the channel capacity R, the minimum schedule length N, for which the scheduling is possible, becomes equal to K(N).*

Proof. Recall that $K(N)$ is equal to the sum of traffic (number of optical slots) arriving or departing from the most congested destination/source. Thus:

Denote the total number of slots sent (t_k) or received (r_k) at node i during the schedule, as:

$$t_i = \sum_{d=1}^{|V|} n_{i,d}, \qquad \forall i \in V; \tag{3.3}$$

$$r_i = \sum_{s=1}^{|V|} n_{s,i}, \qquad \forall i \in V; \tag{3.4}$$

By definition of $n_{s,d}$, we have:

$$t_i = \sum_{d=1}^{|V|} \lceil N \cdot T_{i,d} \rceil, \qquad \forall i \in V; \tag{3.5}$$

$$r_i = \sum_{s=1}^{|V|} \lceil N \cdot T_{s,i} \rceil, \qquad \forall i \in V; \tag{3.6}$$

Suppose, without loss of generality, that $\max_i t_i \geq \max_i r_i$ (the proof can easily be adapted if $\max_k t_k \leq \max_k r_k$). Let $i' = \arg\max_i t_i$ be the index of the node that sends the largest number of slots. By definition we have:

$$K(N) = \sum_{d=1}^{|V|} \lceil N \cdot T_{i',d} \rceil, \qquad \forall i \in V; \tag{3.7}$$

Note that, using a triangular inequality:

$$K(N) = \sum_{d=1}^{|V|} \lceil N \cdot T_{i',d} \rceil \geq \lceil N \cdot \sum_{d=1}^{|V|} T_{i',d} \rceil, \qquad \forall i \in V; \tag{3.8}$$

When the traffic increases we have:

$$\sum_{d=1}^{|V|} T_{i',d} \to 1, \qquad \forall i \in V; \tag{3.9}$$

and the right-hand side of Eq. 3.8 tends to N, i.e., $K(N) \geq N$. As in the same time, we know from the definition of K(N) that $K(N) \leq N$, we conclude that for very high traffic loads at a source/destination, the only possible solution for minimum schedule length is $N = K(N)$.

We propose the `Minimum Schedule Length (MSL)` algorithm to find the minimum schedule length. As will be seen, a key subroutine of `MSL` consists of a 0-1 ILP formulation, which we call `SCHEDULE`. We present those algorithms in the following subsections.

3.2.2 Notations for 0-1 Integer Linear Programmes

Firstly, in this subsection, we define the index notations of the 0-1 ILP formulations that are given in this work (Tab. 3.1). If, in a particular 0-1 ILP formulation a given index is different in respect to the notation defined here, it will be redefined again in the corresponding section.

In the following sections (i.e., Sections 3.2.3, 3.3.2 and 4.2.3) given parameters and output variables will be defined for each particular 0-1 ILP formulation.

Table 3.1 — Index notation.

i	generic node index;
j	generic node index;
w	wavelength index;
k	slot index;
ℓ	link index;
r	path index;
b	backup path index;
s	source node index.
d	destination node index.

3.2.3 General scheduling 0-1 ILP: SCHEDULE

The `SCHEDULE` 0-1 ILP formulation described below returns a schedule S for a TWIN network, given a network graph G, a traffic demand T and a cyclic QoS parameter $\beta_{s,d}$, as formalized below. The problem of computing an optimized schedule is a NP complete problem [91].

We assume that routing is static and that a route is pre-computed and assigned to each (s, d) pair. Alternate routing can easily be added to the formulation, for instance the 0-1 ILP formulations presented in Sections 3.3.2 and 4.2.3, support multiple possible paths from a source s to a destination d.

Given Parameters

- $G(V, E)$: a non-directed graph describing the physical mesh topology, where V is the set of nodes, E is the set of links. $|V|$ is the number of nodes in the network;
- T: traffic matrix: element, $T_{s,d}$, is the traffic rate (≤ 1, i.e., normalized to single wavelength capacity) sent by the source s to the destination d;
- $D_{s,d}$: propagation delay (in number of slots) on pre-computed route experienced by a slot emitted by the source s until it reaches the destination d;
- N: schedule length;
- $\beta_{s,d} \leq N - 1$: maximum latency of a HOL packet (expressed as a number of time slots) of a demand from node s to node d.

Output Variables

- binary $S^k_{s,d}$ equal to 1 if slot k is used for communication from source node s to destination node d, and 0 otherwise; S is the schedule.

Notation

- $C_{k,n}$: set of all subsequences of length k of $[1, \ldots, n]$ or of a circular shift of $[1, \ldots, n]$; in other words: $C_{k,n} = \{[(m \mod n)+1, ((1+m) \mod n)+1, \ldots, ((k-1+m) \mod n)+1] : 0 \leq k \leq n-1\}$.

0-1 ILP formulation

Objective function: when we are given a schedule length, we try to allocate as many slots as possible in order not to leave any slot empty, which would correspond to wasted capacity:

$$\max \sum_{s=1}^{|V|} \sum_{d=1}^{|V|} \sum_{k=1}^{N} S_{s,d}^{k}; \tag{3.10}$$

Constraint ensuring that only one slot is used at the same time toward the same destination: denoted by $k'_{s,d,k} = (k - D_{s,d}) \mod N$, the slot number in which slot k arriving at node d was emitted from node s:

$$\sum_{s=1, s\neq d}^{|V|} S_{s,d}^{k'} \leq 1, \qquad \forall d \in V, \forall k \leq N; \tag{3.11}$$

Constraint ensuring that at most one slot is emitted from the same source, during any time slot:

$$\sum_{d=1, d\neq s}^{|V|} S_{s,d}^{k} \leq 1, \qquad \forall i \in V, \forall k \leq N; \tag{3.12}$$

Constraint ensuring that to each source-destination pair is allocated atleast the demanded capacity:

$$\sum_{k=1}^{N} S_{s,d}^{k} \geq N \cdot T_{s,d}, \qquad \forall s \in V, \forall d \in V, \text{ s.t. } s \neq d; \tag{3.13}$$

Constraint ensuring that there is at least a busy slot from s to d within $\beta_{s,d}+1$ consecutive slots, i.e., the latency of a HOL slot is capped by $\beta_{s,d}$ for a demand from s to d:

$$1 \leq \sum_{k \in C'} S_{s,d}^{k} \quad \forall C' \in C_{\beta_{s,d}+1,N}, \forall s \in V, \forall d \in V \text{ s.t. } T_{s,d} \neq 0; \tag{3.14}$$

Note that the proposed 0-1 ILP `SCHEDULE` does not attempt to minimize the schedule length; rather, the schedule length is an input to `SCHEDULE`.

3.2.4 Algorithm `MSL`: finding the minimum schedule length

In order to find the minimum schedule length, we propose the greedy algorithm `MSL` (Alg. 2), which iteratively attempts to schedule a given demand in a given TWIN network starting with a small schedule length, by increasing the schedule length and trying to obtain a scheduling solution with the `SCHEDULE` 0-1 ILP until a solution is found. In the examples studied in this section, the running time of a single iteration of `MSL` algorithm was typically within few minutes.

REQUIRE: same inputs as 0-1 ILP `SCHEDULE`;
initialize the allocation window: N is equal to the maximum number of sources a network destination has;
while *Eq. 3.1* **do**
  calculate $K(N)$ based on N (use $n_{s,d} = \lceil T_{s,d} * N \rceil$);
  if *N≥K(N)* **then**
    use 0-1 ILP `SCHEDULE` on T and N;
    if *(there is a feasible solution)* **then**
      BREAK;
    else
      N=N+1;
    end
  end
end
RETURN(N);

Algorithm 2: Algorithm `MSL`.

3.2.5 Impact of propagation distance on schedule length

In case of non-equal propagation delay, it is not guaranteed that there is a schedule of length $K(N)$. However, we show here that minimizing the schedule length can lead to the improvement of the QoS performance in the network, such that the schedule length minimization is indeed desirable, by using simulations with both an LP solver [34] used for scheduling calculation, and an ns-2-based event-driven simulator [89] used for latency measurements. It is assumed that the QoS parameters have equal values for all source-destination pairs, i.e., $\beta_{s,d} = \beta$. Without loss of generality, we suppose that the transmitter (and receiver) capacity is $R = 10$ Gb/s. The slot duration is equal to 10 μs. Unless stated differently, in the following examples we consider $\beta = N - 1$, i.e., we relax the QoS constraint.

Firstly, we are interested to evaluate the impact of the network propagation distances, i.e., topology, on the minimum schedule length. The topologies were generated randomly by choosing 5 random node locations in a x-y two-dimensional plane, with coordinates uniformly drawn between 0 and 100 slot durations[2], the fiber connectivity between the nodes remains unchanged and it corresponds to the topology from Fig. 3.2. Each simulation point is an average over 20 random topologies. We consider uniform and symmetric demands $T(s, d) = a$ between all 5 nodes ($s \neq d$). The maximum value for a is 0.25 (normalized to the channel capacity R) and each source/destination sends/receives a traffic of $4aR$.

We run an ILP program to find the minimum value N on which we can schedule $K(N)$. Note that in our program, we attempt to schedule at least the traffic specified by T, but we also schedule more slots if possible.

The optimal results of simulations, obtained by `MSL` algorithm and optimal solution based on 0-1 ILP formulation, are summarized in Fig. 3.1. The figure shows that with

[2]The distance between the network nodes can be expressed in terms of slot durations.

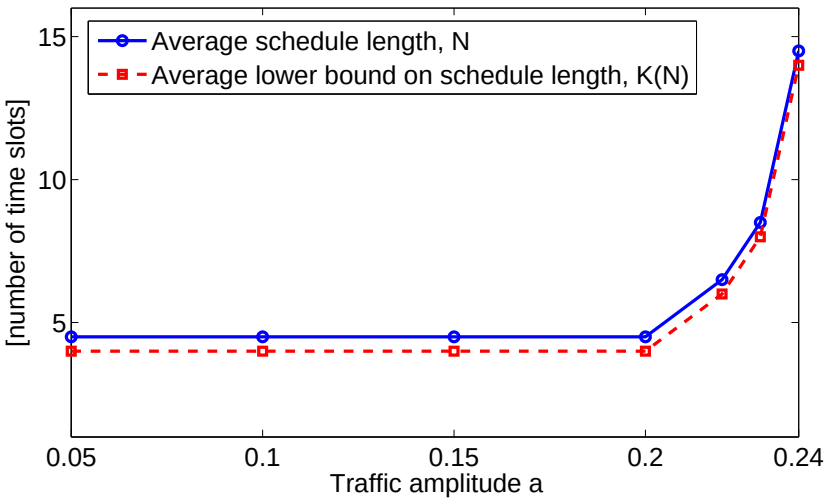


Figure 3.1 — Impact of propagation distances and traffic load on the schedule length.

the increase in traffic load, the size of the minimum schedule length N increases as well. We can also see that the propagation distances have an impact on N. Indeed, for a topology without propagation delays between the nodes, there would not be a gap between the two curves in Fig. 3.1. However, the impact of the topology is very mild, and could be even neglected. Thus, we observe that $K(N) \leq N \leq K(N) + 1$.

With the increase of traffic amplitude a, the gap between N and $K(N)$ decreases (as for more and more points participating in the average score presented in the figure, we have $N = K(N)$). For instance, in point for $a = 0.24$, each station sends/receives a traffic of 9.6 Gb/s, i.e., it uses 96% of its transmitter capacity. According to Theorem 2, $N \approx K(N)$ in this case, which is confirmed in Fig. 3.1.

3.2.6 Impact of traffic pattern on schedule length

In this example, and in the remaining simulations, the network is a geometrical rhombus (Fig. 3.2), with the exterior edge lengths of $b = 3$ time slots. In this topology, the propagation times between node pairs are different (all edge size are rounded to the entire number of slots), and are calculated according to the shortest path between the node pairs.

Here, the focus is on studying the impact that different traffic patterns have on the TWIN's minimum schedule length N. We study the uniform and symmetric traffic, as in the previous case, with an amplitude a, as well as the randomly generated traffic. We generated randomly the traffic between each source-destination node pair, by resorting to an approximate uniform distribution. Firstly, we generate the traffic according to the uniform distribution with the average μ (normalized to total transponder capacity). Then, when needed, we would "biased" the distribution, in order to limit the sum of overall sent/received traffic per station to 100% of the transponder capacity. For

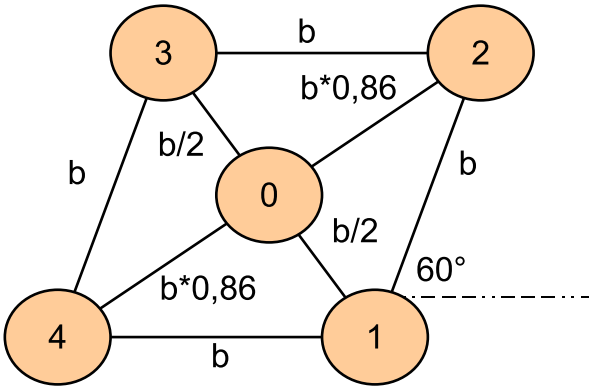


Figure 3.2 — Rhombus with edge size b.

random traffic, the results are averages over 20 different input matrices.

With this setup, the results for minimum schedule length, obtained by use of `MSL` algorithm, are given in Fig. 3.3 (results of `MSL` algorithm). We see clearly that in the random scenario, the minimum schedule length N is larger than $K(N)$ only for low network loads. For uniform and symmetric traffic, N is always strictly greater than $K(N)$. The difference between the last two is very low in both cases, which is due to the chosen topology and traffic pattern.

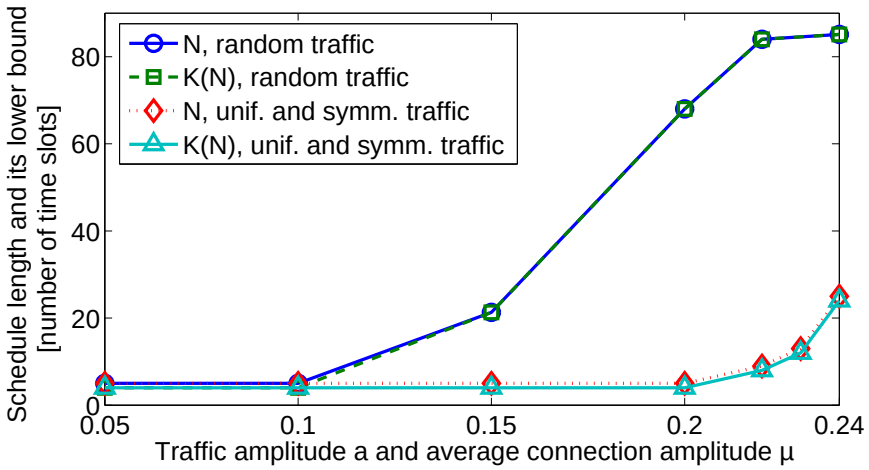


Figure 3.3 — Impact of traffic patterns on schedule length.

The random case is more impressive, as the minimal schedule length increases significantly with the offered traffic. The results have shown that for less symmetric traffic patterns, the minimum schedule length will likely be much higher, than in purely uniform and symmetric scenario.

Recall the definition of slot allocation efficiency at the beginning of Subsection 3.2.1; given the notation introduced in Subsection 3.2.3, the slot allocation efficiency is equal

to:

$$\sum_{s=1}^{|V|}\sum_{d=1}^{|V|}\sum_{k=1}^{N} S_{s,d}^{k}/(N\cdot|V|); \tag{3.15}$$

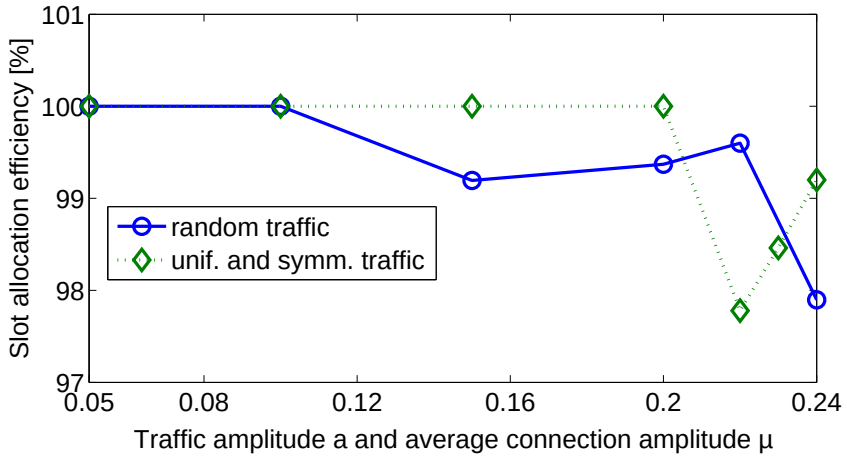


Figure 3.4 — Slot allocation efficiency.

Fig. 3.4 shows the slot allocation efficiency for two traffic profiles (results of `MSL` algorithm). Definitely, the slot allocation efficiency is lower for higher loads, which is expected, since the slot allocation is more and more complex due to the increase of the traffic. The decrease is non-monotonous, since there is no correlation between the LP solver outputs for the neighboring traffic loads. Note also that even for very high wavelength occupancy ratios (e.g.,$a = \mu = 0.24$), the slot allocation efficiency is very high (above 98% of available capacity is used).

3.2.7 Impact of schedule length on latency performance

In the previous subsection, we have shown how the minimum schedule length can be calculated, by use of the `MSL` algorithm, and that both the topology and traffic profile can have an important impact on minimum N.

Here, we study by simulations the performance of network for different schedule lengths. We report latencies (insertion delay) obtained by the use of Network Simulator, ns-2, which was enhanced by the additional functions to support the TWIN slot allocation. The results are at 95% confidence level and with 10% confidence interval. The slot allocations used as the input for ns-2 simulations are either the optimal 0-1 ILP solutions, or they are obtained with a 5% optimality gap. The slot arrival distribution is Bernoulli, and the slot duration is 10 μs.

We consider an asymmetric traffic matrix, obtained from the uniform and symmetric traffic pattern simulated previously, by forcing node 0 (Fig. 3.2) to send 10 times less

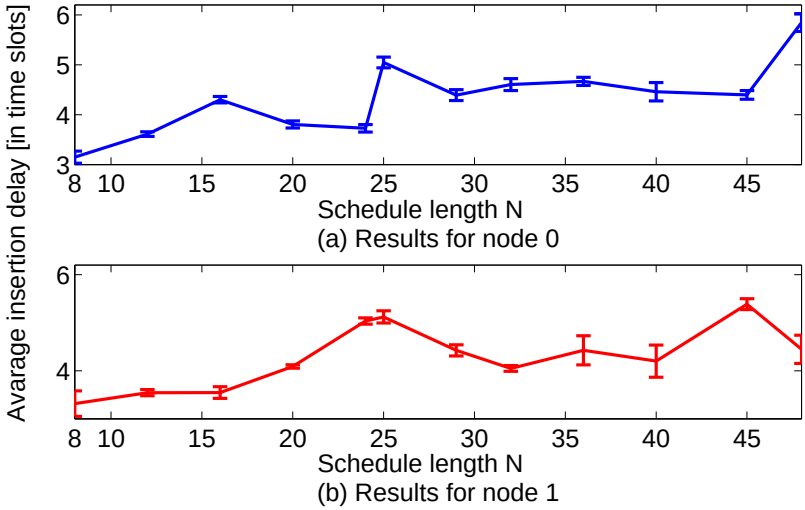


Figure 3.5 — Insertion latency for different schedule lengths.

traffic than the other nodes (i.e., node 0 now sends $a/10$). The asymmetry is supposed to make the scheduling more sensitive to the schedule length size.

The minimum schedule length for $a = 0.24$ can be shown to be $N = 8$ in this case. In Figs. 3.5(a) and 3.5(b), the results of `MSL` algorithm and ns-2 are illustrated. We measure by simulation the average insertion delay for nodes 0 and 1, respectively, when the schedule length is not at its minimum. Each schedule length leads to different slot allocation.

The presented results show that minimizing the schedule length generally improves the network performance. Furthermore, the impact is more clear at node 0 (improvement in delay more than 40%), than at node 1. This is because node 0 has the maximum degree of 4, which increases the impact of the schedule length.

3.2.8 Impact of parameter β on latency performance

Next, we evaluate the impact that QoS parameter β has on the node's insertion process. We suppose that network traffic is uniform and symmetric, with fixed amplitude $a = 0.24$. The minimum schedule length for $a = 0.24$ is $N = 25$ in this case (Fig. 3.3). Since the traffic amplitude is fixed, we measure how the insertion delay changes with β. The results are shown in Figs. 3.6(a) and 3.6(b), for nodes 0 and 1, (results of `MSL` algorithm and ns-2). The minimum β for which the 0-1 ILP optimization gives a feasible solution is 6. The maximum for β is equal to $N - 1 = 24$, since for a larger β the network performance does not change.

Figs. 3.6(a) and 3.6(b) show that with use of β, the insertion latency can be further reduced, below the value achieved by minimum schedule length. For instance, for node 0 (Fig. 3.6(a)), the latency is reduced from 4.6 to 3.3 time slots when using β, which is a reduction in delay greater than 20%, however, considering the propagation

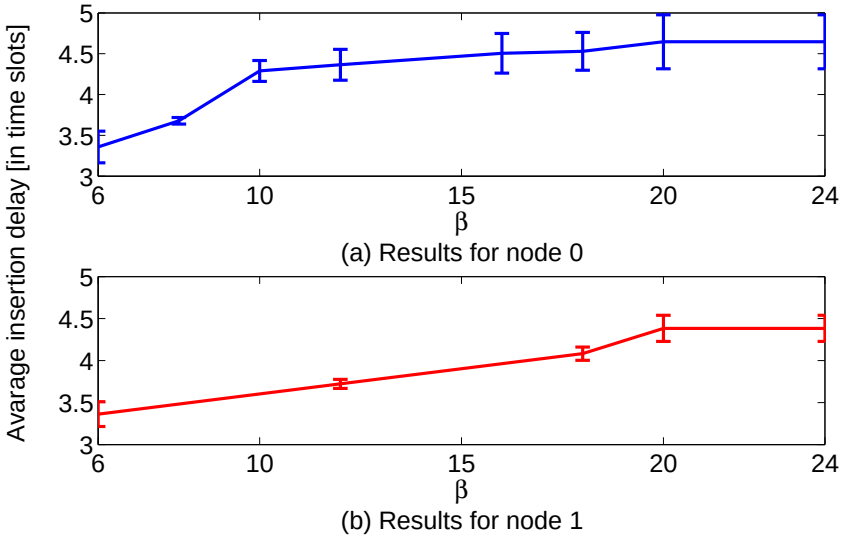


Figure 3.6 — Insertion latency for different values of parameter β.

delay the reduction becomes smaller for large scale networks. Once again, the node degree has an impact on the solution: a higher degree implies the better results.

3.2.9 Conclusion on minimum schedule length

We have proposed a novel slot allocating method for Time-Domain Wavelength Interleaved Network (TWIN), which improves the network performance by minimizing the scheduled length used for resource allocation. We developed the `Minimum Schedule Length (MSL)` algorithm to find the minimum schedule length. A key subroutine of `MSL` consists of a 0-1 Integer Linear Programming (0-1 ILP) formulation, which we call `SCHEDULE`.

We presented those algorithms in this section. The 0-1 ILP formulation method also includes a cyclic Quality of Service (QoS) constraint, which is used to provide a desired level of maximum queuing latency. To study the performance of network configurations, we have tested the solutions in the ns-2-based network simulator.

We have shown that minimizing the schedule length only can lead to latency reduction for more than 40%, while the use of cyclic QoS constraint can additionally improve the network performance.

3.3 Optimal cost dimensioning and resource allocation

In previous section, we proposed two ways of improving the slot allocation efficiency in Time-Domain Wavelength Interleaved Network (TWIN): by minimizing the schedule length and by enforcing a cyclical Quality of Service (QoS) constraint on the allocated slots. In this section, we extend the study by simultaneously addressing the Routing and Wavelength Assignment (RWA) and scheduling problem in TWIN, these two problems shall be solved together to minimize the total network cost. The proposed dimensioning tool is a 0-1 Integer Linear Programming (0-1 ILP) formulation.

3.3.1 Cost model for network dimensioning

A detailed description of the TWIN architecture and main dimensioning problems is given in the Sections 1.2.3, 1.3.3 and 1.3.3 of the state of the art chapter. In this section we focus on the cost model which is used in the developed dimensioning tools.

The objective function that we use in our 0-1 ILP formulations (Sections 3.3.2 and 4.2.3) is composed of three main costs: the transponder cost C_t, the cost of wavelength use per km of fiber C_{lw}, and the cost of gate C_s (only for Packet Optical Add Drop Multiplexer (POADM)).

When estimating the cost of a transponder (TRX) for TWIN and POADM, we consider that it shall be similar to the cost of an Ethernet port. The price of 10G SFP+ (Ethernet) TRX is around $ 1000 [26]. However, since Ethernet is well standardized but the POADM/TWIN encapsulation layers are not, the real price for a Wavelength Division Multiplexing (WDM) POADM/TWIN TRX may be significantly bigger. Having in mind this, and Moore's law impact on the equipment price, we can estimate that the final cost for the TRX is between $ 1000-2000.

The cost of wavelength use per km of fiber length (hereafter denoted simply by "wavelength cost") is primarily the cost of wavelength leasing, when the same fibers are shared by several operators. This cost varies from $C_{lw}(LB)$ (lower bound) to $C_{lw}(UB)$ (upper bound), with values expressed in km^{-1}. The lower bound on C_{lw} we set to:

$$C_{lw}(LB) = \frac{C_t}{\sum_\ell L_\ell \cdot W}; \tag{3.16}$$

where $\sum_\ell L_\ell$ is the total fiber length in the network (obtained as the sum of individual link lengths L_ℓ, over all links ℓ) and W is the total number of wavelengths. Such lower bound guarantees that the wavelength cost is negligible w.r.t. the TRX cost, representing a scenario where the network price is dominated by the transponder cost, such that the transponder cost is minimized firstly, and wavelength utilization is minimized as a secondary goal. The upper bound is taken to be:

$$C_{lw}(UB) = 0.1 \cdot C_t; \tag{3.17}$$

which is derived from [94], where the ratio between the wavelength cost per ring and the receiver cost is estimated on a ring circumference of 100 km.

Finally, we account for the gate cost when POADM rings are used. This cost is considered to be in the range from $ 50-100 , when gates are implemented with reflective Semiconductor Optical Amplifiers (SOAs). Here, we consider the cost of a discrete device, without taking into account the integration of the components (which could lower the price), nor the cost of the electronic drivers (which could increase the price). We would like to stress that the proposed cost model is not ideal as it only takes partially into consideration the node architecture in TWIN and POADM. Here, the goal is, above all, to compare the two technologies in terms of the efficient transponder and wavelength use.

3.3.2 0-1 ILP formulation for cost dimensioning, scheduling and RWA

The 0-1 ILP solution presented next, assesses the RWA and the problem of optimal slot allocation simultaneously. We use the traffic matrix expressed in Gb/s, and we allocate the minimum needed number of slots to each traffic demand. Both, the size of the schedule K and the input traffic matrix are the input parameters in the formulation.

Given Parameters

- $G(V, E)$: a non-directed graph describing the physical mesh topology, where V is the set of nodes, E is the set of links;
- T_d: normalized number of slots to be allocated for demand d.
- $\mathcal{R}_d = \{R_{d,1}, R_{d,2}, \ldots, R_{d,|\mathcal{R}_d|}\}$: set of routing paths $R_{d,r}$ which can be used to carry demand d, see below;
- $R_{d,r}$: set of links which are part of a tree and are used to route a demand d;
- start($R_{d,r}$): the (set of) link(s) of $R_{d,r}$ that are attached to source nodes of demand d;
- end($R_{d,r}$): the (set of) link(s) of $R_{d,r}$ that are attached to destination nodes of demand d;
- W: maximum number of wavelengths per fiber;
- C_{lw}: wavelength leasing cost per km of the used fiber (link);
- C_t: transponder cost;
- C_{slot}: unallocated slot cost (used only in the objective function from Subsection 3.3.3);
- L_ℓ: length of link ℓ in km;

- $D_{d,\ell,r}$: delay (in real or integer number of slots) experienced by a slot emitted by the source src of demand d to reach link ℓ, by using the route r. Let dst be the destination of demand d. Let z be the node on route r, such that $z \in r$ and $z = end(\ell)$. We define: $v_{z,\ell,r} = D_{d,\ell,r} - \lfloor D_{d,\ell,r} \rfloor$, $(0 < v_{z,\ell,r} < 1)$;
- η_z: set of src with property $\eta_z = \arg\min_{src} v_{z,\ell,r}$;
- K: number of slots used for the allocation, i.e., the "schedule length" (a number at least as large as the number of edge nodes, i.e., sources and destinations);
- A: number of demands;
- B: maximum number of alternate routes in a structure $\mathcal{R}_d$, (i.e., $\max_d \mathcal{R}_d$);

Output Variables

- binary $x_{\ell,w}^{k,d,r}$ equal to 1 if slot k is used to carry demand d on wavelength w, by using link ℓ and route r from $\mathcal{R}_d$, and 0 otherwise;
- binary $u_{i,w}$ equal to 1 if a TRX with fast tunable laser and a receiver for fixed wavelength w is deployed at node i, and 0 otherwise;
- binary $e_{\ell,w}$ equal to 1 if link l on wavelength w is used, and 0 otherwise;
- binary $p_{d,r}$ equal to 1 when route r is used to carry the traffic for demand d, and 0 otherwise;

Objective function, minimizing the number of used channels on the links and number of transponders, is:

$$\min\Big(C_{lw} \cdot \sum_{\ell}^{E} \sum_{w=1}^{W} L_\ell \cdot e_{\ell,w} + C_t \cdot \sum_{i=1}^{|V|} \sum_{w=1}^{W} u_{i,w}\Big); \tag{3.18}$$

General constraints

Capacity constraint, ensuring the allocation of sufficient number of slots:

$$\sum_{r=1}^{|\mathcal{R}_d|} \sum_{\substack{\exists \ell \in E \\ \ell \in r}}^{E} \sum_{w=1}^{W} \sum_{k=1}^{K} x_{\ell,w}^{k,d,r} = T_d, \qquad \forall d \leq A; \tag{3.19}$$

Slot-wavelength continuity constraint (i.e., ensuring that an allocated slot is on the same wavelength, at the same slot location, all the way from a source node to any destination node):

$$x_{\ell,w}^{k,d,r} = x_{\ell',w}^{k,d,r}, \qquad \forall w \leq W, \forall d \leq A, \forall r \leq |\mathcal{R}_d|, (\forall \ell, \ell' \in R_{d,r}), \forall k \in K; \tag{3.20}$$

Routing and wavelength assignment constraints

Link utilization variable constraint (used to determine which wavelength is used on which links):

$$\sum_{d=1}^{A}\sum_{r=1}^{|\mathcal{R}_d|}\sum_{k=1}^{K} x_{\ell,w}^{k,d,r} \leq A \cdot B \cdot K \cdot e_{\ell,w}, \qquad \forall \ell \in E; \tag{3.21}$$

Constraints ensuring that each traffic demand d is carried over a single path (the absence of multiple paths prevents the reordering of bursts at the reception):

$$\sum_{k=1}^{K}\sum_{\ell \in r}\sum_{w=1}^{W} x_{\ell,w}^{k,d,r} \leq W \cdot K \cdot |E| \cdot p_{d,r}, \qquad \forall d \leq A, \forall r \leq |\mathcal{R}_d|; \tag{3.22}$$

$$\sum_{r=1}^{|\mathcal{R}_d|} p_{d,r} \leq 1, \qquad \forall d \leq A; \tag{3.23}$$

Constraint ensuring that two nodes do not share the same color, which is a native TWIN property:

$$\sum_{i=1}^{|V|} u_{i,w} \leq 1, \qquad \forall w \leq W; \tag{3.24}$$

Slot allocation constraints

Fast wavelength-tunable TX capacity constraint (allocates the TXs at each node based on sent slots):

$$\sum_{\substack{d=1,\\ i \in d}}^{A}\sum_{\ell}^{E}\sum_{\substack{r=1,\\ \ell=\text{start}(R_{d,r})}}^{|\mathcal{R}_d|}\sum_{w=1}^{W} x_{\ell,w}^{k,d,r} \leq \sum_{w=1}^{W} u_{i,w}, \qquad k \leq K, \forall i \leq |V|; \tag{3.25}$$

Colored RX capacity constraint (allocates the RXs at each node based on received slots), accounts for non-integer (in time slots) propagation delay:

$$\begin{aligned} &\sum_{\substack{d=1,\\ i=dest(d)}}^{A}\sum_{r=1}^{|\mathcal{R}_d|}\sum_{\substack{\ell \in r,\\ i=\text{end}(\ell)}}^{E} B_1(x) \leq u_{i,w}, \qquad \forall w \leq W, \forall k \leq K, \forall i \leq |V|; \\ &\sum_{\substack{d=1,\\ i=dest(d)}}^{A}\sum_{r=1}^{|\mathcal{R}_d|}\sum_{\substack{\ell \in r,\\ i=\text{end}(\ell)}}^{E} B_2(x) \leq u_{i,w}, \qquad \forall w \leq W, \forall k \leq K, \forall i \leq |V|; \end{aligned} \tag{3.26}$$

define:
$\forall l \in E, \forall d \leq A, \forall r \leq |\mathcal{R}_d|, \forall w \leq W, \forall z \in end(l)$,

$$B_1(x) = \begin{cases} x_{l,w}^{k',d,r}, & \text{for } src \in \eta_z \\ x_{l,w}^{k'(k-1),d,r}, & \text{otherwise} \end{cases}$$

$$B_2(x) = \begin{cases} x_{l,w}^{k',d,r}, & \text{for } src \in \eta_z \\ x_{l,w}^{k'(k+1),d,r}, & \text{otherwise} \end{cases}$$

where:
$k' \equiv \lfloor k - D_{d,\ell,r} \rfloor \mod K$.

The slot indexes k' is calculated to account for different propagation delays of traffic flows taking different propagation paths between sources towards the same destination.

Constraint which avoids link collision by preventing several demands from using the same link on the same wavelength and the same slot:

$$\begin{aligned} &\sum_{d=1}^{A} \sum_{\substack{r=1, \\ \ell \in r}}^{|\mathcal{R}_d|} B_1(x) \leq 1, \qquad \forall \ell \in E, \forall w \leq W, \forall k \leq K; \\ &\sum_{d=1}^{A} \sum_{\substack{r=1, \\ \ell \in r}}^{|\mathcal{R}_d|} B_2(x) \leq 1, \qquad \forall \ell \in E, \forall w \leq W, \forall k \leq K; \end{aligned} \tag{3.27}$$

where $k' \equiv \lfloor k - D_{d,\ell,r} \rfloor \mod K$.

3.3.3 Impact of real propagation delay on resource allocation efficiency

The 0-1 ILP formulation described in previous subsection returns the optimal cost for a TWIN network, given a network graph G and a traffic demand T. In this subsection we assume that routing is static and that a single route is pre-computed and assigned to each (s, d) pair. We are interested in the size of the "unallocated capacity" in case of propagation delay expressed in real number of slots. For this study, the following objective function is used:

$$\min\Big(C_{slot} \cdot \sum_{\ell}^{E} \sum_{\substack{r=1, \\ \ell \in r}}^{|\mathcal{R}_d|} \sum_{k=1}^{K} \sum_{w=1}^{W} (1 - x_{\ell,w}^{k,d,r})\Big) \tag{3.28}$$

Here, we suppose a full uniform and symmetric traffic matrix, with fixed traffic amplitude $\alpha = 0.1$. For this traffic and network (Fig. 3.2), the lower bound on the schedule length is K(N)=4. The propagation distances, for the network in study are randomly

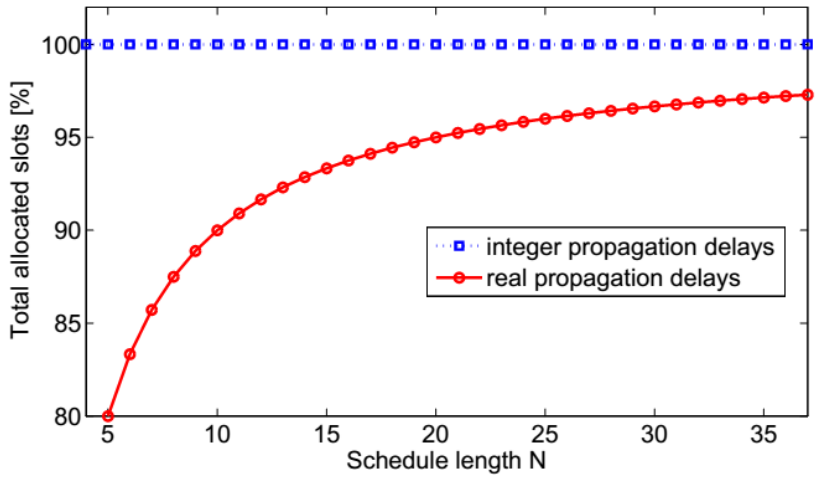


Figure 3.7 — Slot allocation efficiency in case of real propagation delay.

generated, by following the uniform distribution with mean value of 5 slots. We compare the total number of allocated slots, for different schedule sizes N, in Fig. 3.7.

From Fig. 3.7, we can see that for integer propagation delays the number of allocated slots is 100%, for all schedule lengths N. For non-integer propagation delays, the number of allocated slots falls to 80% of all available slots, for very small sizes of schedule length. It is because, at the source each slot can overlap with only 1 slot of different traffic demand with the same destination. However, with the increase of the schedule length, the percentage of unallocated slots recovers quickly. For sufficiently large schedule lengths, the total number of allocated slots approaches 100% efficiency.

We can also notice that for $N = 4$, no feasible solution exists, in case of real propagation delays, because a slot cannot be assigned to a TRX, since it is blocked with the previous slot toward a different destination (Fig. 3.8).

In the following, two solutions to overcome this issue are proposed:

Link Calibration

In order to have a perfect alignment at the source, each source-destination propagation time should be equal to an integer number of the time slot. Thus $V(V-1)$ paths should be calibrated. Therefore, this can be done by calibrating the paths by means of variable optical delay lines [41]. However, if each link is calibrated individually, then all the $V(V-1)$ paths are calibrated as well. Consequently, the total number of delay lines required for the calibration is equal to the number of links in the network which is always less than $V(V-1)$ connections.

Doubled TRX

In order to calibrate the network, the above solution requires a number of optical variable delay lines equal to the number of links. However, even if the network is

not calibrated, it is possible to fill all the allocated slots to a destination by using two tunable lasers. The slot overlap (Δt) is always less than the slot duration (t_k), therefore, the maximum number of overlapping slots cannot be higher than two (Fig. 3.8). Thus, two tunable lasers are sufficient to avoid blocking at the source.

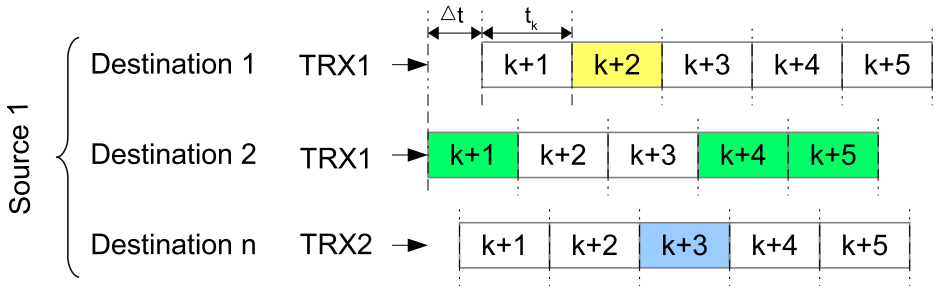


Figure 3.8 — Alignment at source, in case of real propagation delay.

In order to reduce the time complexity of the 0-1 ILP formulations, in the remaining of this document, only the integer propagation delays are considered.

3.4 Scalable solution for dimensioning, scheduling, RWA and virtualization

Optimally solving the Routing and Wavelength Assignment (RWA) and scheduling sub-problems in Time-Domain Wavelength Interleaved Network (TWIN) cannot be done in polynomial time, so the 0-1 Integer Linear Programming (0-1 ILP) tool proposed in previous sections are limited to small examples and cannot be used for real-scale networks. In this section, we propose an algorithm [69], called Heuristic Method for Virtual TWIN Dimensioning (HMVTD), that as primary concern provides a scalable solution for the RWA and scheduling optimization, a problem treated in Sections 3.3.2. In addition, for the first time, the proposed algorithm allows the creation of the arbitrary number of virtual domains over the same physical TWIN network, at minimum cost, and thus, assesses the emerging challenge of enabling the network level virtualization for optical transport technologies [49],[43],[55].

The network virtualization is an important building block in Software-Defined Networking (SDN) [49], and is a technology that shall facilitate the centralized control and the automation of the high capacity optical infrastructure. The optical network virtualization enables the abstraction and sharing of optical interfaces, channels and links between different virtual network domains, and offers to the service providers the possibility to disjointly operate such domains over the same optical infrastructure. Separate virtual domains usually support separate Quality of Service (QoS) guarantees and service contracts (e.g., see [45]). To the best of our knowledge, a virtualization enabling algorithm has not been proposed so far for a TWIN network. HMVTD defines the way the optical interfaces are shared between different virtual domains, and maps the time slots to each of the interfaces, according to the calculated global scheduling.

The remainder of the section is organized as follows. Subsection 3.4.1 is devoted to describe the proposed heuristic method and its variants. Subsections 3.4.2-3.4.5 show the detailed numerical results. Final subsection concludes this study and highlights the significant contribution of this work.

3.4.1 Heuristic method for virtual TWIN dimensioning

The Capital Expenditures (CAPEX) cost of the TWIN network, or the total network cost, is composed of the wavelength leasing cost per km of the used fiber, (C_{lw}), and cost of transponder, TRX (C_t). The objective of the dimensioning problem considered here is to minimize the total network cost, over all the network virtual domains. Since the wavelength cost is related to the link length, traffic routes over greater distances have higher price. Therefore, we are able to encompass the impact that the physical topology has on the final network cost.

We use the traffic matrix to identify the number of needed slots to each traffic demand. Both the size of the TWIN's schedule K and the input traffic matrix are the input parameters in the heuristic. An important aspect of the heuristic is to satisfy the constraints that prevents slot blocking at the sources, and the bursts collisions at the

destinations, and all intermediate nodes. We assume that the propagation delays are integer number of slot durations.

Finally, we suppose that the slots that are the object of allocation are enumerated by numbers $0, 1, ..., K-1$ (mod K), and that the scheduling that is the result of the optimization is periodically repeated.

The full list of input parameters and variables used by HMVTD algorithm is given next.

Input Parameters

- $G(V, E)$: a non-directed graph describing the physical mesh topology, where V is the set of nodes and E is the set of links;
- $T^v_{i,j}$: traffic demand in number of slots (normalized to schedule length) to be allocated for flow between the source i and destination j, for virtual domain v;
- $\ell_{i,j}$: length in km of a link from node i to node j;
- D_k: time slot duration; ν: speed of light in the fiber;
- W: maximum allowed number of wavelengths per fiber;
- TRX: maximum allowed number of transponders per node;
- C_{lw}: wavelength leasing cost per km of the used fiber;
- C_t: transponder cost;
- K: number of slots used for the allocation, (i.e., the "schedule length");
- $DemandOrderingPolicy(v)$: either MLC, MLS, MLD, LCF or RD, for different virtual domains v, with the acronyms defined below;
- $DemandServingPolicy(v)$: either ED or PD, for different virtual domains v, with the acronyms defined below;
- $SlotSelectionPolicy(v)$: either FFS or RS, for different virtual domains v, with the acronyms defined below;
- N: number of iterations (constant ≥ 1 if $DemandOrderingPolicy(v)$=RD or $SlotSelectionPolicy(v)$=RS, else equal to 1);
- D: maximum number of overlayed virtual networking domains in the final solution;
- $NetworkConfiguration$: contains the mapping of slots and virtual domains to the physical network (it can be a given parameter, or an empty set).

Output Variables

- $G'(V, E')$: a non-directed graph describing the physical tree topology, where V is the set of nodes, E' is the set of links;
- $R^v_{i,j}$: set of links used to route the demand from source i to destination j, in virtual domain v;
- $L^v_{i,j}$: length of path from source i to destination j, in virtual domain v;
- $\delta^v_{i,j}$: propagation delay (in number of slots) experienced by a slot emitted by the source i until it reaches the destination j, for virtual domain v;
- binary $S^{t,w,v}_{i,j}(k)$ equal to 1, if slot k is used to carry demand from source i to destination j, in virtual domain v, by using transmitter t, wavelength w, and equal to 0 otherwise;
- binary tx^t_i equal to 1, if a transmitter t with fast tunable laser is deployed at node i, and 0 otherwise;
- binary rx^w_j equal to 1, if a fixed wavelength receiver w is deployed at node j, and 0 otherwise;
- integer N_W, denotes total number of wavelengths used in the network;
- real C_W denotes total wavelength leasing cost in the network;
- real C_{TRX} denotes total transponder cost in the network;
- real C_{TOT} denotes total design cost of the network;

The pseudo-code for the algorithm HMVTD is given in Alg. 3, and is composed of three basic parts:

1. tree and path construction,
2. resource allocation,
3. cost calculation.

The algorithm is repeated N times in order to calculate the overlapping slots and define the optical resource sharing for all virtual domains. The details of all algorithm parts are defined next.

1) Tree and path construction

In this step, first a Minimum Spanning Tree (MST) is found for a given virtual domain. Several algorithms find the MST, like Kruskal's [46] and Steiner's algorithms. While Kruskal algorithm spans all vertices (V) of a given graph, the Steiner algorithm spans a given subset of vertices ($V' \subset V$). Therefore, for the current implementation, the Kruskal's algorithm is chosen.

Kruskal's algorithm finds a MST for a connected weighted graph $G(V, E)$, i.e., it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. For current implementation, the weight is expressed in the link length $\ell_{i,j}$.

Next, the routing paths $R^v_{i,j}$ are found on the MST: the problem is solved by Breadth-First Search (BFS) algorithm. BFS finds the path on the graph starting from the source i and inspects all the neighboring nodes until it finds the destination j. At the end of this step, the delay $\delta^v_{i,j}$ (in integer number of slots) is calculated.

2) Resource allocation

For the resource allocation, we study various variants of demand ordering policies, demand serving policies, and slot selection policies, as described next.

Demand Ordering Policies

- Most Loaded Connection (MLC): the demands i, j are sorted in the decreasing order based on the load from a source i towards a destination j;
- Most Loaded Source (MLS): the demands i, j are sorted in the decreasing order based on the total load from a source i;

$$max(\sum_{j}^{|V|} T_{i,j}), \qquad \forall i \leq |V|; \tag{3.29}$$

- Most Loaded Destination (MLD): the demands i, j are sorted in the decreasing order based on the total load towards destination j;

$$max(\sum_{i}^{|V|} T_{i,j}), \qquad \forall j \leq |V|; \tag{3.30}$$

- Longest Connection First (LCF): the demands i, j are sorted in the decreasing order based on the path length $L^v_{i,j}$ between source i and destination j;
- Random Demands (RD): the demands i, j are sorted in random order.

Note that for the above policies, in the case of equal values, the demand with lower index is ordered first.

for *all* $v \leq D$ **do**
  sub. 1: Tree and path construction:
  Run the algorithm MST on $G(V, E)$;
  // Now we have a new graph $G'(V, E')$;
  Run the algorithm BFS on $G'(V, E')$, to calculate $R^v_{i,j}$ and $L^v_{i,j}$, for all $i, j \in V$;
  set $\delta^v_{i,j} = L^v_{i,j}/(\nu * D_k)$ $(\forall i, j \in V)$;
  sub. 2: Resource allocation:
  Create the $List(Demands)$ containing the traffic demands $T^v_{i,j}$, sorted in the order defined by $DemandOrderingPolicy(v)$;
  for *iteration* $\leq N$ **do**
    set $S^{t,w,v}_{i,j}(k) = 0$, $(\forall k)(1 \leq k \leq K)$;
    for *all* $T^v_{i,j}$ *taken in the order of* $List(Demands)$ **do**
      while $T^v_{i,j} > 0$ **do**
        for *all* t, w *taken in the first-fit order* **do**
          for *all values of* k*, chosen according to* $DemandServingPolicy(v)$ *and* $SlotSelectionPolicy(v)$ **do**
            if *NO SLOT BLOCKING neither at source nor at destination* **then**
              $S^{t,w,v}_{i,j}(k) = 1$; $T^v_{i,j} = T^v_{i,j} - 1$; proceed with next $T^v_{i,j}$;
            end
          end
        end
      end
    end
    sub. 3: Cost calculation:
    COST UPDATE;
    $iteration = iteration + 1$;
  end
  $v = v + 1$; Update $NetworkConfiguration$;
end

Algorithm 3: Heuristic Method for Virtual TWIN Dimensioning.

Demand Serving Policies

- Entire Demand (ED): once a demand is selected, the entire number of required slots k are allocated;
- Partial Demand (PD): once a demand is selected, a single slot k out of required number is allocated, and the following demand is served.

Slot Selection Policies

- First Fit Slot (FFS): the first available slot k is allocated;
- Random slot (RS): random available slot k is allocated.

Note that the slot selection policy can include the QoS specifications (e.g., as done in Section 3.2), which allows to set the desired QoS level in each virtual domain.

To allocate a slot k for a demand i, j on wavelength w, the slot availability at the source side and destination side is checked. In Eq. 3.32, the difference in propagation delay between i, j and i', j pairs, $\delta_{i,j} - \delta_{i',j}$, are considered. The "NO SLOT BLOCKING" condition from Alg. 3 is defined by the equations:

$$\sum_{j'}^{|V|} \sum_{w}^{W} \sum_{v}^{D} S_{i,j'}^{t,w,v}(k) = 0, \quad \forall k \leq K, \forall t \leq TRX, \forall i \leq |V|; \tag{3.31}$$

$$\sum_{i'}^{|V|} \sum_{t}^{TRX} \sum_{v}^{D} S_{i',j}^{t,w,v}(k + (\delta_{i,j} - \delta_{i',j}) \text{mod} \ \ K) = 0,$$

$$\forall k \leq K, \forall w \leq W, \leq K, \forall i \leq |V|, \forall j \leq |V|; \tag{3.32}$$

Once a slot k is allocated to nodes i, j, the traffic demand $T_{i,j}^{v}$ is decreased by 1. The resource allocation sub-problem repeats until all the traffic demands are satisfied.

3) Cost calculation

The "COST UPDATE" procedure from Alg. 3 consists of updating all the relevant network costs. In this phase, the total design cost C_{T0T} is calculated, composed from the cost of wavelength and the cost of transponders.

The total number/cost of transponders C_{TRX} is calculated (Eq. 3.35) as sum of the maximum between the number of transmitters (Eq. 3.33) and receivers (Eq. 3.34) at each node:

$$D \cdot |V| \cdot W \cdot K \cdot tx_i^t \geq \sum_{v}^{D} \sum_{j}^{|V|} \sum_{w}^{W} \sum_{k}^{K} S_{i,j}^{t,w,v}(k), \quad \forall t \leq TRX, \forall i \leq |V|; \tag{3.33}$$

$$D \cdot |V| \cdot TRX \cdot K \cdot rx_j^w \geq \sum_{v}^{D} \sum_{i}^{|V|} \sum_{t}^{TRX} \sum_{k}^{K} S_{i,j}^{t,w,v}(k), \quad \forall w \leq W, \forall j \leq |V|; \tag{3.34}$$

$$C_{TRX} = C_t \cdot \sum_{i}^{|V|} max(\sum_{t}^{TRX} tx_i^t, \sum_{w}^{W} rx_i^w); \tag{3.35}$$

The total number of wavelengths N_W and the total wavelength cost C_W:

$$N_W = \sum_{j}^{|V|} \sum_{w}^{W} rx_j^w; \tag{3.36}$$

$$C_W = C_{lw} \cdot \sum_{v}^{D} \sum_{i}^{|V|} \sum_{j}^{|V|} \sum_{w}^{W} (L_{ij}^v \cdot rx_j^w); \tag{3.37}$$

For the combination which uses the random policies (slot selection policy (RS) and/or demand ordering policy (RD)), the HMVTD computes the total cost N times (for each virtual domain v), and out of N costs it keeps the value of the minimum total cost, C_{TOT}.

Note that each sub-problem has as input the results from the previous sub-problems. Therefore, the order of sub-problems in the HMTD are not interchangeable.

Following subsections report the results obtained by using the HMVTD, which is implemented in C++ programming language, and runs on Linux computer with 12 central processing units and 24GB of random-access memory. The results are compared with the results obtained from a commercially available LP solver (CPLEX [34]), in which the 0-1 ILP formulation from Section 3.3.2 has been implemented (these results are given with the optimality gap of 10%).

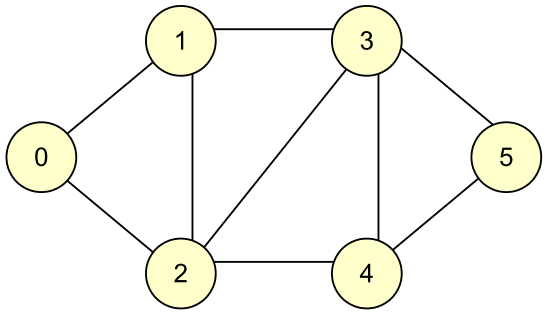


Figure 3.9 — Topology of a 6-node mesh network.

We consider the 6-node physical topology with 9 links as depicted in Fig. 3.9. The propagation delay on each link corresponds to 5 slots duration, where one time slot is equal to 10 μs. The wavelength channel and TRX data rate is supposed to be 10 Gb/s.

The schedule length is supposed to be $K = 5$ and the number of wavelengths in the network is not greater than $W = 40$. It is also possible to account for non-integer number of slots, but it is outside the scope of this section.

The cost is expressed in arbitrary units (a.u.). We set the TRX cost to $C_t = 1$ (a.u.), while the wavelength cost varies from $C_{lw} = 10^{-4}$ (a.u.)·km^{-1} (lower bound) to $C_{lw} = 0.1$ (a.u.)·km^{-1}. The lower bound on C_{lw} is calculated as a value at which a single transponder cost is equal to the total possible wavelength leasing cost, $C_t/(\sum_\ell L_\ell \cdot W)$; this ensure that no transponder can be traded for a wavelength (Subsection 3.3.1). The assumption for the upper bound on C_{lw} is derived from [94], where the ratio between the wavelength cost per ring and the receiver cost is estimated on a small size metropolitan ring.

In the following, we assess the performance of different heuristic combinations in terms of complexity and design cost, first, for the single virtual domain (in Subsections 3.4.2, 3.4.3 and 3.4.4). Then, in Subsection 3.4.5, we evaluate the gain of virtualization considering multiple virtual domains.

3.4.2 HMVTD performance evaluation

We first study the efficiency of the HMVTD using different heuristic combinations, when performing the resource allocation. For this study, we considered a uniform and symmetric traffic between each pair of nodes, with fixed total sent/received traffic of $\alpha = 2c$ at each node (where c is the channel capacity). The TRX and wavelength costs are, respectively: $C_t = 1$ (a.u.), $C_{lw} = 0.1$ (a.u.) km^{-1}. The difference in cost between the HMVTD and the 0-1 ILP optimal solution is called "the cost penalty".

Note that for the heuristics which use randomness in slot selection policy (i.e., RS) and in demand ordering policy (i.e., RD), the number of iterations N is equal to 100. All the results are averaged over 20 simulations, and we verify that the confidence intervals are sufficiently small with regard to the system model of this study.

Tab. 3.2 shows that selecting the demands according to the MLC and MLS achieves worse performance than other algorithm variants. LCF and RD policies have similar cost penalty, while RD policy has slightly increased time complexity due to the presence of iterations.

ED-based demand serving policy outperforms the PD in all cases. This proves that serving the demands according to round robin process leads to resource wastage.

RS policy outperforms the FFS in most of the cases. However, FFS achieves low cost penalty (27.51%) when it is coupled with random demand ordering. This is explained by the fact that random process explores multiple possibilities and keeps the configuration with minimum cost.

In addition to the cost penalty, performance of heuristics is also related to time complexity. In HMVTD, time complexity depends on number of virtual domains (D), network size ($|V|, |E|$), schedule length (K), and number of iterations (N) as an additional parameter for the random-based algorithms.

Table 3.2 — Cost penalty for different resource allocation policies

Demand Ordering Policy	Demand Serving Policy	Slot Selection Policy	Time Complexity	Cost Penalty [%]
MLC	ED	FFS	$O(D\|V\|^3K^2)$	68.75
		RS	$O(D\|V\|^2K^3N)$	30.21
	PD	FFS	$O(D\|V\|^3K^2)$	89.58
		RS	$O(D\|V\|^2K^3N)$	46.66
MLS	ED	FFS	$O(D\|V\|^3K^2)$	68.75
		RS	$O(D\|V\|^2K^3N)$	30.16
	PD	FFS	$O(D\|V\|^3K^2)$	89.58
		RS	$O(D\|V\|^2K^3N)$	45.51
MLD	ED	FFS	$O(D\|V\|^3K^2)$	41.66
		RS	$O(D\|V\|^2K^3N)$	29.58
	PD	FFS	$O(D\|V\|^3K^2)$	81.25
		RS	$O(D\|V\|^2K^3N)$	44.37
LCF	ED	FFS	$O(D\|V\|^3K^2)$	39.58
		RS	$O(D\|V\|^2K^3N)$	27.51
	PD	FFS	$O(D\|V\|^3K^2)$	47.91
		RS	$O(D\|V\|^2K^3N)$	44.37
RD	ED	FFS	$O(D\|V\|^3K^2N)$	27.51
		RS	$O(D\|V\|^2K^3N^2)$	29.58
	PD	FFS	$O(D\|V\|^3K^2N)$	44.58
		RS	$O(D\|V\|^2K^3N^2)$	40.62

The MST and BFS have the complexity of $O(|E|\log|V|)$ [46] and $O(|V|\sqrt{E})$ [27], respectively. The complexity of resource allocation sub-problem varies according to the heuristics combinations.

The cost calculation sub-problem has negligible complexity. HMVTD resolves the three sub-problem sequentially, therefore, its global complexity is dominated by the resource allocation algorithm that has the highest complexity. All algorithms have a polynomial complexity of degree between 6 and 8. As in a real network $K >> V$, the algorithms with the lowest order of K (e.g., MLC-ED-FFS and LCF-ED-FFS) are less complex than the others.

In the following, we carry out a detailed study of the performance considering the most pertinent heuristics having the cost within 30% of the optimal solution.

3.4.3 Impact of traffic intensity on final network price

In this subsection, we assess the total design cost as a function of traffic intensity with fixed TRX and wavelength cost. We set the TRX cost to $C_t = 1$ (a.u.). For such C_t, the wavelength cost is taken at its upper bound, and is fixed to $C_{lw} = 0.1$ (a.u.) km^{-1}. In the following, distributed and centralized traffic are considered.

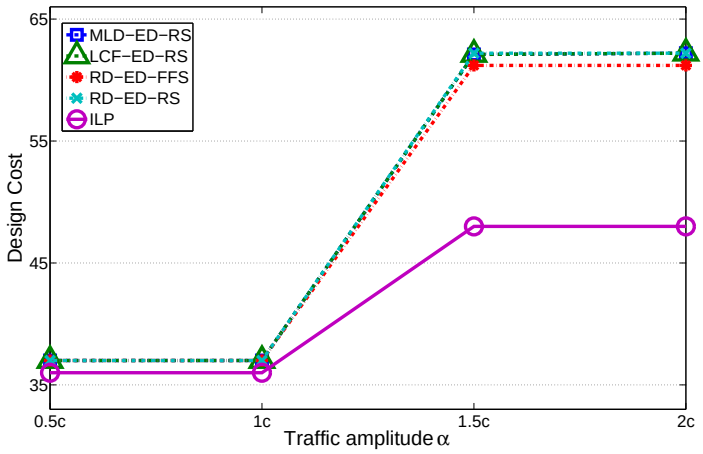


Figure 3.10 — Design cost in case of distributed traffic load.

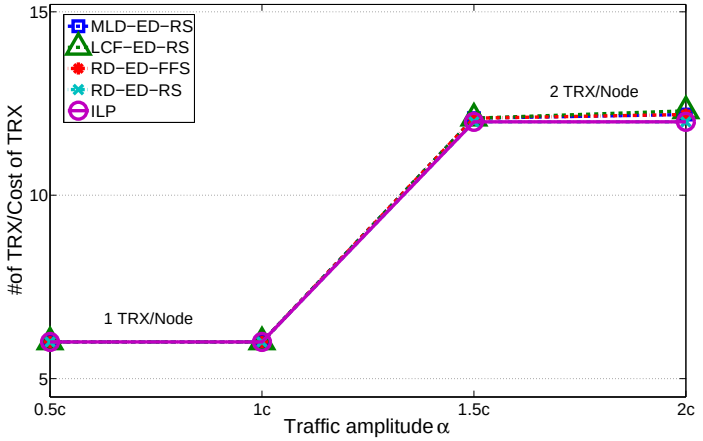


Figure 3.11 — Transponder cost in case of distributed traffic load.

Distributed Traffic

The traffic matrix is assumed to be uniform and complete as in the previous subsection (i.e., there are 30 traffic flows in the network). The total sent/received traffic by a node has amplitude α.

The design cost is plotted in Fig. 3.10. For traffic loads between $0.5c$ and c, the design cost is close to the optimal solution. For traffic higher than c, the design cost is within 27% that of the optimal solution. To explain this difference, we plot the TRX cost and wavelength cost separately.

Figs. 3.11 and 3.12 show that the number of transponders and wavelengths, respectively, are very close to the optimal solution. However, Fig. 3.13 illustrates that the difference in design cost between heuristics and 0-1 ILP is mainly due to the difference in wavelength cost, especially when traffic exceeds the wavelength capacity ($\alpha > c$). The penalty due to wavelength cost is the result of the simplified wavelength assignment process. Indeed, when a wavelength is completely filled in, HMVTD assigns

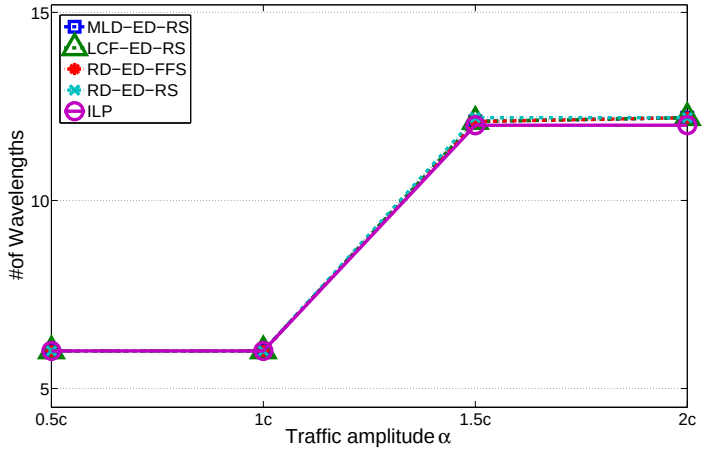


Figure 3.12 — Number of wavelengths in case of distributed traffic load.

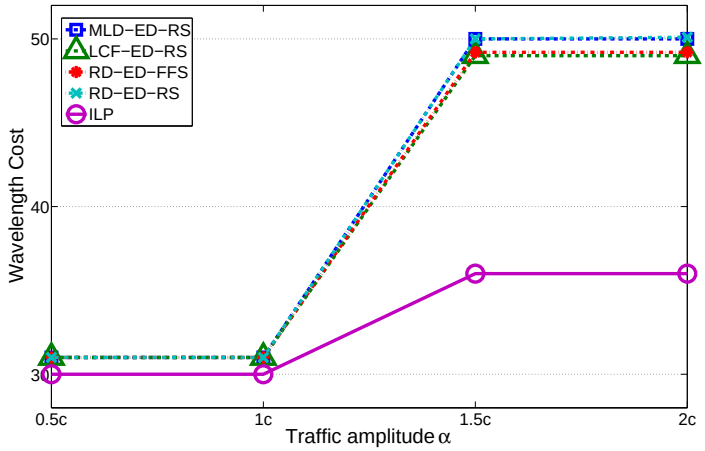


Figure 3.13 — Wavelength cost in case of distributed traffic load.

another wavelength and continues the resource allocation process for the remaining demands without minimizing the total wavelength per fiber length. The chosen wavelength assignment process helps to reduce significantly the heuristic complexity, since it avoids performing the resource allocation process from the beginning each time a new wavelength is needed. However, this is achieved at price of the increased total network cost.

Centralized Traffic

Centralized traffic is the case where one of the nodes in the network plays the role of a gateway to the backbone network. In our example, the gateway is located at node 1, and it is supposed that this node sends and receives back the traffic with the same amplitude from all the nodes in the network (i.e., α is the total sent/received traffic amplitude by the gateway). This scenario is close to the current operational metropolitan network.

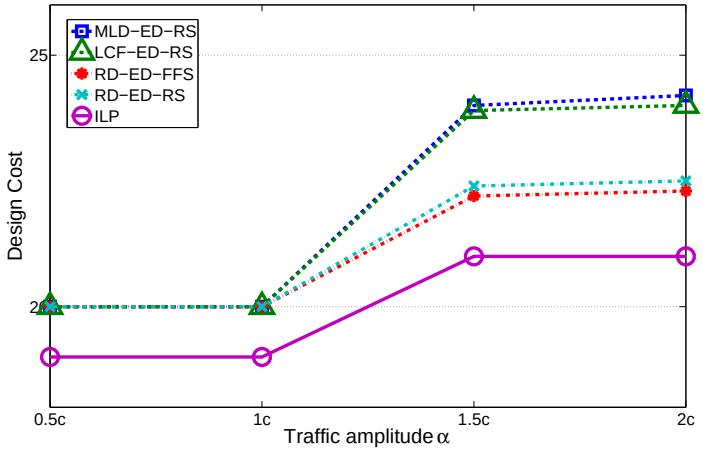


Figure 3.14 — Design cost in case of centralized traffic load.

The design cost comparison is presented in Fig. 3.14. The combinations using the RD policy have almost the same performance, with design cost within 6% that of the optimal solution. The MLD and LCF policies have similar performance and the average design cost is within 15% that of the optimal solution for $\alpha = 2c$. Similar to the distributed case, the cost penalty comes from the wavelength leasing cost over the network links.

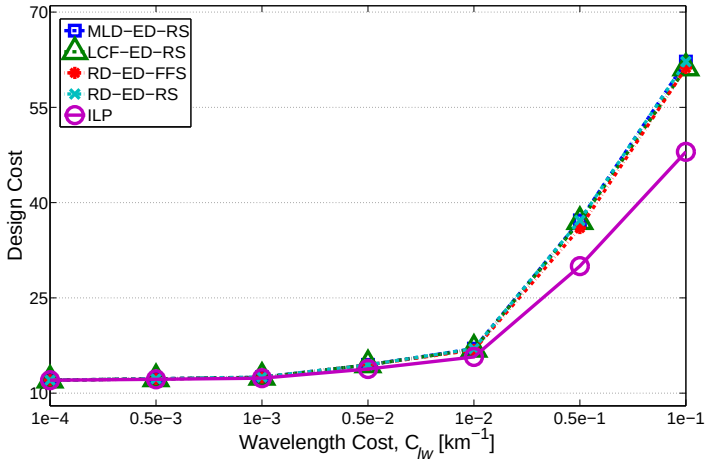


Figure 3.15 — Network design cost as a function of the wavelength cost

3.4.4 Impact of the wavelength cost on the final network price

The previous simulations show that the penalty in design cost (w.r.t. the optimal solution) is manly due to wavelength cost. In order to get a more complete understanding of the behavior for different policies, we study the impact of the wavelength cost C_{lw} on design cost. We consider for this study a distributed traffic profile, with fixed am-

plitude $\alpha = 2c$, as at this load the heuristics cost penalty is the most significant. The C_{lw} changes in the range, between 10^{-4} (a.u.) km^{-1} and 0.1 (a.u.) km^{-1}.

The results plotted in Fig. 3.15 show that for small values of wavelength cost, all four policies are very close to the optimal solution. However, for the sufficiently high value of C_{lw} (greater than 1% of the TRX cost), the design cost penalty of the heuristics increase, until it reaches 27% of the optimal solution.

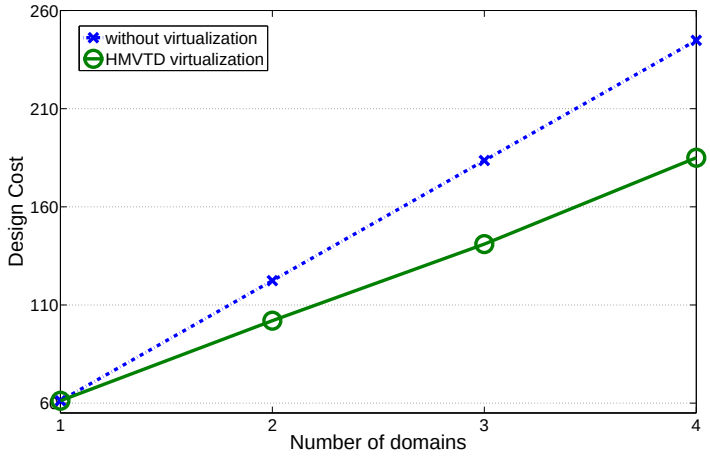


Figure 3.16 — Impact of virtualization on the design cost for distributed traffic load.

3.4.5 Impact of virtualization on the final network price

In this subsection, we assess the performance of the HMVTD for multiple virtual domains using the RD-ED-FFS heuristic combination, that has the best performance according to the previous study. The distributed scenario with traffic amplitude $\alpha = 2c$ is used.

Fig. 3.16 shows that sharing the resources between multiple virtual domains as defined by HMVTD leads to up to 30% savings in design cost comparing to the network dimensioning without virtualization, when resources are separately allocated for each domain.

3.4.6 Heuristic performance on large scale network: Deutsche Telekom

In this subsection, we assess the total design cost as a function of distributed traffic intensity with fixed TRX and wavelength cost. We set the TRX cost to $C_t = 1$ (a.u.). For such C_t, the wavelength cost is taken at its upper bound, and is fixed to $C_{lw} = 0.1$ (a.u.) km^{-1}.

For this case Deutsche Telekom network (Fig. 3.17) with 17 nodes and 26 links, is used. In order to have a metro network scenario, the links length is rescaled by factor of 10.

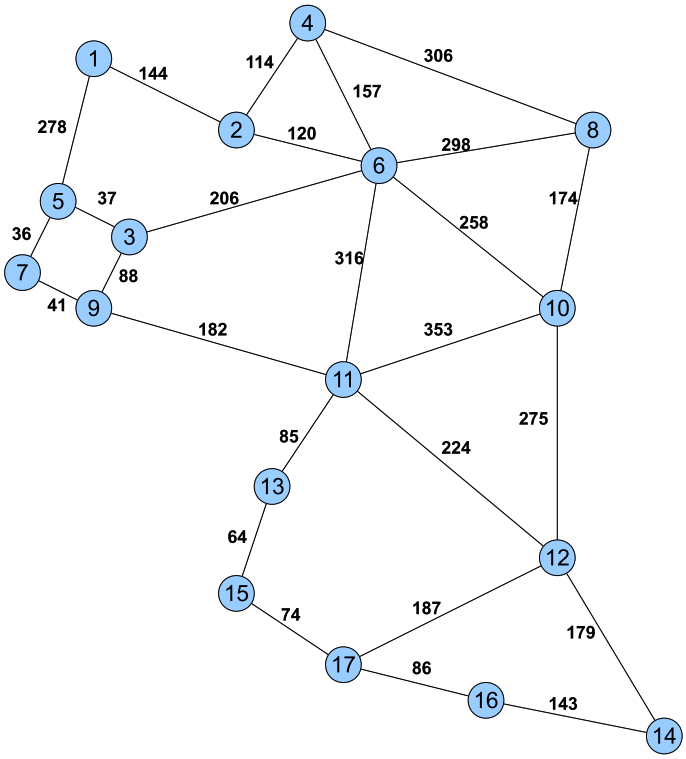


Figure 3.17 — Deutsche Telekom network.

The traffic matrix is assumed to be uniform and complete (i.e., there are 272 traffic flows in the network). The schedule length is supposed to be $K = 100$.

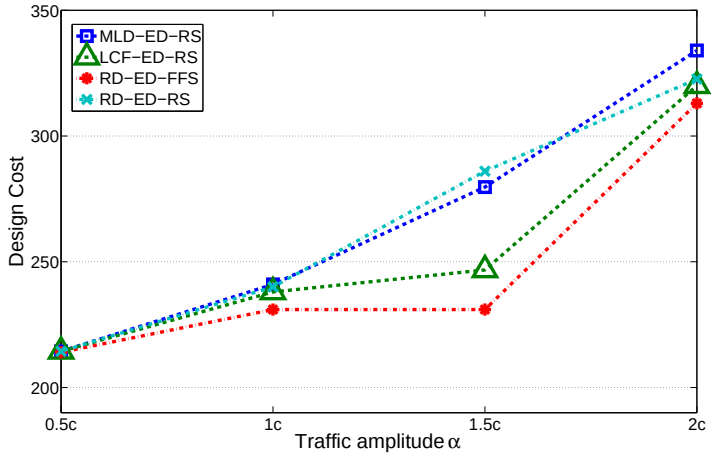


Figure 3.18 — Design cost in case of distributed traffic load.

Similar to previous subsections, the same set of results were obtained. However, the computation time is in order of units of minutes and not seconds. The total design cost is illustrated in Fig. 3.18 and it has the same trend as in Subsection 3.4.3, with slightly better performance for RD-ED-FFS policy.

3.4.7 Conclusion on TWIN cost dimensioning

In this section, we have presented the comprehensive algorithm called Heuristic Method for Virtual TWIN Dimensioning (HMVTD), that is a scalable solution for routing, scheduling and virtualization in Time-Domain Wavelength Interleaved Network (TWIN) optical burst switching network. The results show that the most pertinent heuristics combinations can reach a cost within 27% that of the optimal solution. This penalty is mainly due to wavelength leasing cost (driven by the simplified wavelength assignment), since the number of used transponders and wavelengths are close to the optimal. We also demonstrated that the impact of the wavelength leasing cost is dominant, when the wavelength cost per km is superior to 1% of the transponder cost. In addition, applying HMVTD to network virtualization leads to 30% saving in design cost w.r.t. the case without virtualization. In addition, the heuristic performances where checked on a large scale network (Deutsche Telekom), and the obtained results were as expected.

3.5 Conclusion

In this chapter, we have proposed a novel slot allocating method for Time-Domain Wavelength Interleaved Network (TWIN), which improves the network performance by minimizing the scheduled length used for resource allocation. We developed the `Minimum Schedule Length (MSL)` algorithm to find the minimum schedule length. A key subroutine of `MSL` consists of a 0-1 Integer Linear Programming (0-1 ILP) formulation, which we call `SCHEDULE`. The 0-1 ILP formulation method also includes a cyclic Quality of Service (QoS) constraint. We have shown that minimizing the schedule length can only lead to latency reduction for more than 40%, while the use of cyclic QoS constraint can additionally improve the network performance.

We have presented the comprehensive algorithm called Heuristic Method for Virtual TWIN Dimensioning (HMVTD), that is a scalable solution for routing, scheduling and virtualization in TWIN optical burst switching network. HMVTD breaks down the original NP-complete problem of TWIN dimensioning into three sub-problems: tree and path construction, resource allocation and cost calculation.

We used Kruskal's algorithm to construct the tree and Breadth-First Search (BFS) in order to find out the paths. The resource allocation is based on three policies: demand ordering, demand serving and slot selection. Several heuristic algorithms are proposed to perform each of the previous policies. Hence, many possible heuristics combinations are obtained. The final sub-problem, the cost calculation, evaluates the cost of the obtained solution taking into consideration the transponders and the wavelengths leasing costs.

We first assessed the performance of each heuristics combination in terms of time complexity and cost penalty compared with the optimal solution. Besides, we evaluated the design cost of the most pertinent combinations using different traffic intensities and different wavelength leasing cost. We also considered two traffic matrices: distributed and centralized.

Results show that the most pertinent heuristics can reach a cost within 27% that of the optimal solution. This penalty is mainly due to wavelength leasing cost, since the number of used transponders and wavelengths are close to the optimal.

Finally, we proved that HMVTD could be a promising candidate for TWIN network design considering multiple virtual domains and large scale networks. Considering these results, we plan to use HMVTD in different network scenarios and assess the network performance taking into account its QoS requirements.

CHAPTER 4

Design Comparison of Sub-wavelength Optical Switching Networks

In this chapter, we compare the cost of the optical transport solutions with sub-wavelength switching granularity. Two scenarios are considered, survivable metro networks and Long Term Evolution (LTE) mobile backhauling, with the focus on Time-Domain Wavelength Interleaved Network (TWIN) and Packet Optical Add Drop Multiplexer (POADM) technologies. Both technologies apply optical slot switching: TWIN on a physical mesh topology, and POADM on a physical ring. We use a single network dimensioning model that supports protection and multicast traffic, for the two technologies. The network planning solution is based on 0-1 Integer Linear Programming (0-1 ILP) formulation, and addresses the Routing and Wavelength Assignment (RWA) and the optimal schedule calculation.

4.1 Introduction

In this chapter, we focus on two such sub-wavelength-granularity switching technologies, Time-Domain Wavelength Interleaved Network (TWIN) [103], presented in Section 1.2.3, adapted to physically meshed networks, and Packet Optical Add Drop Multiplexer (POADM) [13], presented in Section 1.2.3, which applies to physical rings only. In addition to efficient bandwidth utilization thanks to the sub-wavelength switching granularity, TWIN and POADM are designed to minimize electronic switching and power consumption by resorting to the optical transparency for transit traffic [61]. In TWIN and POADM, nodes can communicate directly, without the need for a centralized traffic hub that electronically processes all data transmitted in the network.

An appealing property of TWIN is its entirely passive core infrastructure: the electronic processing of the traffic is performed only when inserting or extracting the optical bursts. This allows to have a core part of the network composed of inexpensive, and passive optical elements, such as couplers, splitters and wavelength blockers which consume no or little energy. Although its intermediate nodes are not entirely passive, POADM is more energy efficient than traditional Reconfigurable Optical Add Drop Multiplexer (ROADM) and Ethernet over Wavelength Division Multiplexing (WDM) solutions [61].

From the bandwidth efficiency point of view, in Section 3.3.3 and [95], the achievable throughputs of TWIN and POADM are studied, respectively, and it is shown that the network in both cases can achieve the capacity use efficiency of more than 90%. Both networks have a synchronous, slotted operation; the scheduling (resource allocation, term interchangeable with anterior) problem consists in efficiently allocating the time slots to different source-destination pairs. In contrast to POADM, where both opportunistic and non-opportunistic slot access can be used for scheduling slots, TWIN needs to use some kind of slot reservation or grant mechanism [92]. In both cases, the slot allocation is non-separable from the network planning, and the network configuration has a strong impact on the network performance, such as latency [98],[95].

Another example of a difference between POADM and TWIN is the way they use the WDM spectrum. In the original version of TWIN, different destinations must use different subsets of wavelengths, while in POADM the wavelengths can be shared by different destinations. This is an important advantage of POADM, which differentiates it from other metro technologies [6],[37].

Because of their differences, TWIN and POADM may have different Capital Expenditures (CAPEX)/Operational Expenditures (OPEX) cost and Quality of Service (QoS) performance. In the current chapter, for the first time we propose a dimensioning model for the two technologies which allows us to get initial conclusions on their CAPEX price.

In this chapter, the comparison solution is based on the use of a 0-1 Integer Linear Programming (0-1 ILP) formulation, and supports, for the first time, network planning for both TWIN and POADM, taking POADM as a special case of TWIN dimensioning where the underlying topology is a ring. The 0-1 ILP formulation solves the scheduling problem while minimizing the total number of transponders to be deployed and the price of wavelength per km. For POADM, the cost of fast wavelength blockers (e.g.,

Semiconductor Optical Amplifier (SOA) optical gates) utilization is also accounted. Note that the cost function is approximate, as it approximates the real cost of the main components present in TWIN and POADM. However, the model allows to estimate the efficiency of two solutions when using the basic network resources (transponders and wavelengths). In fact, all the cost components that depend on dimensioning are accounted for in the model. All the other costs are per-node costs, which do not depend on dimensioning, since nodes are known as an input. Furthermore, the cost model is strengthened by the fact that we consider wide ranges for each cost component, that are estimated based on available data.

The remainder of the chapter is organized as follows. In Section 4.2 the first scenario is considered, where the cost of survivable TWIN and POADM is compared. In particular, we include the protection model in our solution. Indeed, when planning the network, the resource needs for the protection shall be taken into account and can significantly increase the final network cost [71]. In Section 4.3, a detailed study of TWIN and POADM in Long Term Evolution (LTE) mobile backhauling scenario is presented. Finally, we conclude the chapter in the last section.

4.2 Scenario 1: survivable metro networks

There are numerous studies on the transport network protection mechanisms. For instance, Resilient Packet Ring (RPR) offers two protection mechanisms against a link failure event, "wrapping" and "steering" which differ in the protection triggering speed and in the efficiency of the bandwidth use [85]. Multiprotocol Label Switching - Transport Profile (MPLS-TP), is expected to have an improved set of resilience features in comparison to MPLS, mainly by introducing novel Operations, Administration, and Maintenance (OAM) mechanisms for protection triggering and by addressing ring topology protection with wrapping and steering methods [56]. In paper [71], the authors presented an extensive study of different protection methods for Optical Circuit Switching (OCS) networks.

The protection in TWIN is first considered in [103], where the possibility of using multiple trees to provide 1 : 1 protection is studied. In that paper, the process of allocating the slots is based on an approximative algorithm, while the TWIN-tree calculation problem has been resolved separately from the slot allocation. The cost of 1+1 protection was not considered. In [8], the authors discuss the possible protection methods for a variant of TWIN using the coherent technology, but they do not present the numerical results on protection. In our paper [68], we studied the cost of protection in TWIN, by using an optimal algorithm that resolves the Routing and Wavelength Assignment (RWA) and the scheduling problem, in the protection settings. Protection in bidirectional POADM rings is addressed in [99].

In this section, the comparison solution is based on the use of a 0-1 ILP formulation, and supports, for the first time, network planning for both TWIN and POADM, taking POADM as a special case of TWIN dimensioning where the underlying topology is a ring. The 0-1 ILP formulation solves the scheduling problem while minimizing the total number of transponders to be deployed and the price of wavelength per km. Note

that the cost function is approximate, as it approximates the real cost of the main components present in TWIN and POADM.

However, the model allows to estimate the efficiency of two solutions when using the basic network resources (transponders and wavelengths). In fact, all the cost components that depend on dimensioning are accounted for in the model. All the other costs are per-node costs, which do not depend on dimensioning, since nodes are known as an input.

Furthermore, the cost model is strengthened by the fact that we consider wide ranges for each cost component, that are estimated based on available data.

In Subsection 4.2.2, we review the protection methods for metro networks. Subsection 4.2.3 is devoted to the joint, 0-1 ILP dimensioning method, for TWIN and POADM. In the result Subsections 4.2.6 and 4.2.7, the cost models for the two networks are compared in different simulation scenarios, for different protection methods and the performances are discussed [67].

4.2.1 Studied network architectures

In this subsection, the characteristics of POADM and TWIN are detailed, with the focus on node components, which are part of the optimization problem.

Time-Domain Wavelength Interleaved Network (TWIN)

In TWIN, each TRX is equipped with a fast wavelength-tunable transmitter and a burst-mode wavelength-colored receiver. Relatively inexpensive non-coherent (e.g., 10 Gb/s) technology can be leveraged to implement those edge nodes, but coherent solution can be also envisaged. TWIN is expected to be realized over ROADM architecture exploiting Wavelength Selective Switches (WSSs).

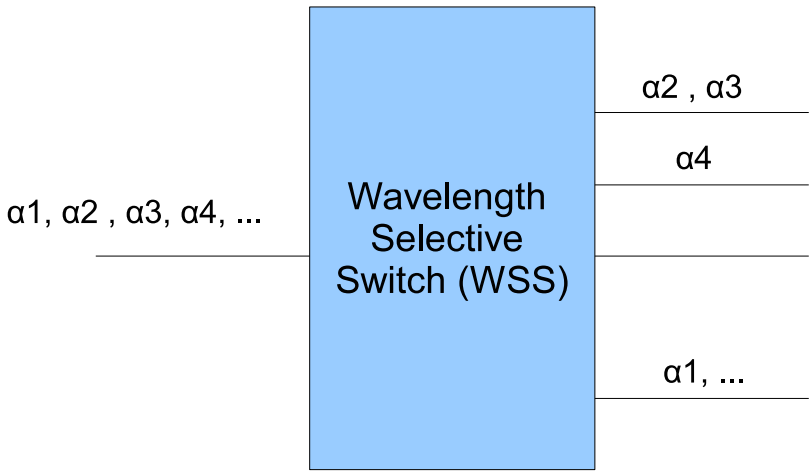


Figure 4.1 — Wavelength Selective Switch functionality.

Wavelength Selective Switches (WSS) are recently the core of current reconfigurable optical networks. It can be dynamically reconfigured to route a wavelength or block of certain wavelengths. The Wavelength Selective Switch (WSS) makes it possible for either of the wavelengths (or set of disjoint wavelengths) on either incoming fiber to be connected to either output fiber (Fig. 4.1).

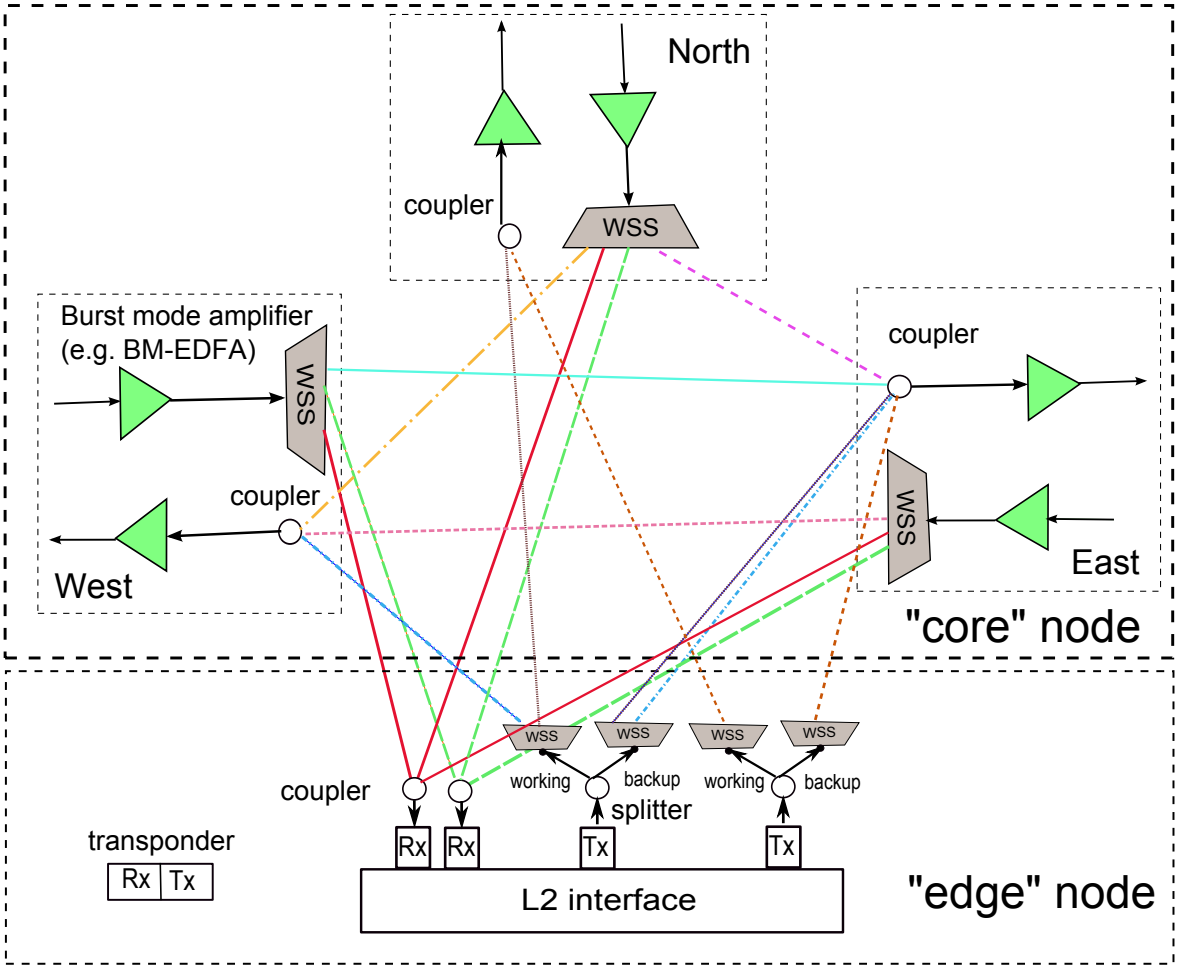


Figure 4.2 — TWIN core and edge node architecture.

In this chapter, for the dimensioning study, we propose a TWIN node architecture illustrated in Fig. 4.2. The degree of the plotted TWIN node is three. The most significant cost in TWIN are the transponders, which are composed of a transmitting (TX) and a receiving part (RX). At the emission side, the same wavelengths can be used for (disjoint) working and backup path, therefore, two WSS per TX are required. At the reception side, a dedicated RX for each wavelength is used, thus, a coupler per RX is sufficient.

The core node architecture in TWIN differs from the edge node in absence of any add/drop ports, i.e., the interface to the second layer of ISO/OSI (International Organization for Standardization/Open Systems Interconnection) network layer model ("L2

interface") and the transponders are absent (Fig. 4.2).

In this section, we assume that optical bursts in TWIN have a fixed duration: the channels are slotted and synchronous i.e., slots propagating on different wavelengths are time-aligned. Contrary to Section 3.3.2, in this section for the sake of simplicity, we assume that all links have a length corresponding to an integer number of slots.

The "schedule size", K, is an input parameter in this study (e.g., K can be computed based on the algorithm presented in Section 3.2.3). The slot allocation repeats after these K slots, during network operation.

In Section 3.3.2, we proposed a dimensioning solution for TWIN which resolves the scheduling problem together with the RWA problem. In the protection configuration, the slot allocation is needed simultaneously for both working and backup paths.

Packet Optical Add Drop Multiplexer (POADM)

In POADM, the logical topology of the network is a bidirectional ring, where each network node can use any fiber and any wavelength for the insertion or the extraction of traffic [13]. The wavelengths for the insertion/extraction process might be shared by different network destinations. The only condition is that the destinations must have the transponders able to receive on these wavelengths.

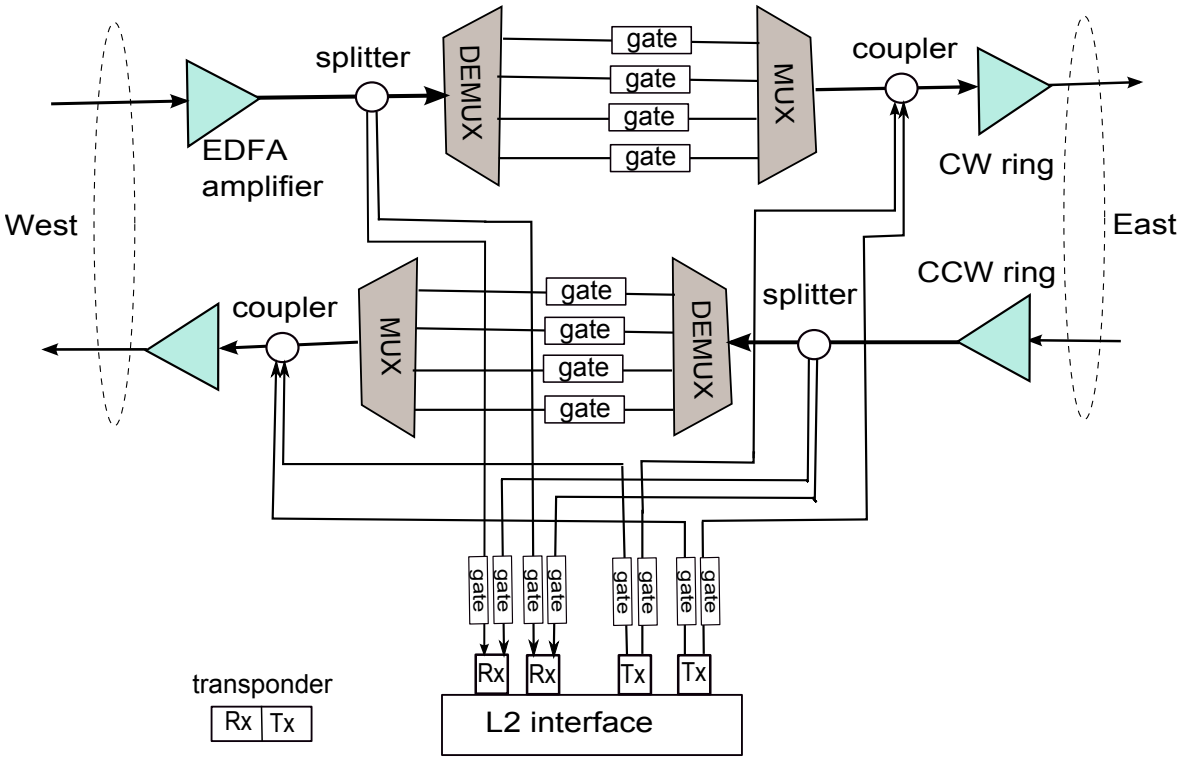


Figure 4.3 — POADM node architecture.

The simplified node architecture of POADM is illustrated in Fig. 4.3. Wavelength blocking is achieved with fast optical gates that can selectively erase any slot. These gates can be based on SOAs. The transponders share the same architecture as in TWIN, and the channels are slotted and synchronous at the slot level.

In order to minimize the network cost, for low loads, the transponders can be shared between different ring directions, with the condition that a proper scheduling is applied. This problem is addressed in the 0-1 ILP presented in Subsection 4.2.3.

The most expensive component in POADM, apart from the transponders, is the slot blocker. The number of gates in POADM is equal to the number of wavelengths used in the network (per ring direction).

For protection purpose, we consider two POADM configurations. The first one is named POADM version 1, where one fiber ring is dedicated to the working traffic and the opposite fiber is used exclusively for carrying the backup traffic. The second one, is POADM version 2, where both working and backup traffic can be carried in any directions. The advantage of POADM v.1 is a simple restoration mechanism, while the advantage of POADM v.2 is in its more efficient capacity use.

4.2.2 Comparison scenarios

We first define shared and dedicated protection. Then, we introduce a cost model, which is based on TWIN and POADM node architectures.

Classic protection methods in metro networks

We consider the problem of protecting the traffic flowing through a network against a single link failure. The basic type of protection consist in doubling the used channel capacity which is called **"dedicated"** protection. [1] The traffic between two nodes is reserved over two link-disjoint routes. For instance, if K_{p1} is the rate of a flow between nodes A and B (on the left in Fig. 4.4), on the shortest path, then, for protection, a flow of the same rate $K_{p2} = K_{p1}$ will be reserved from A to B on another path. Such protection is the most expensive in terms of capacity utilization, but it is the most effective, because the irregularities due to delays and packet loss in case of failure of any part of the network are minimized.

The second type is the protection when in the case of failure the same capacity is shared simultaneously by more than one backup path (e.g., link AC is shared for protection of paths A-B and A-E on the right in Fig. 4.4), we are speaking about "**shared**" protection. Such protection is used for example in Synchronous Optical Networking/Synchronous Digital Hierarchy (SONET/SDH) [72]. The advantage of this form of protection is the lower price, because the capacities designed to preserve the connections in case of different link failures are not used at the same time, since they can be shared by several (source, destination) node pairs. The disadvantages of

[1] More specifically, here we consider the so-called "path" protection, meaning that we use completely disjoint paths for the backup. "Link" protection also exists, but is out of scope of this study.

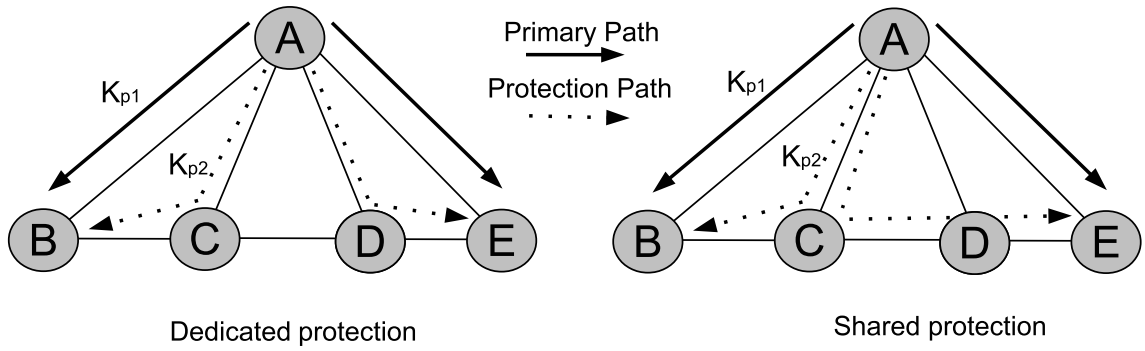


Figure 4.4 — Basic protection methods for mesh network.

this kind of protection are the higher delays needed to restore a route going through a failed link, and hence, a higher data loss upon protection switching. For both protection methods, only the traffic is protected and not the transponders.

Note that in TWIN and POADM v.2 we can consider both shared and dedicated protection. In POADM v.1, the focus is on simplifying the restoration procedure by switching all traffic between one direction and the opposite direction in case of failure. Both directions are dimensioned to carry the whole traffic matrix, which implies that the notion of shared protection does not apply.

4.2.3 0-1 ILP with protection support for TWIN and POADM

The dimensioning problem usually consists in allocating the wavelengths and the equipment (i.e., transmitters, receivers) in order to support the given traffic demands, with the objective of minimizing the CAPEX cost. As defined by the cost model in the Section 3.3.1, the CAPEX cost of the network is composed of wavelength cost, (C_{lw}), cost of transponder, TRX, (C_t), and cost of gates, (C_s). Since the wavelength cost is defined per link, we are able to encompass the impact that the physical topology has on the final network cost, as the traffic routes over greater distances have higher price. A wavelength can be shared between working and backup paths if they are path disjoint, this will lead to saving in wavelength cost. The formulation can easily be adapted to include fiber installation cost rather than wavelength leasing. When allocating the transponders, we suppose that each of them is composed of a transmitting part (TX), and a receiving part (RX), with a fixed wavelength receiver.

The 0-1 ILP solution presented next, assesses the RWA and the problem of optimal slot allocation simultaneously. We use the traffic matrix expressed in Gb/s, and we allocate the minimum needed number of slots to each traffic demand. Both, the size of the schedule K and the input traffic matrix are the input parameters in the formulation. Three different variations of the formulation are given, enabling us to study the dedicated protection, the shared protection and the case without any protection.

An important aspect of the solution is to satisfy the constraints that will prevent the collisions of the optical slots at destinations, sources, and everywhere in the network. As

in Chapter 3, we suppose that the slots that are the object of allocation are enumerated by numbers $0, 1, ..., K-1$ and that at a given moment in time, all the sources send the slots enumerated with the same number.

0-1 ILP formulation

The input parameters to the formulation are listed in Tab. 4.1. The list of output binary variables that are used is given in Tab. 4.2. The full list of 0-1 ILP constraints is given next.

Table 4.1 — Given parameters.

Name	Definition
$G(V, E)$	a non-directed graph describing the mesh topology, where V is the set of nodes, E is the set of links;
T_d	normalized number of slots to be allocated for demand d;
$\mathcal{R}_d$	set of routing paths $\{R_{d,1}, R_{d,2}, \ldots, R_{d,\lvert\mathcal{R}_d\rvert}\}$, where $R_{d,r}$ can be used to carry demand d, see below;
$R_{d,r}$	set of links which are part of a tree and are used to route a demand d;
$\text{start}(R_{d,r})$	the (set of) link(s) of $R_{d,r}$ that are attached to source nodes of demand d;
$\text{end}(R_{d,r})$	the (set of) link(s) of $R_{d,r}$ that are attached to destination nodes of demand d;
W	maximum number of wavelengths per fiber;
C_{lw}	wavelength leasing cost per km of the used fiber (link);
C_t	transponder costs;
C_s	gate costs;
L_ℓ	length of link ℓ in km;
$D_{d,\ell,r}$	delay (in number of slots) experienced by a slot emitted by the source of demand d to reach link ℓ, by using the route r;
K	number of slots to allocate (the "schedule length");
A	number of traffic demands d;
B	the maximum number of alternate routes in $\mathcal{R}_d$ (i.e., $\max_d \mathcal{R}_d$).

Objective function, minimizing the number of used channels on the links, number of transponders and number of optical gates, is:

$$\min \left(C_{lw} \cdot \sum_{\ell \in E} \sum_{w=1}^{W} L_\ell \cdot e_{\ell,w} + C_t \cdot \sum_{i=1}^{|V|} \sum_{w=1}^{W} u_{i,w} + 2 \cdot C_s \cdot |V| \cdot \sum_{w=1}^{W} y_w \right); \qquad (4.1)$$

Table 4.2 — Output variables.

Name	Definition
$x_{\ell,w}^{k,d,r}$	equal to 1 if slot k is used to carry demand d on wavelength w, by using link ℓ and route r from $\mathcal{R}_d$, and 0 otherwise;
$z_{\ell,w}^{k,d,r,b}$	equal to 1 if slot k on wavelength w, link ℓ and backup route b is used for protecting demand d on route r $(r, b \in \mathcal{R}_d)$, and 0 otherwise;
y_w	equal to 1 if the wavelength w is used in the network; and 0 otherwise;
$u_{i,w}$	equal to 1 if a TRX with fast tunable laser and a receiver for fixed wavelength w is deployed at node i, for all traffic type, and 0 otherwise;
$e_{\ell,w}$	equal to 1 if link l on wavelength w is used, and 0 otherwise;
$a_{d,r}$	equal to 1 when route r is used to carry the primary traffic for demand d, and 0 otherwise;
$c_{d,b}$	equal to 1 when route b is used to carry the backup traffic for demand d, and 0 otherwise.

Capacity constraint, ensuring the allocation of sufficient number of slots:

$$\sum_{r=1}^{|\mathcal{R}_d|} \sum_{\substack{\ell \in E, \\ \ell \in r}} \sum_{w=1}^{W} \sum_{k=1}^{K} x_{\ell,w}^{k,d,r} = T_d, \qquad \forall d \leq A; \tag{4.2}$$

Constraint ensuring that there is enough backup capacity for each working path:

$$\sum_{w=1}^{W} \sum_{\ell \in E} \sum_{k=1}^{K} x_{\ell,w}^{k,d,r} / L_\ell = \sum_{w=1}^{W} \sum_{\ell \in E} \sum_{k=1}^{K} \sum_{b=1}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b} / L_\ell, \qquad \forall d \leq A, \forall r \leq |\mathcal{R}_d|; \tag{4.3}$$

Slot-wavelength continuity constraint (i.e., ensuring that an allocated slot is on the same wavelength, at the same slot location, all the way from a source node to any destination node):

$$x_{\ell,w}^{k,d,r} = x_{\ell',w}^{k,d,r}, \qquad \forall w \leq W, \forall d \leq A, \forall r \leq |\mathcal{R}_d|, (\forall \ell, \ell' \in R_{d,r}), \forall k \in K; \tag{4.4}$$

$$z_{\ell,w}^{k,d,r,b} = z_{\ell',w}^{k,d,r,b}, \qquad \forall w, \forall d \leq A, r \leq |\mathcal{R}_d|, b \leq |\mathcal{R}_d|, (\forall \ell, \ell' \in R_{d,b}), \forall k \in K; \tag{4.5}$$

Wavelength variable constraint (used to compute the number of wavelengths):

$$\sum_{\ell \in E} \sum_{d=1}^{A} \sum_{\substack{r=1, \\ \ell \in r}}^{|\mathcal{R}_d|} \sum_{k=1}^{K} \left(x_{\ell,w}^{k,d,r} + \sum_{\substack{b=1, \\ \ell \in r}}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b} \right) \leq 2 \cdot K \cdot A \cdot |E| \cdot B^2 \cdot y_w, \qquad \forall w \leq W; \tag{4.6}$$

Link utilization variable constraint (used to determine which wavelength is used on which links):

$$\sum_{d=1}^{A}\sum_{\substack{r=1\\ \ell\in r}}^{|\mathcal{R}_d|}\sum_{k=1}^{K}\left(x_{\ell,w}^{k,d,r}+\sum_{\substack{b=1,\ell\notin r,\\ \ell\in b}}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b}\right)\leq 2\cdot A\cdot B^2\cdot K\cdot e_{\ell,w},\qquad \forall \ell\in E,\forall w\leq W;\ (4.7)$$

Constraint ensuring that two nodes do not share the same color, which is a native TWIN property (only for TWIN):

$$\sum_{i=1}^{|V|} u_{i,w}\leq 1,\qquad \forall w\leq W; \tag{4.8}$$

Constraints ensuring that each traffic demand d is carried over a single working path and protected with a single backup path (the absence of multiple paths prevents the reordering of bursts at the reception):

$$\sum_{k=1}^{K}\sum_{\substack{\ell\in E,\\ \ell\in b}}\sum_{w=1}^{W} x_{\ell,w}^{k,d,r}\leq W\cdot K\cdot |E|\cdot a_{d,r},\qquad \forall d\leq A,\forall r\leq |\mathcal{R}_d|; \tag{4.9}$$

$$\sum_{k=1}^{K}\sum_{\substack{\ell\in E,\\ \ell\in b}}\sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|}\sum_{w=1}^{W} z_{\ell,w}^{k,d,r,b}\leq B\cdot W\cdot K\cdot |E|\cdot c_{d,b},\qquad \forall d\in A,\forall b\leq |\mathcal{R}_d|; \tag{4.10}$$

$$\sum_{r=1}^{|\mathcal{R}_d|} a_{d,r}\leq 1,\qquad \forall d\leq A; \tag{4.11}$$

$$\sum_{b=1}^{|\mathcal{R}_d|} c_{d,b}\leq 1,\qquad \forall d\leq A; \tag{4.12}$$

Dedicated path protection

Fast wavelength-tunable TX capacity constraint (allocates the TXs at each node based on sent slots):

$$\sum_{w=1}^{W}\sum_{\substack{d=1,\\ i\in d}}^{A}\sum_{\substack{\ell\in E,\\ i=\mathrm{start}(\ell)}}\left(\sum_{\substack{r=1,\\ \ell=\mathrm{start}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k,d,r}+\sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|}\sum_{\substack{b=1,\\ \ell=\mathrm{start}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b}\right)\leq \sum_{w=1}^{W} u_{i,w},$$

$$\forall k\leq K,\forall i\leq |V|; \tag{4.13}$$

Colored RX capacity constraint (allocates the RXs at each node based on received slots):

$$\sum_{\substack{d=1,\\ i\in d}}^{A} \sum_{\substack{\ell\in E,\\ i=\mathrm{end}(\ell)}} \left(\sum_{\substack{r=1,\\ \ell=\mathrm{end}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k',d,r} + \sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell=\mathrm{end}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k'',d,r,b} \right) \leq u_{i,w},$$

$$\forall w \leq W, \forall k \leq K, \forall i \leq |V|; \tag{4.14}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

The slot indexes k' and k'' are calculated to account for different propagation delays of traffic flows taking different propagation paths between a source and a destination.

Constraint which avoids link collision by preventing several demands from using the same link on the same wavelength and the same slot:

$$\sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\in r}}^{|\mathcal{R}_d|} \sum_{k'=1}^{K} x_{\ell,w}^{k',d,r} + \sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell\in b}}^{|\mathcal{R}_d|} \sum_{k''=1}^{K} z_{\ell,w}^{k'',d,r,b} \leq 1, \qquad \forall \ell \in E, \forall w \leq W, \forall k \leq K; \tag{4.15}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

Shared path protection

Fast wavelength-tunable TX capacity constraint (allocates the TXs at each node based on sent slots):

$$\sum_{w=1}^{W} \sum_{\substack{d=1,\\ i\in d}}^{A} \sum_{\substack{\ell\in E,\\ i=\mathrm{start}(\ell),\\ \ell\neq e}} \left(\sum_{\substack{r=1,\\ \ell=\mathrm{start}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k,d,r} + \sum_{\substack{r=1,\\ \ell\notin r,\\ e\in r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,e\notin b\\ \ell=\mathrm{start}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b} \right) \leq \sum_{w=1}^{W} u_{i,w},$$

$$\forall k \leq K, \forall e \in E, \forall i \leq |V|; \tag{4.16}$$

$$\sum_{w=1}^{W} \sum_{\substack{d=1,\\ i\in d}}^{A} \sum_{\substack{\ell\in E,\\ i=\mathrm{start}(\ell)}} \left(\sum_{\substack{r=1,\\ \ell=\mathrm{start}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k,d,r} + \sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell=\mathrm{start}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k,d,r,b} \right) \leq A \cdot \sum_{w=1}^{W} u_{i,w},$$

$$\forall k \leq K, \forall i \leq |V|; \tag{4.17}$$

Colored RX capacity constraint (allocates the RXs at each node based on received slots):

$$\sum_{\substack{d=1,\\ i\in d}}^{A} \sum_{\substack{\ell\in E,\\ i=\mathrm{end}(\ell),\\ \ell\neq e}} \left(\sum_{\substack{r=1,\\ \ell=\mathrm{end}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k',d,r} + \sum_{\substack{r=1,\\ \ell\notin r,\\ e\in r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,e\notin b\\ \ell=\mathrm{end}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k'',d,r,b} \right) \leq u_{i,w},$$

$$\forall w \leq W, \forall k \leq K, \forall e \in E, \forall i \leq |V|; \tag{4.18}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

$$\sum_{\substack{d=1,\\ i\in d}}^{A} \sum_{\substack{\ell\in E,\\ i=\mathrm{end}(\ell)}} \left(\sum_{\substack{r=1,\\ \ell=\mathrm{end}(R_{d,r})}}^{|\mathcal{R}_d|} x_{\ell,w}^{k',d,r} + \sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell=\mathrm{end}(R_{d,b})}}^{|\mathcal{R}_d|} z_{\ell,w}^{k'',d,r,b} \right) \leq A \cdot u_{i,w},$$

$$\forall w \leq W, \forall k \leq K, \forall i \leq |V|; \tag{4.19}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

Constraint which avoids link collision by preventing several demands from using the same link on the same wavelength and the same slot:

$$\sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\in r}}^{|\mathcal{R}_d|} \sum_{k'=1}^{K} x_{\ell,w}^{k',d,r} + \sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\notin r,\\ e\in r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell\in b,\\ e\notin b}}^{|\mathcal{R}_d|} \sum_{k''=1}^{K} z_{\ell,w}^{k'',d,r,b} \leq 1,$$

$$(\forall \ell, e \in E)\ (\ell \neq e), \forall w \leq W, \forall k \leq K; \tag{4.20}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

$$\sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\in r}}^{|\mathcal{R}_d|} \sum_{k'=1}^{K} x_{\ell,w}^{k',d',r} + \sum_{d=1}^{A} \sum_{\substack{r=1,\\ \ell\notin r}}^{|\mathcal{R}_d|} \sum_{\substack{b=1,\\ \ell\in b}}^{|\mathcal{R}_d|} \sum_{k''=1}^{K} z_{\ell,w}^{k'',d,r,b} \leq A,$$

$$\forall \ell \in E, \forall w \leq W, \forall k \leq K; \tag{4.21}$$

where $k' \equiv (k - D_{d,\ell,r}) \mod K$ and $k'' \equiv (k - D_{d,\ell,b}) \mod K$.

No protection case

This case is obtained from "dedicated path protection" case, by removing the variables $z_{\ell,w}^{k,d,r,b}$ from all the constraints.

4.2.4 Applicability of the 0-1 ILP model to TWIN and POADM

Note that all the equations of the 0-1 ILP model are the same for TWIN and POADM, except the Eq. (4.8), which applies only for TWIN to account for the separation of the wavelength sets used at different destinations.

The output of the model gives the optimal slot allocation (scheduling) for both TWIN and POADM. Scheduling in POADM is thus supposed to use the static slot reservation, like in TWIN, although POADM performance might be improved by resorting to an opportunistic slot access. On the other hand, opportunistic slot access presents a specific stability problem, as shown in [95]. As the scope of the present study is to compare TWIN with POADM, we chose to use a unified dimensioning method.

4.2.5 Schedule update in TWIN

Dynamic resource allocation in TWIN has a positive impact on the energy consumption of the TRXs devices [93], and the gain in OPEX/CAPEX depends on the frequency of schedule update. In this subsection the impact of schedule update in TWIN is addressed.

In the non protection case (see the 0-1 ILP formulation proposed in Sections 3.3.2) we suppose that all destinations have the same starting time of the schedule. In case of traffic variation, a schedule is recomputed, and a source i receives the new grant from the control entity. The source starts emitting slots after a period of time equal to $\delta_{\widetilde{S_D},D} - \delta_{i,D}$, where $\delta_{\widetilde{S_D},D}$ is the propagation time from the farthest source $\widetilde{S_D}$ to the destination D, and $\delta_{i,D}$ propagation time from the source i to the destination D. This delay in emission avoids the slot collision between the slots from the updated schedule with the slots from the precedent schedule, from different sources towards the same destination. In Fig. 4.5 an example of slot collision as result of schedule update is illustrated.

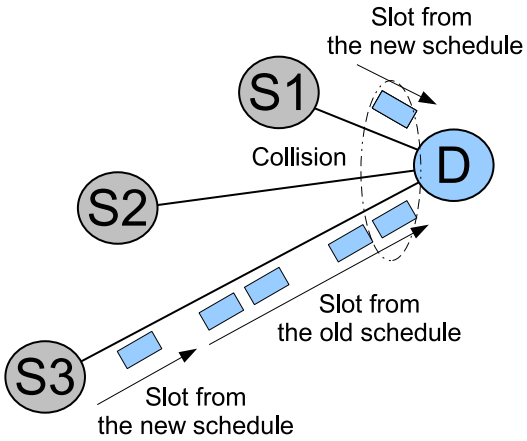


Figure 4.5 — Example of slot collision as result of schedule update.

The period of time between the moment of reception of the new grants and the moment that all the sources start emitting bursts on new schedule (here called switching time, t_{sw}) is given by:

$$t_{sw} = max(\delta_{\widetilde{S_D},D} - \delta_{i,D}), \qquad \forall i \leq |V|;$$

The t_{sw} is the negative aspect of the schedule update, and in large size networks can lead to the significant service interruption.

In addition, in case of the protection schemes where the slots used for normal operation mode are reused for backup mode, the problem of schedule update becomes even more complex. In this case, in addition to the collision of slots between sources towards the same destination, also the collision of slots between working and backup path of the same connection should be avoided.

However, in the protection schemes proposed in this thesis the slots are not shared between working and backup paths. In addition, at the destinations the collisions are avoided between working and backup paths (see Eq. 4.14 and Eq. 4.18). Thus, upon a failure the sources can automatically start emitting bursts, since there is no risk of collision. This is the one of the decisive advantage of proposed schemes.

In general case, this issue does not appear in POADM, because POADM relies on ring network, and there is no risk of collision at the destination. However, for our low cost implementation, the collision should be avoided at the destination, because the TRX are shared for low load.

4.2.6 Impact of traffic pattern on the cost efficiency

The goal of the present subsection is to evaluate the cost of different technologies and protection methods, for different traffic intensities, and to separately study the impact of different cost components on the final solution.

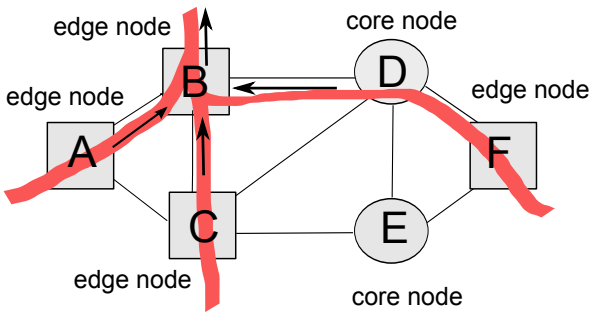


Figure 4.6 — TWIN network.

The present subsection reports the results from a commercially available LP solver (CPLEX [34]), in which the 0-1 ILP formulation has been implemented. All the results are given within an optimality gap of 10%. Other assumptions are: TRX (and channel)

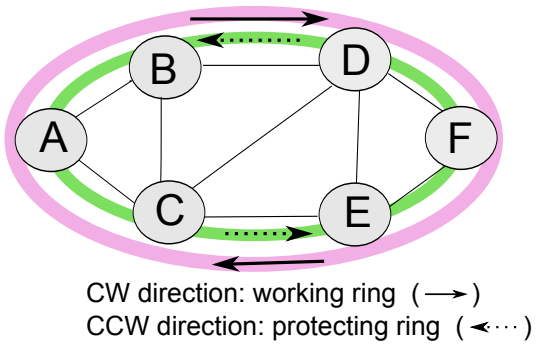


Figure 4.7 — POADM version 1.

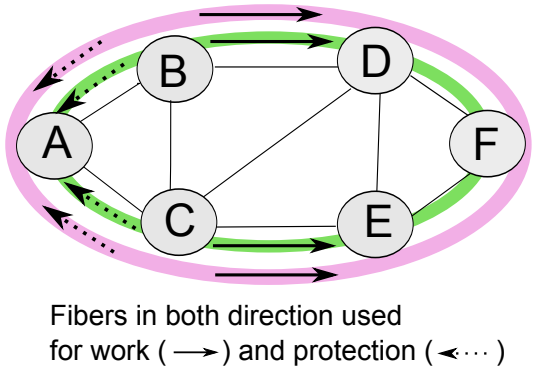


Figure 4.8 — POADM version 2.

capacity of 10 Gb/s, the link lengths expressed in integer number of slots[2], and the schedule length $K = 5$. The number of alternative paths between each pair of nodes is $B = 5$, and the available number of WDM channels is $W = 80$.

In the first set of simulations, we set the TRX cost to $C_t = 1$ (arbitrary units, a.u.), as a normalized value corresponding to \$ 1000 in reality. For such C_t, the wavelength cost is taken at its upper bound, and is fixed to $C_{lw} = 0.1$ (a.u.) km^{-1} . Finally, gate cost is fixed to $C_s = 0.05$ (a.u.), corresponding to \$ 50.

We first study the cost of the protection in TWIN and POADM for the increasing values of traffic amplitude α, in case of distributed and centralized traffic load. For readability of the results, we separately plot the results for the shared and the dedicated protection. Note that the only protection mode considered for POADM v.1 is dedicated protection.

Distributed traffic

We consider the 6-node logical topology, with 9 links, depicted in Fig. 4.6 (TWIN on native mesh topology) and Figs. 4.7 and 4.8 (POADM on native ring topology, where the POADM ring is formed on the links AB, BD, DF, FE, EC and CA). The traffic matrix is symmetric and non-centralized. It is supposed that nodes A and B (Figs. 4.6, 4.7 and 4.8) exchange traffic with same amplitude α (normalized to the TRX capacity) with nodes E and F in both directions (e.g., there are 8 traffic flows in the network).

The design cost in this scenario is plotted in Figs. 4.9 and 4.10. In the considered settings, TWIN is less cost efficient, than both versions of POADM: in the case of shared protection (Fig. 4.9) TWIN is more expensive than POADM v.1 by 10% in terms of design cost, and more expensive than POADM v.2 by more than 20%. POADM is more efficient since it can reuse the wavelengths towards any destination. In the case

[2]Note that the length of a link can be interchangeably expressed in terms of km or integer number of slots. It is also possible to account for non-integer number of slots (e.g., the 0-1 ILP formulation presented in Section 3.3.2 accounts for non-integer number of slots), but it is outside the scope of this section.

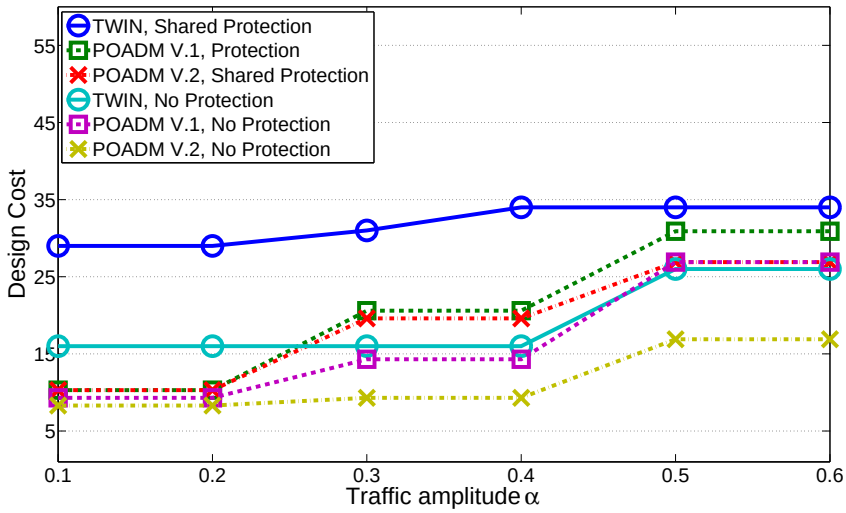


Figure 4.9 — Design cost in case of distributed traffic load, shared protection.

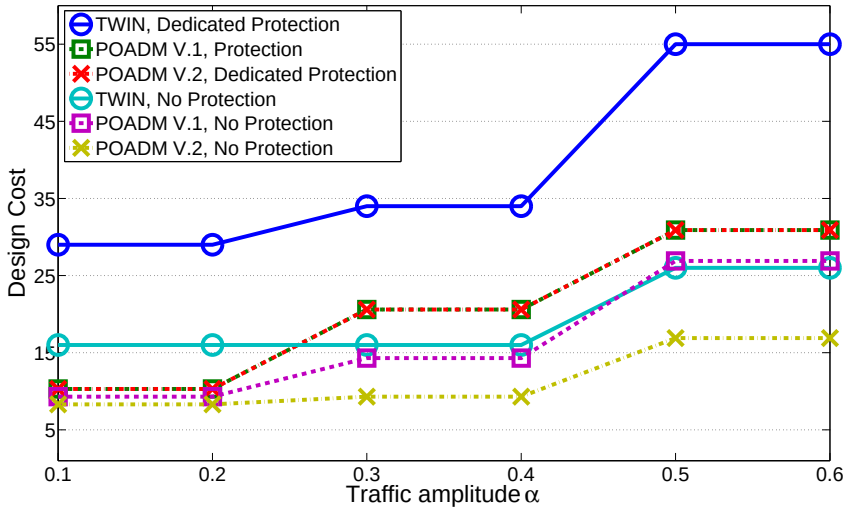


Figure 4.10 — Design cost in case of distributed traffic load, dedicated protection.

of dedicated protection (Fig. 4.10), TWIN is even more expensive: POADM v.1 and 2 are less expensive by more than 45%.

As expected, POADM v.1 protection is more expensive than POADM v.2 shared (more than 10%), and is as expensive as POADM v.2 dedicated. For all technologies, the dedicated protection method is more expensive than shared (35% more for TWIN, and 20% more for POADM v.2). A very important remark is that the cost of TWIN's dedicated protection increases much faster than of POADM. We observe that TWIN is less efficient for the increased traffic loads.

To understand better the reasons for the previous results, for the same experiment we plotted the transponder cost in Figs. 4.11 and 4.12. The results on these diagrams show that POADM v.2 requires the highest number of TRXs in shared protection case.

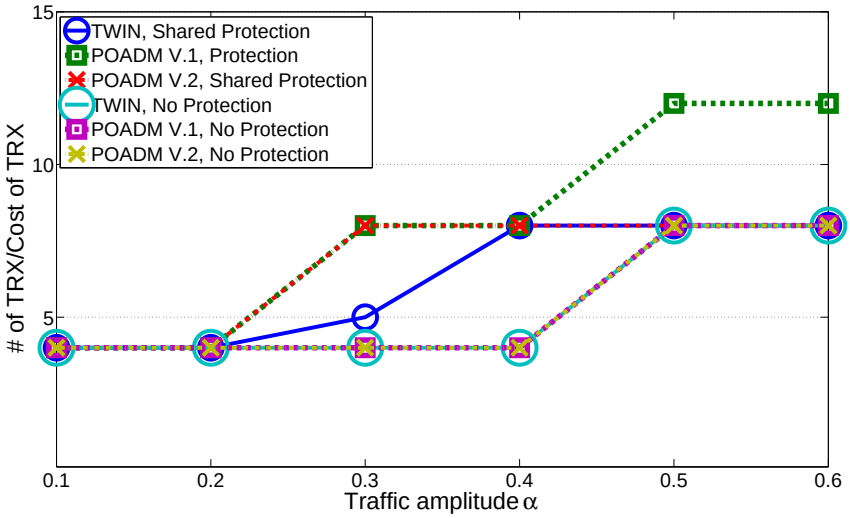


Figure 4.11 — Transponder cost in case of distributed traffic load, shared protection.

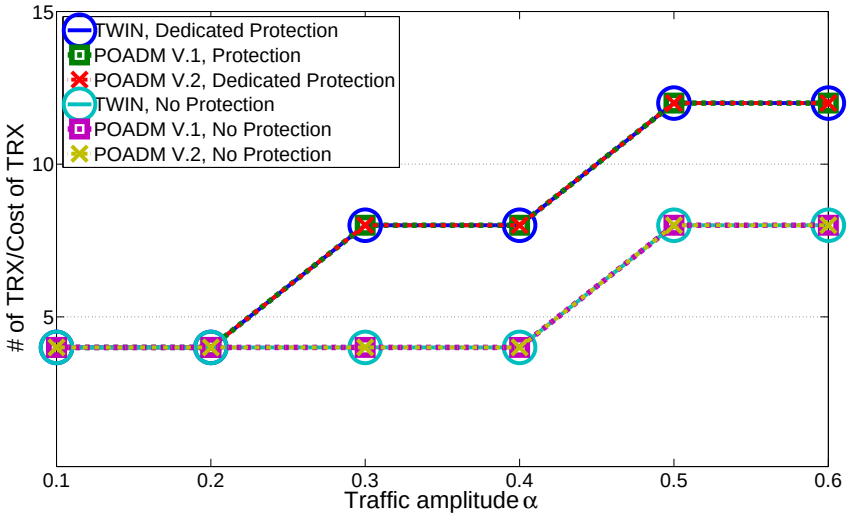


Figure 4.12 — Transponder cost in case of distributed traffic load, dedicated protection.

However, TWIN and POADM are equally efficient in terms of the allocated number of TRXs, for dedicated protection. Thus, the number of transponders is not the actual reason for the better POADM performance.

The second cost element, the wavelength cost, is plotted in Figs. 4.13 and 4.14. POADM has better performance as result of a more efficient wavelength use. The wavelength use in TWIN is poor, because of different wavelength allocations: in TWIN each destination employs a separate subset of wavelengths. This assumption is used to simplify the network core, but increases the network cost if the wavelength cost is not negligible. Let us note that TWIN's wavelength cost increases more and more rapidly at higher loads.

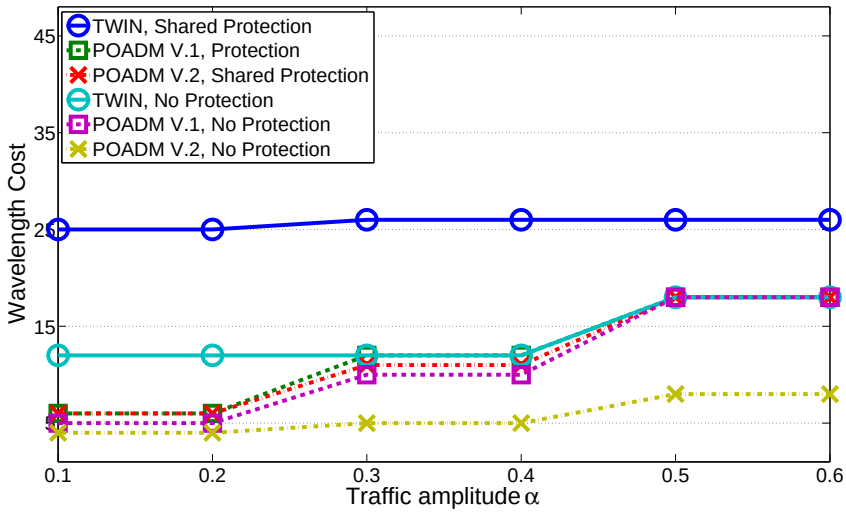


Figure 4.13 — Wavelength cost in case of distributed traffic load, shared protection.

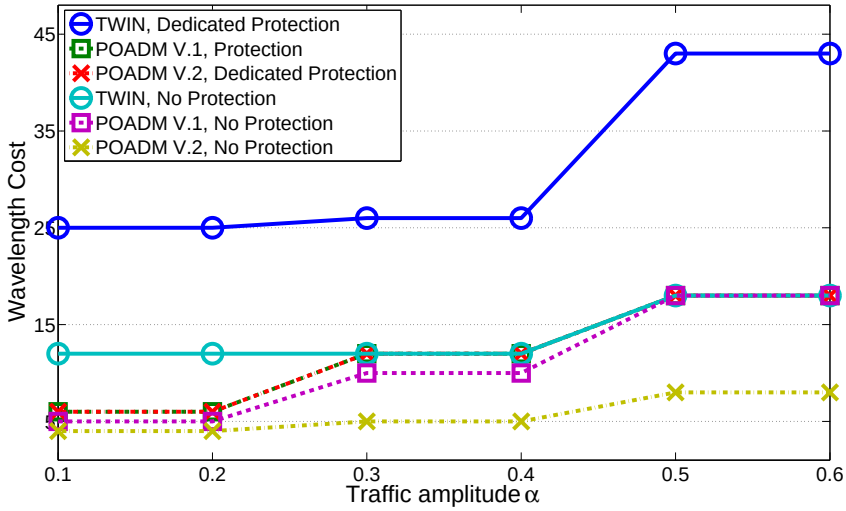


Figure 4.14 — Wavelength cost in case of distributed traffic load, dedicated protection.

The previous reasoning is confirmed with the number of wavelengths plots, given in Figs. 4.15 and 4.16. The number of wavelengths in TWIN is doubled (for shared protection) or tripled (for dedicated protection) compared with POADM. Although the cost of gates is included in the cost model only for POADM, the total design cost is much lower for POADM (Figs. 4.9 and 4.10).

Centralized traffic

Centralized traffic is a case where one of the nodes in the network Figs. 4.6 and 4.8 (in our case node A), plays the role of a gateway to the backbone network. The traffic model follows the concentration-distribution scheme, where node A sends to other 3

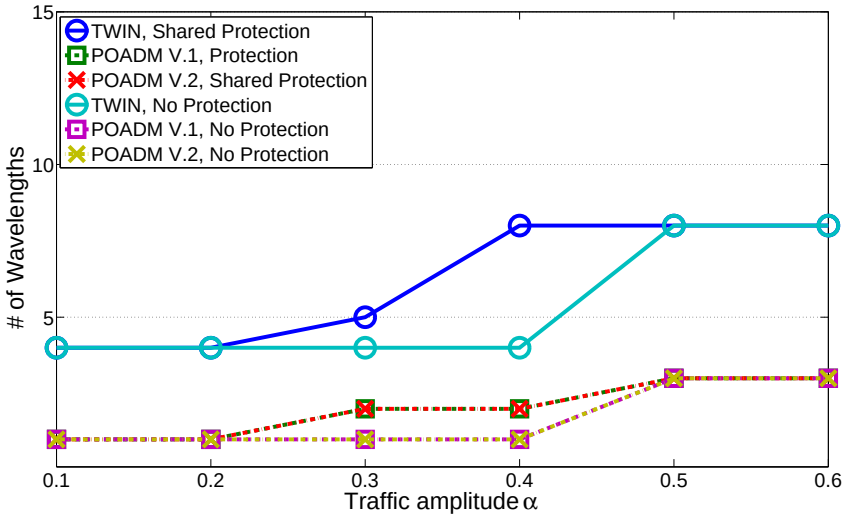


Figure 4.15 — Number of wavelengths for distributed traffic load, shared protection.

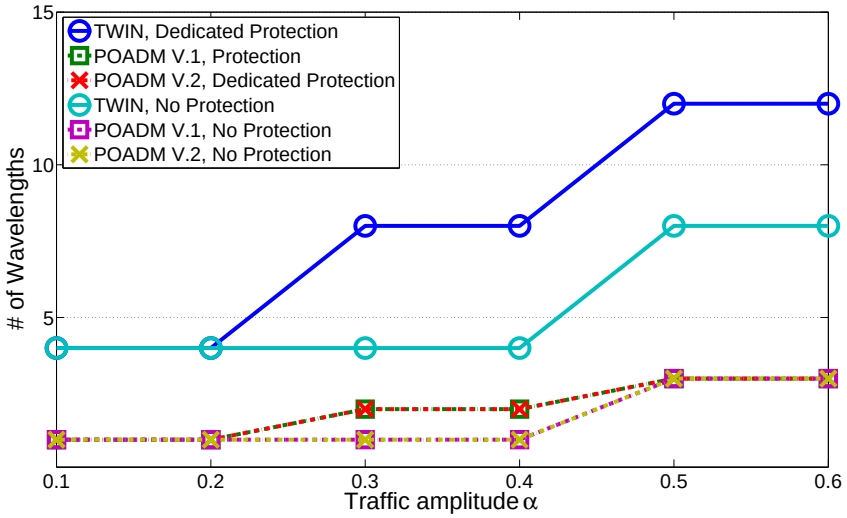


Figure 4.16 — Number of wavelengths for distributed traffic load, dedicated protection.

edge nodes a traffic of amplitude α_2, and receives $\alpha_2/10$ from them. Such client-server service model is very usual in today's metro networks.

Design cost comparison is presented in Figs. 4.17 and 4.18. The shared and dedicated protection have very similar performance for TWIN, as in distributed case, studied previously, but this is not the case for POADM. However, for shared protection at high load, TWIN is 25% more expensive than dedicated protection POADM v.1, and 80% more expensive than shared protection POADM v.2.

In the following, we compare TWIN and POADM for wide range of particular costs. We are particularly interested to see how the wavelength cost C_{lw} impacts the TWIN performance.

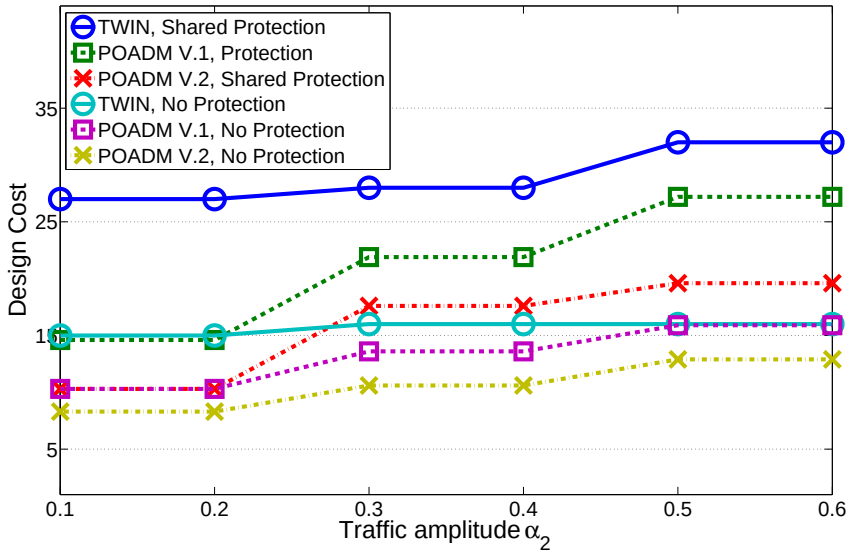


Figure 4.17 — Design cost in case of centralized traffic load, shared protection.

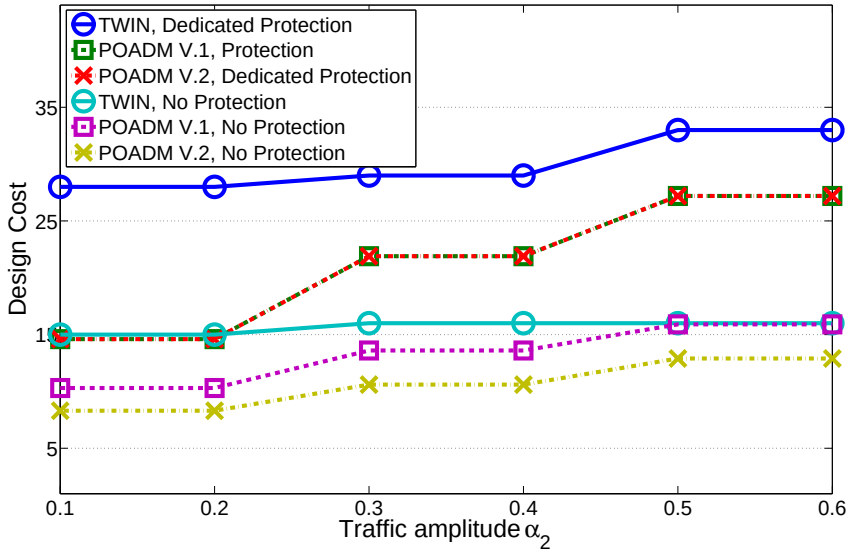


Figure 4.18 — Design cost in case of centralized traffic load, dedicated protection.

4.2.7 Impact of particular cost components on the protection efficiency

Here, the traffic amplitude α has the fixed value of 0.6 (at this load the protection has highest cost), in the same traffic model as previously. We define three scenarios in order to study separately the impact of different cost components, like C_{lw}, C_t and C_s.

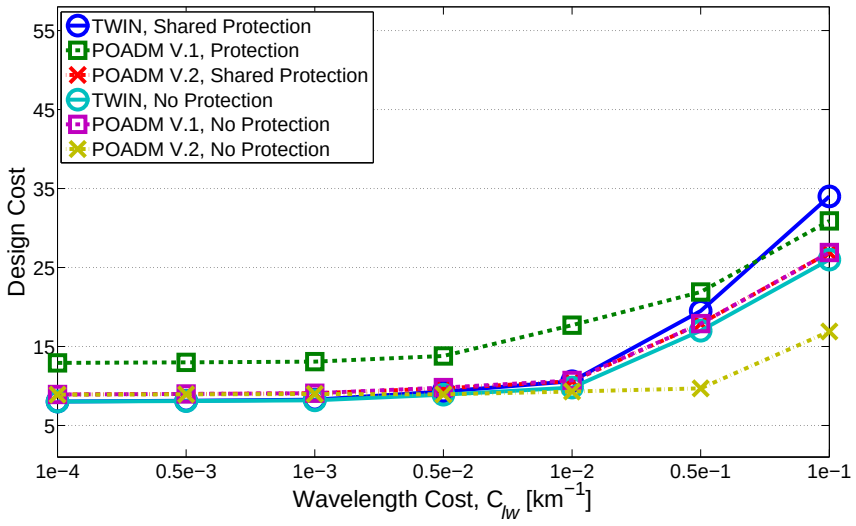


Figure 4.19 — Design cost for different values of C_{lw}, shared protection.

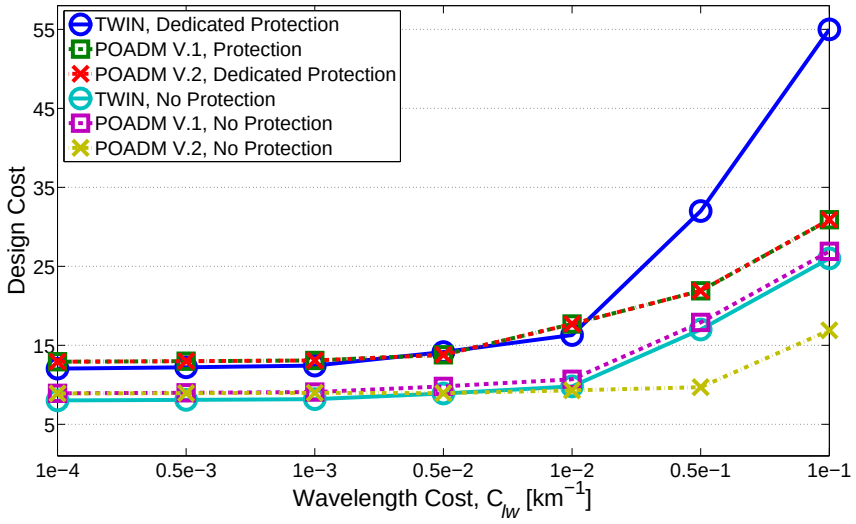


Figure 4.20 — Design cost for different values of C_{lw}, dedicated protection.

Impact of wavelength cost

It is supposed that $C_t = 1$, and $C_s = 0.05$, while C_{lw} changes in the range, between 1e-4 (a.u.) km^{-1} and 0.1 (a.u.) km^{-1}, defined by the cost model (Section 3.3.1).

Figs. 4.19 and 4.20 present the design cost for shared and dedicated protection, respectively. For shared protection, POADM v.1 is the most expensive for C_{lw} up to 0.07. For dedicated protection, however, there is no difference between different technologies for C_{lw} up to 0.01. In many real network situations, where the fibers are already installed, and there is no need for leasing of wavelengths, the values of C_{lw} can be set to very low values and in this case TWIN's wavelength inefficiency will not be penalized.

In addition, for negligible wavelength cost, in no protection case, TWIN is slightly less expensive than POADM v.1 and POADM v.2.

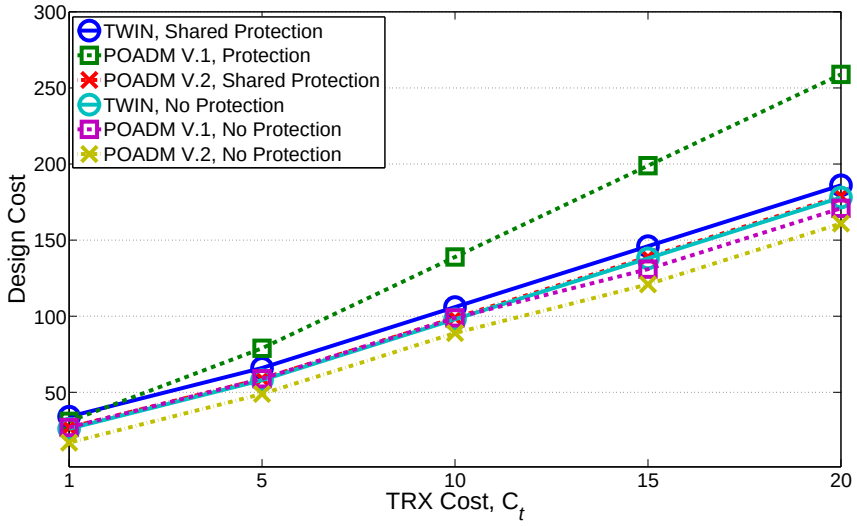


Figure 4.21 — Design cost for different values of C_t, shared protection.

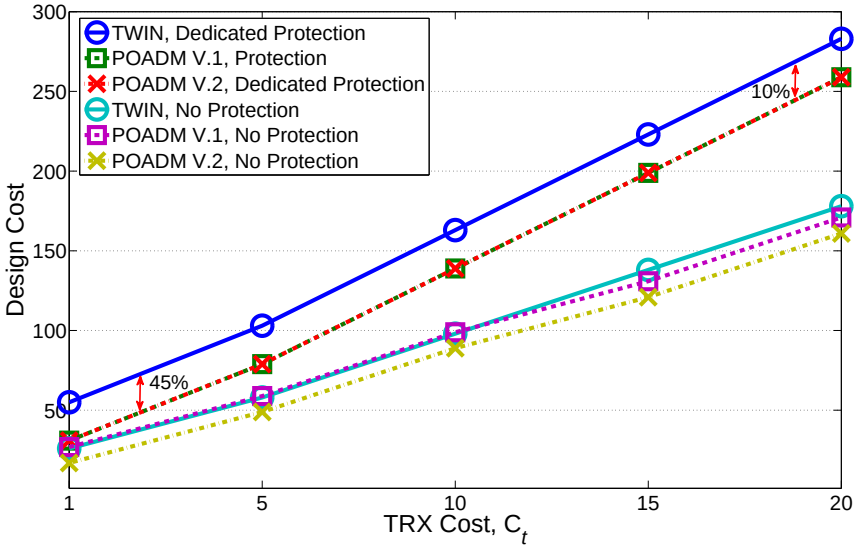


Figure 4.22 — Design cost for different values of C_t, dedicated protection.

Impact of TRX cost

Here, we suppose that C_t changes in the range [1,20] (we consider a large range of the cost model in order to account for different transmission technologies, modulation formats, bitrates and prices) and it is normalized to $ 1000, while the other cost components are fixed to $C_s = 0.05$ and $C_{lw} = 0.1$. The resulting design cost is shown in Figs. 4.21 and 4.22.

According to the simulation results on Fig. 4.21, POADM v.1 is more expensive than other protection methods in shared scenario, while POADM v.2 and TWIN have the same cost. The cost increases linearly for all technologies.

The linearity is the sign that the transponder cost does not affect the final optimization

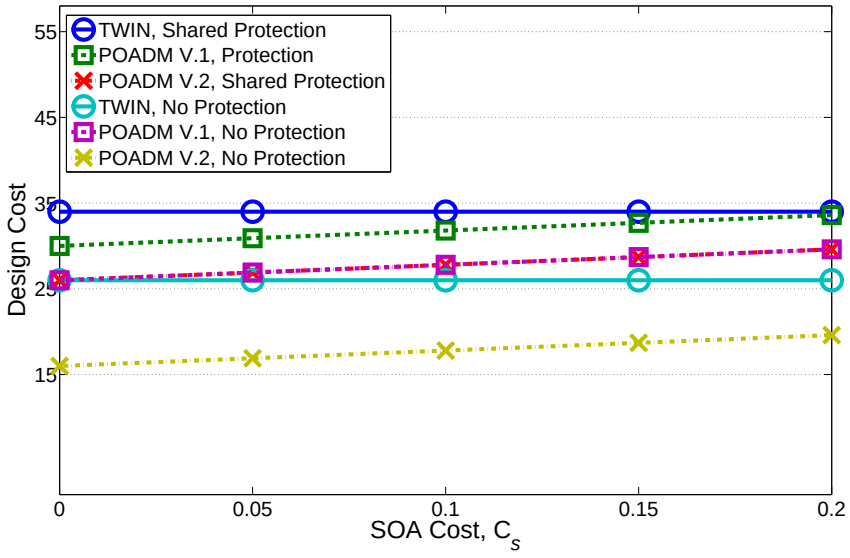


Figure 4.23 — Design cost for different values of C_s, shared protection.

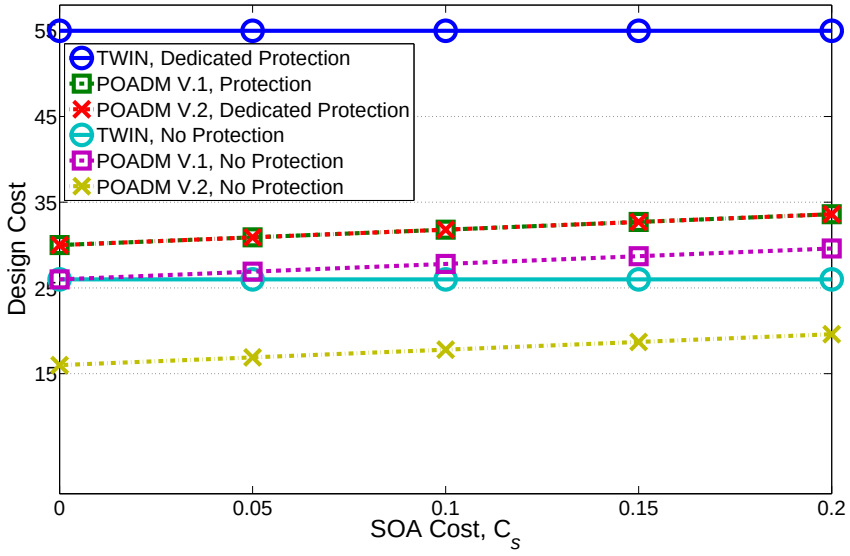


Figure 4.24 — Design cost for different values of C_s, dedicated protection.

configuration, which remains unchanged. For dedicated protection, TWIN remains more expensive than POADM (Fig. 4.22), however, the relative difference between them decreases from 45% (for $C_t = 1$) to 10% (for $C_t = 20$).

Impact of gate cost

It is supposed that $C_t = 1$, and $C_{lw} = 0.1$, while C_s changes in the extent range defined by the cost model (from 0 to 0.2). In Figs. 4.23 and 4.24 the design cost for shared and dedicated protection, respectively, is presented. TWIN is not impacted by the change of C_s. Next, the change of costs for POADM is linear, meaning, that the change of

optical gates cost does not imply a change in the network configuration.

In no protection case, TWIN is more expensive if the penalty in wavelength cost C_{lw} is higher than $2 \cdot C_s \cdot V$, and vice versa.

For both shared and dedicated protection, since the range for C_s is small, the POADM cost increases only slightly, for few percentages. TWIN remains more expensive, and the ratios of costs are almost like in the Subsection 4.2.6.

4.2.8 Conclusion on protection cost

A new joint method for network dimensioning allowing to allocate the needed resources for dedicated and shared path protection was proposed for two metro technologies, Time-Domain Wavelength Interleaved Network (TWIN), which operates on mesh topology, and Packet Optical Add Drop Multiplexer (POADM), which operates on ring topology.

For the comparison, we used a Capital Expenditure (CAPEX) cost model for the two networks, which takes into account the cost of the transponders, the cost of the wavelength, and the cost of optical gates for POADM. All the cost components that depend on dimensioning are accounted for in the model. All other costs are per-node costs, which do not depend on dimensioning, since the nodes are known as an input.

According to our results, the number of allocated transponders is similar in TWIN and POADM. The transponder cost increase does not change the network configuration, but decreases the relative difference between the two technologies. TWIN network is more sensitive to the wavelength cost, since it allocates more wavelengths than POADM, and thus, is less cost efficient than POADM, generally. Gate cost (accounted for POADM) has a very limited impact on the final solution.

4.3 Scenario 2: LTE mobile backhauling

With the evolution of the optical transport technologies, energy efficient solutions for the transport in the radio access part of cellular backhaul networks have become available. In this section, we consider the use of fine granularity optical transport solutions, optical slot switching, to interconnect the eNodeBs in the Long Term Evolution (LTE) mobile backhaul network.

Mobile traffic continues to grow fast with the increasing popularity of smartphones and video-based service. The Cisco VNI report [36] shows that the global mobile data traffic grew 70% in 2012. Meanwhile, new radio access technologies, e.g., Coordinated Multi-Point (CoMP) [50] or Cloud Radio Access Network (C-RAN), will introduce additional traffic in the mobile backhaul. This pushes mobile network operators to seek for backhaul solutions with higher capacity, lower Capital Expenditures (CAPEX)/Operational Expenditures (OPEX), and better energy efficiency.

In addition, CoMP can improve the experience for mobile users at the edge of mobile cells, through the cooperation of several neighboring eNodeBs. This however requires the exchange of large amount of data (up to 1 Gb/s) within each group of cooperating eNodeBs, which may involve 3 or more nodes, effectively requiring multicast communication.

Here, optical slot switching is applied at the metro level for backhaul networks with a few to a few dozens of nodes (eNodeBs): Fig. 4.25 shows LTE eNodeBs which should be interconnected with a mobile backhaul network. We focus on two sub-wavelength-granularity switching technologies for the backhaul: Time-Domain Wavelength Interleaved Network (TWIN), for physically meshed networks; and Packet Optical Add Drop Multiplexer (POADM), for physical rings only.

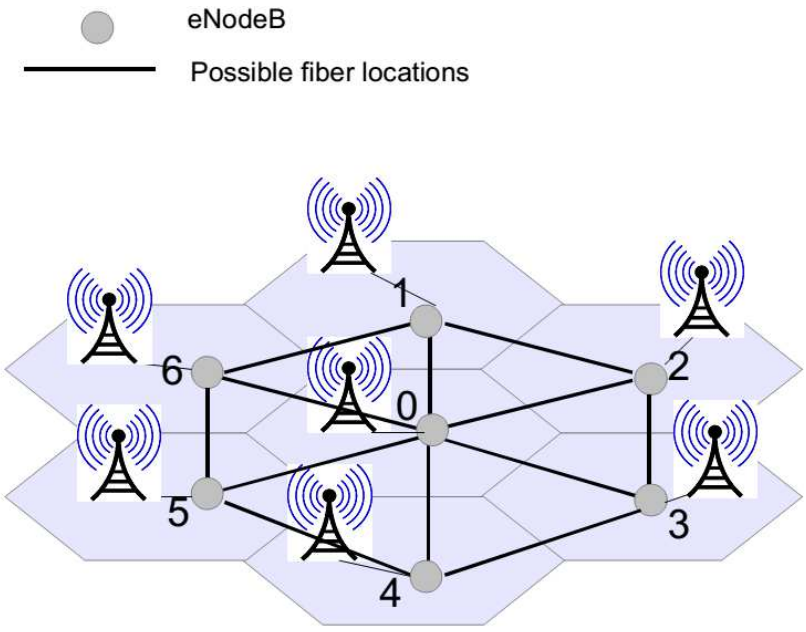


Figure 4.25 — A cellular LTE topology.

TWIN and POADM provide sub-wavelength switching granularity, and are designed to minimize the power consumption thanks to optical transparency for the transit traffic [61]. Direct node-to-node communication [102] can be efficiently leveraged to transport CoMP transmission traffic, which is typically multicast.

Differently from the previous section, this section proposes, dimensioning strategies for LTE backhaul networks that take multicasting into account and considers specific traffic profiles. We assume that TWIN and POADM exchange data according to a schedule, which can be computed in a central node. We compare a TWIN and a POADM network, in terms of the network dimensioning in a mobile backhaul scenario, and their suitability for the use in the LTE mobile backhaul network. The comparison study is based on the use of a 0-1 Integer Linear Programming (0-1 ILP) formulation, which can support network planning for both TWIN and POADM, taking POADM as a special case of TWIN where the underlying topology is a ring.

The 0-1 ILP solves the scheduling problem while minimizing the total number of transponders to be deployed and the price of a wavelength used per km. Then, the two networks are compared in a typical LTE backhaul scenario, with their ability to support various multicast traffic implementations is discussed.

In the remainder of the section we describe the specifics of the two optical transport solutions (TWIN and POADM) to be compared (Subsection 4.3.1), multicast option (Subsection 4.3.3), mathematical model and finally we give the numerical results on the comparison of the two optical transport technologies [62].

4.3.1 Studied network architectures

TWIN (Fig. 4.26) consists of simple core nodes based on passive devices, and of intelligent edge nodes that are equipped with transponders (TRX). By tuning its emitter to the appropriate wavelength, a source can send a burst to any destination. The routing at intermediate nodes is only based on the burst wavelength and thus purely passive. Thus, the logical topology of TWIN is an overlay of optical multipoint-to-point trees, one per destination.

Fig. 4.26 shows one such tree that can be used to reach eNodeB 1. In this section channels are slotted and synchronous i.e., slots propagating on different wavelengths are time-aligned and several nodes may listen on the same wavelength, which allows optical multicast.

In POADM (Fig. 4.27), the physical topology of the network is a bidirectional ring, where each network node can use any fiber wavelength for the insertion or the extraction of traffic. The wavelengths for the insertion/extraction process might be shared by different network destinations. As in TWIN each receiver is colored and can receive traffic on a single, predetermined wavelength. The node architecture of POADM for one of the directions is shown in Fig. 4.27. Wavelength blocking is achieved with fast optical gates that can selectively erase any slot. Transponders are similar to those in TWIN, and the channels are synchronous.

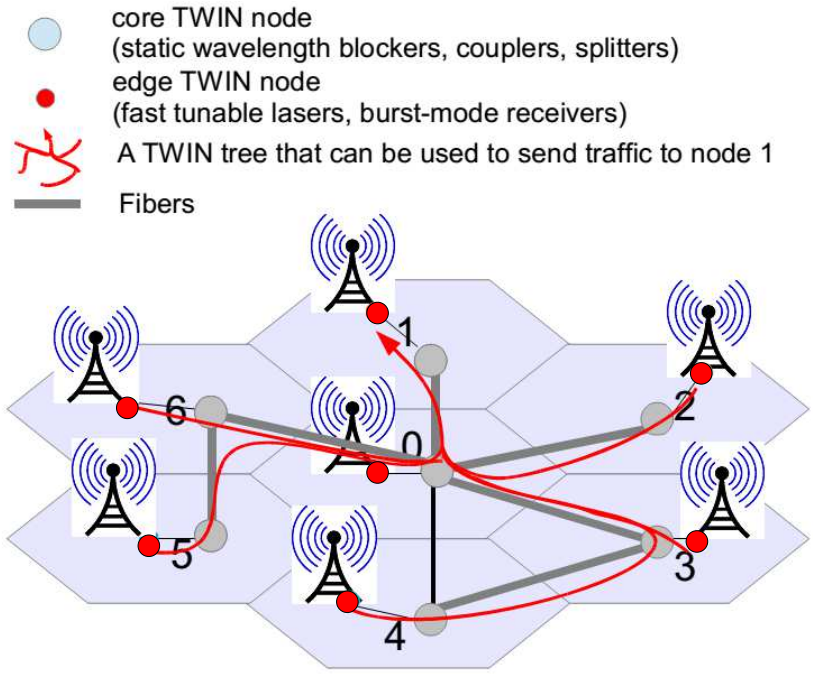


Figure 4.26 — TWIN network on LTE topology.

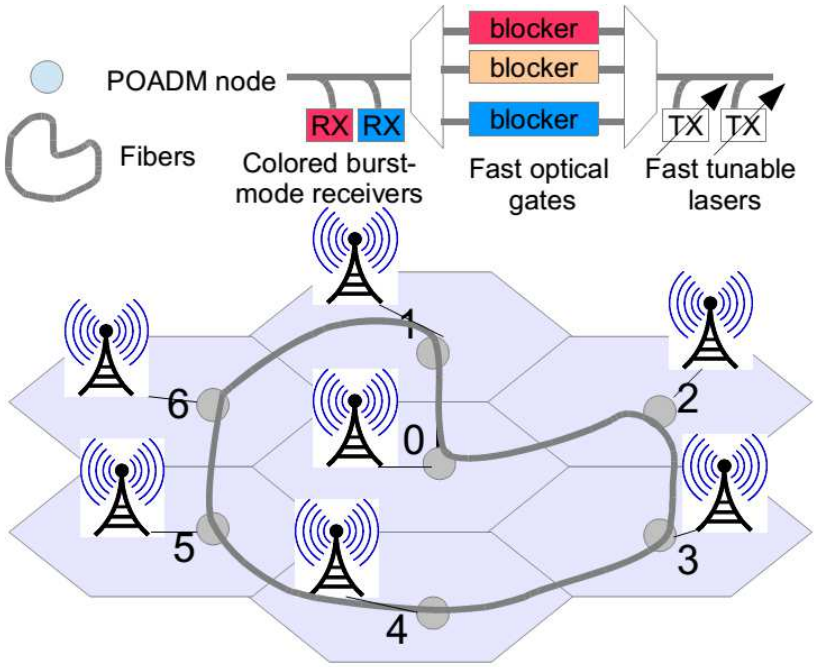


Figure 4.27 — POADM network on LTE topology.

4.3.2 Resource allocation and cost dimensioning for multicast

The 0-1 ILP formulation presented in Section 3.3.2 supports the multicast traffic. The only difference are the input structures R_d, which are sets of routing structures $R_{d,r}$, that can be used to carry demand d. Routing structure is a route (d unicast) or a tree (d multicast).

4.3.3 Multicast options

The multicast options (Fig. 4.28) that we consider in this study are:

- (MUL1): multicast and unicast traffic carried on separated wavelengths;
- (MUL2): multicast and unicast traffic carried on shared wavelengths;
- (MUL3): multicast traffic is emulated with several unicast streams: the source node emitter sends the same slot several times.

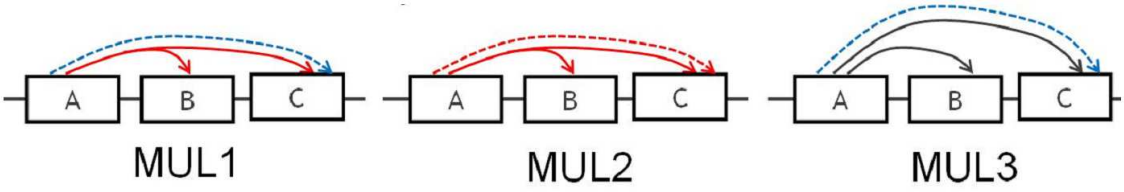


Figure 4.28 — Multicast options.

Observe that, in MUL1 and MUL2, multicasting is achieved directly at the optical layer using optical splitters (native in POADM architecture). In addition, MUL1 and MUL2 was firstly introduced in [80], however, the difference between these options was not numerically quantified.

4.3.4 Impact of traffic pattern on the cost efficiency

In the present subsection, we consider the 7-node topology depicted in Fig. 4.25, where neighboring cells can be connected with fibers. For this study the 0-1 ILP formulation presented in Sections 4.2.3 is used. All results reported here are given within an optimality gap of 15% using a commercial LP solver (CPLEX [34]). Other assumptions are: TRX (and channel) capacity of 10 Gb/s; distance between neighboring cells: 2 km (1 slot duration i.e., 10 μs); schedule length: $K = 5$. The number of alternative paths/trees between each pair of nodes is 3 (resp., 2) for both unicast and multicast traffic for TWIN (resp., POADM). For POADM, the traffic is sent on the bidirectional ring depicted in Fig. 4.27. We set the TRX cost to $C_t = 1$ (a.u.), while the wavelength cost varies from $C_{lw} = 10^{-4}$ (a.u.) km^{-1} (lower bound; guarantees that no TRX can be exchanged for a wavelength) to $C_{lw} = 0.1$ (a.u.) km^{-1} (upper bound Section 3.3.1). Note that we used a predetermined ring for POADM, while set of links actually used in TWIN is an output of the dimensioning algorithm.

We consider the LTE traffic demand for 3 cells per site (eNodeB) consisting of down/up-stream traffic (3 Gb/s, 1.5 Gb/s) and direct eNodeB-eNodeB ("X2") traffic (0.1 Gb/s) as shown in Tab. 4.3. We assume that node 0 is the eNodeB which is connected to the backbone, such that traffic to and from the backbone is included in the demand.

Table 4.3 — Unicast demand $[Gb/s]$.

Destinations \ Sources	0	1	2	3	4	5	6
0	0	3.4	3.4	3.4	3.4	3.4	3.4
1	1.9	0	0.1	0	0	0	0.1
2	1.9	0.1	0	0.1	0	0	0
3	1.9	0	0.1	0	0.1	0	0
4	1.9	0	0	0.1	0	0.1	0
5	1.9	0	0	0	0.1	0	0.1
6	1.9	0.1	0	0	0	0.1	0

Apart from the unicast traffic, there are 6 multicast groups, within which each group member sends CoMP traffic of amplitude 1 Gb/s to the other group members (these groups are: $\{0,1,2\}, \{0,3,4\}, \{0,5,6\}, \{2,3\}, \{4,5\}, \{6,1\}$, for a total of 15 flows). In addition eNodeB 0 broadcasts 3 Gb/s IPTV traffic to all other nodes. The overall reference demand is 51 Gb/s.

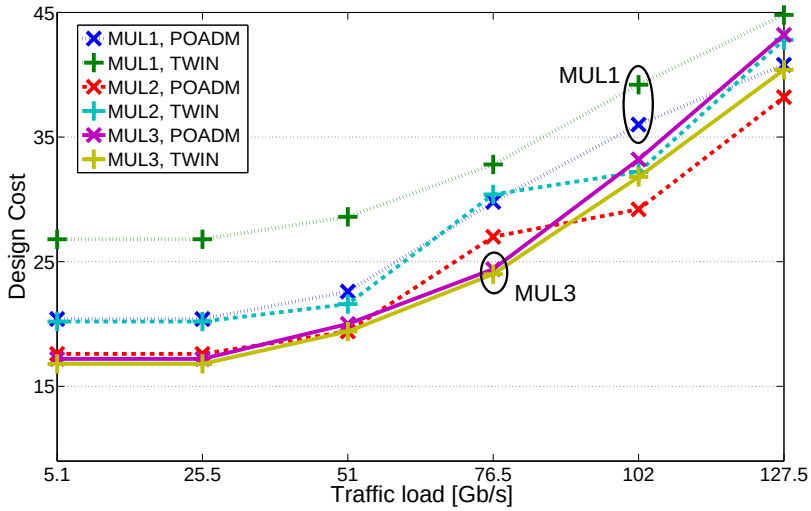


Figure 4.29 — Design cost in case of different traffic load, ($C_{lw} = 0.1$).

Firstly, in Fig. 4.29, we consider different multicast options for $C_{lw} = 0.1$, and where the reference demand is uniformly scaled. MUL1 is the most cost expensive because it requires separate TRX for the unicast and the multicast channels. For multicast options MUL1 and MUL2, POADM has a lower design cost than TWIN, while for MUL3, the two technologies have virtually the same cost. Interestingly, for sufficiently

high traffic loads, the gap in the cost for supporting multicasting options MUL1, 2 and 3 significantly decreases.

In Figs. 4.30 and 4.31, we give the design cost components (TRX cost and wavelength cost) for each technology. Fig. 4.30 indicates that the better performance of POADM is due to its more efficient TRX use: number of TRX is lower in POADM than in TWIN for MUL1 and MUL2, while for MUL3, POADM and TWIN have almost the same number of TRXs.

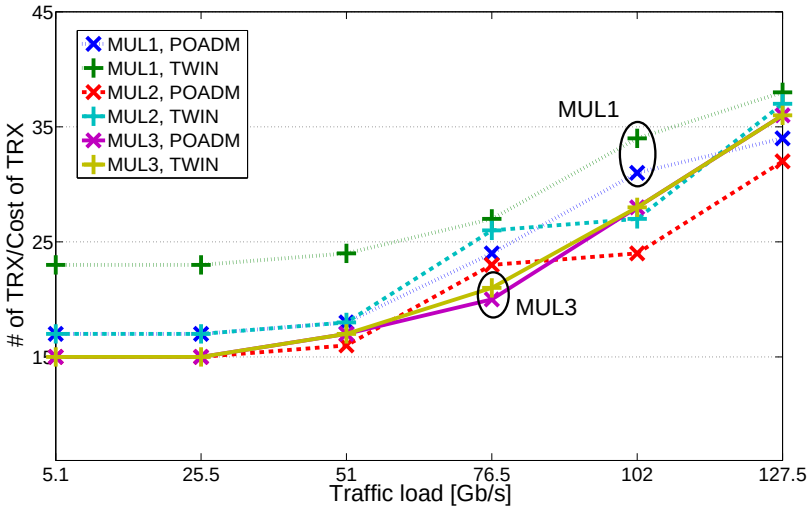


Figure 4.30 — TRX cost in case of different traffic load, ($C_{lw} = 0.1$).

TWIN's lower efficiency in accommodating the multicast traffic is due to the high number of multicast groups between the neighboring nodes in the LTE scenario, and the central position of the most congested node 0. On the other hand, TWIN's mesh topology is efficient for supporting the unicast traffic, since the shortest paths can be more often used than in POADM ring, and thus the wavelength cost can be minimized. This is confirmed by Fig. 4.31, showing that the two technologies have the similar wavelength costs for MUL1 and MUL2, with TWIN being more efficient when multicasting is emulated with several unicast streams, i.e., MUL3.

Slot replication on multicast trees can lead to increasing capacity and network costs; indeed, when slots are replicated in the optical domain (as in MUL1 and MUL2) it is not possible to schedule each copy independently after replication due to the lack of optical memory. Such physical constraints make the scheduling difficult and indeed lead to inefficient scheduling and the requirement for additional transponders and wavelengths. For this reason, MUL3 can be slightly less expensive than MUL2 for TWIN in Fig. 4.29 despite requiring sending more slots by source nodes. The gap closes when the load increases as the aforementioned inefficiency is offset by the quantity of replicated traffic to carry. Overall, the costs for MUL2 and MUL3 are similar.

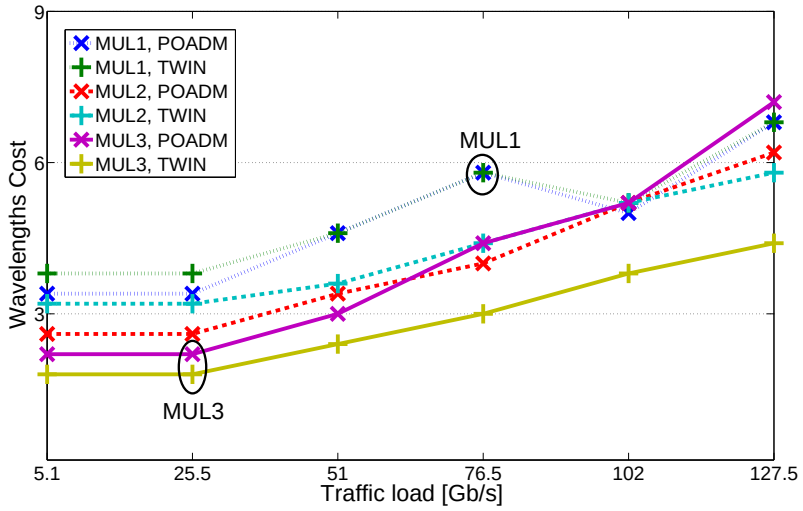


Figure 4.31 — Wavelength cost in case of different traffic load, ($C_{lw} = 0.1$).

4.3.5 Impact of the wavelengths cost on the multicast efficiency

Finally, Fig. 4.32 shows how the design cost difference between POADM and TWIN evolves with the increase of C_{lw}. Small C_{lw} applies to brownfield (no additional cost for using a wavelength) while high C_{lw} applies to greenfield (using a wavelength is included in the deployment cost), or wavelength leasing in a brownfield scenario. The conclusions drawn above apply to both scenarios to a large extent.

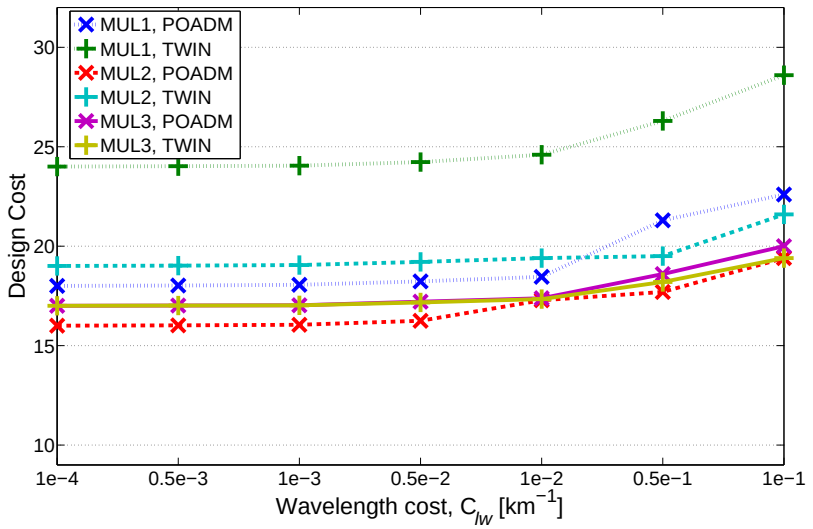


Figure 4.32 — Impact of C_{lw} on total design cost, (load=51 Gb/s).

4.3.6 Conclusion on multicast design

We studied the cost of the optical transport networks used to support Long Term Evolution (LTE)-like traffic (including next-generation features such as Coordinated MultiPoint (CoMP)) in a metro area. In this context, the multicast-enabled Packet Optical Add Drop Multiplexer (POADM) networks outperform Time-Domain Wavelength Interleaved Network (TWIN) network in terms of transponder and wavelength (or fiber) utilization. The results hold whether in a greenfield or a brownfield environment.

Moreover, we have shown that for a typical LTE network scenario, supporting the multicast traffic directly at the optical layer has almost the same cost as supporting the multicast traffic as unicast connections. The impact of latency constraint on the network cost will be the topic of further research.

4.4 Conclusion

In this chapter, we have described the Time-Domain Wavelength Interleaved Network (TWIN) and Packet Optical Add Drop Multiplexer (POADM) cost components, their characteristics and their node structure. Then, based on the current operator network, we have proposed some scenarios where we can deploy these two technologies. For the first time, a single network dimensioning model supporting protection and multicast traffic was proposed for these technologies. The 0-1 Integer Linear Programming (0-1 ILP) formulation was used to solve the problem of network planing, which includes the Routing and Wavelength Assignment (RWA) problem and the optimal schedule calculation. Finally, we have shown the results of comparative studies between TWIN and POADM concerning network survivability and Long Term Evolution (LTE) mobile backhauling.

This work was done in the frame of the CELTIC-Plus project SASER-SaveNet and it is still evolving, until the end of this year. As next steps, we aim to carry out a comparative study with larger network topology and more realistic traffic traces. As future work, we also intend to further investigate the potential of other network scenarios and other parameters. It would be particularly interesting to evaluate POADM and TWIN performance using different control planes and scenarios and indicate which scenarios are the most convenient for each technology.

Conclusion and Perspectives

THIS study addressed the problems related to the Sub-Lambda Photonically Switched Networks (SLPSN) design. We focused on Time-Domain Wavelength Interleaved Network (TWIN) paradigm since it is a lossless solution and it is designed to be deployed on a mesh topology. TWIN is a promising Optical Packet Switching (OPS)/Optical Burst Switching (OBS) technology, because of its node structure simplicity and its bandwidth flexibility. In addition, only the edge nodes perform electronic buffering, while the intermediate nodes consist of passive optical components and operate at full optical capacity without any electronic processing.

The contribution of this thesis was divided in two main parts. First contribution concerns the development of dimensioning tool that simultaneously addresses the Routing and Wavelength Assignment (RWA) and scheduling (resource allocation) problem in TWIN. The problem of scheduling in TWIN depends on path and wavelength selection. This is why, in this thesis, these problems are studied together, in order to try to understand the mutual impact and to find the optimal cost of the network design.

The second contribution is dedicated to the comparison of sub-wavelength optical switching networks in case of different scenarios. Two scenarios are considered, survivable metro networks and Long Term Evolution (LTE) mobile backhauling with the focus on TWIN and Packet Optical Add Drop Multiplexer (POADM) technologies.

A summary of the main results obtained in this thesis is as follows:

- An overview on optical switching networks was provided in Chapter 1, with the focus on optical sub-wavelength switching technologies. It briefly explained the dimensioning issues in optical networks, and the existing work on TWIN dimensioning.

- A new method for the control and the synchronization in a TWIN network was proposed in Chapter 2. This method facilitates TWIN's control and operation with an out-of-band control channel carried on one or multiple Virtual TWIN Rings (VTRs), that cover all the nodes in the network. The control scheme is centralized and the control channel accommodates the operational messages regarding the resource requests and allocations, in addition to the functional information related to the synchronization, network configuration and failures.

  The proposed control scheme for TWIN relies on the virtual rings, while the data paths are the regular TWIN trees, within the mesh topology.

  Since TWIN architecture relies on a precise knowledge of propagation delays between source-destination pairs, the main benefit of the proposed solution is to provide combined means of control, protection recovery and synchronization.

We have proposed and implemented a heuristic algorithm to choose the control rings, with a maximum ring circumference as an input parameter, and with a goal of keeping the total number of virtual rings as low as possible. Indeed, limiting the number of rings makes the passage of the administrative and control messages less complex and limits the overhead traffic on the control channel between the rings.

- In Chapter 3, we have proposed a novel slot allocating method for TWIN, that improves the network performance by minimizing the schedule length used for resource allocation.

  Therefore, the `Minimum Schedule Length (MSL)` algorithm was developed in order to find the minimum schedule length. A key subroutine of `MSL` consists of a 0-1 Integer Linear Programming (0-1 ILP) formulation, that we call `SCHEDULE`. The 0-1 ILP formulation method also includes a cyclic Quality of Service (QoS) constraint, which is used to provide a desired level of maximum queuing latency.

  To study the performance of network configurations, we have tested the solutions in the ns-2-based network simulator.

  In addition, in Section 3.4 we have presented the comprehensive algorithm called Heuristic Method for Virtual TWIN Dimensioning (HMVTD), that is a scalable solution for routing, scheduling and virtualization in TWIN optical burst switching network. HMVTD breaks down the original NP-complete problem of TWIN dimensioning into three sub-problems: tree and path construction, resource allocation and cost calculation.

  We used Kruskal's algorithm to construct the tree and Breadth-First Search (BFS) to find out the paths. The resource allocation was based on three policies: demand ordering, demand serving and slot selection. Several heuristic algorithms were proposed to perform each of the previous policies. Hence, many possible heuristics combinations are obtained. The final sub-problem, the cost calculation, evaluated the cost of the obtained solution taking into consideration the transponders and the wavelengths leasing costs.

  Firstly, the performance of each heuristics combination was assessed in terms of time complexity and cost penalty compared with the optimal solution. Besides, we evaluated the design cost of most pertinent combinations using different traffic intensities and different wavelength leasing cost. The study was extended to a real size network, considering Deutsche Telekom network.

- In Chapter 4, the cost of the optical transport solutions with sub-wavelength switching granularity was compared. Two scenarios were considered, the survivable metro networks and the LTE mobile backhauling:

  - In the first scenario, Section 4.2, a new joint method for network dimensioning was proposed for two metro technologies, allowing to allocate the needed resources for dedicated and shared path protection. One of the main advantage of the proposed protection schemes is that in case failure once the grant

is received the sources can automatically start emission on the new schedule without the risk of slot collision.

For the comparison, we used a Capital Expenditures (CAPEX) cost model for the two networks, that takes into account the cost of the transponders, the cost of the wavelength, and the cost of optical gates for POADM. All the cost components that depend on dimensioning are accounted in the model. All other costs were per-node costs, which do not depend on dimensioning, since nodes are known as an input.

According to our results, the number of allocated transponders is similar in TWIN and POADM. The transponder cost increase does not change the network configuration, but it decreases the relative difference between the two technologies. TWIN network is more sensitive to the wavelength cost, since it allocates more wavelengths than POADM.

– In the second scenario, Section 4.2, we studied the cost of the optical transport networks used to support LTE-like traffic (including next-generation features such as Coordinated MultiPoint (CoMP)) in a metro area. In this context, multicast-enabled POADM networks outperform TWIN network in terms of transponder and wavelength (or fiber) utilization. The results hold whether in a greenfield or a brownfield environment.

  Moreover, it was shown that a typical LTE network scenario, supporting the multicast traffic directly at the optical layer, has almost the same cost as supporting the multicast traffic as unicast connections. The impact of latency constraint on the network cost will be the topic of further research.

This thesis is part of the CELTIC-Plus project SASER-SaveNet that ends in December 2015. The obtained results can be a basis for a further research of the SLPSN networks, which could take some of the following directions:

- We intend to further investigate the network dimensioning considering different cost parameters for TWIN and POADM.
- It would be particularly interesting to evaluate POADM and TWIN performance using different control planes and scenarios and indicate which scenarios are the most convenient for each technology.
- HMVTD could be a promising candidate for TWIN network design with multiple virtual domains. Considering the above results, we plan to use HMVTD to dimension different size networks and assess the network performance taking into account their QoS requirements.
- A future step is to study TWIN performance and cost, considering new optical transport technologies, like flexible grid and coherent detection. This will provide a higher spectrum efficiency and flexibility.
- Moreover, TWIN can be implemented in a Software Defined Network (SDN) framework in which burst emission schedules can be periodically recomputed taking into

account dynamic traffic variation. This will lead to a flexible control plane and will provide additional functionality such as service setup and teardown, service parameter modification and service events and alarming to a TWIN infrastructure.

- In addition, we intend to further investigate the protection schemes in TWIN. For instance, for shared protection, we will consider not only slot sharing, but also, the emitter and receivers sharing among working and backup trees.
- Finally, a future study, could be the study on schedule and resource adaptation to the daily traffic load variation. This is very important, since such adaptation has a positive impact on the energy consumption of the Tx/Rx devices deployed in the TWIN network, and it will lead to OPEX minimization at the price of periodical schedule recomputation.

Bibliography

[1] Alcatel-Lucent Bell Labs. Metro network traffic growth: An architecture impact study. *Alcatel-Lucent Bell Labs Strategic White Paper*, 2013. 7

[2] Immanuel M Bomze, Marco Budinich, Panos M Pardalos, and Marcello Pelillo. The maximum clique problem. In *Handbook of combinatorial optimization*, pages 1–74. Springer, 1999. 15

[3] Edoardo Bonetto, Ahmed Triki, Esther Le Rouzic, Bernard Arzur, and Paulette Gavignet. Circuit switching and time-domain optical sub-wavelength switching technologies: Evaluations on the power consumption. In *Software, Telecommunications and Computer Networks (SoftCOM), 2012 20th International Conference on*, pages 1–5. IEEE, 2012. 8

[4] Gabriella Bosco, Vittorio Curri, Andrea Carena, Pierluigi Poggiolini, and Fabrizio Forghieri. On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers. *Lightwave Technology, Journal of*, 29(1):53–61, 2011. 28, 29

[5] Andrew Brzezinski, Iraj Saniee, Indra Widjaja, and Eytan Modiano. Flow control and congestion management for distributed scheduling of burst transmissions in time-domain wavelength interleaved networks. In *Optical Fiber Communication Conference, 2005. Technical Digest. OFC/NFOEC*, 2005. 34

[6] Andrea Carena, Vito De Feo, Jorge M Finochietto, Roberto Gaudino, Fabio Neri, Chiara Piglione, and Pierluigi Poggiolini. RingO: an experimental wdm optical packet network for metro applications. *Selected Areas in Communications, IEEE Journal on*, 22(8):1561–1571, 2004. v, xix, 8, 9, 72

[7] Ramon Casellas, Raul Muñoz, Josep M Fabrega, Michela Svaluto Moreolo, Ricardo Martinez, Lei Liu, Takehiro Tsuritani, and Itsuro Morita. Design and experimental validation of a GMPLS/PCE control plane for elastic CO-OFDM optical networks. *Selected Areas in Communications, IEEE Journal on*, 31(1):49–61, 2013. 7, 14

[8] Giorgio Cazzaniga, Christian Hermsmeyer, Iraj Saniee, and Indra Widjaja. A new perspective on burst-switched optical networks. *Bell Labs Technical Journal*, 18(3):111–131, 2013. vii, 16, 73

[9] Center for Applied Internet Data Analysis. The CAIDA anonymized 2012 Internet traces. http://www.caida.org/data/passive/passive_2012_dataset.xml. accessed on: March. 2015. 132

[10] Center for Applied Internet Data Analysis. The CAIDA anonymized 2014 Internet traces. https://data.caida.org/datasets/passive-2014/. accessed on: March. 2015. 122

[11] Center for Applied Internet Data Analysis. The CAIDA traffic summary statistics. http://www.caida.org/data/realtime/passive/. accessed on: March. 2015. 122

[12] Dominique Chiaroni, Gema Buforn, Christian Simonneau, S Coronado Etienne, and Jean-Christophe Antona. Optical packet add/drop systems. In *Optical Fiber Communication Conference*, page OThN3. Optical Society of America, 2010. v, 10

[13] Dominique Chiaroni, Géma Buforn Santamaria, Christian Simonneau, Sophie Etienne, Jean-Christophe Antona, Sébastien Bigo, and Jesse Simsarian. Packet OADMs for the next generation of ring networks. *Bell Labs Technical Journal*, 14(4):265–283, 2010. xix, 10, 11, 72, 76

[14] Mirai Chino, Ion Popescu, Takahiro Miyazaki, Yoshihiro Isaji, Shoichi Nagata, Kunisada Ashizawa, Satoru Okamoto, and Naoaki Yamanakai. Application centric energy efficient Ethernet for minimum delay. In *IEICE Technical Committee on Communication System, 2015 (written in Japanese)*. 120

[15] Ken Christensen, Pedro Reviriego, Bruce Nordman, Michael Bennett, Mehrgan Mostowfi, and Juan Antonio Maestro. IEEE 802.3 az: the road to energy efficient Ethernet. *Communications Magazine, IEEE*, 48(11):50–56, 2010. xxi, 119, 120

[16] Konstantinos Christodoulopoulos, Ioannis Tomkos, and Emmanouel Varvarigos. Elastic bandwidth allocation in flexible OFDM-based optical networks. *Journal of Lightwave Technology*, 29(9):1354–1366, 2011. 14

[17] Antonio Cianfrani, Vincenzo Eramo, Marco Listanti, and Marco Polverini. An OSPF enhancement for energy saving in IP networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pages 325–330. IEEE, 2011. 119, 125, 133, 135

[18] Pablo Coll, Javier Marenco, Isabel Méndez Díaz, and Paula Zabala. Facets of the graph coloring polytope. *Annals of Operations Research*, 116(1-4):79–90, 2002. 15

[19] Filippo Cugini, Gianluca Meloni, Francesco Paolucci, Nicola Sambo, Marco Secondini, Luca Gerardi, Luca Potì, and Piero Castoldi. Demonstration of flexible optical network based on path computation element. *Journal of Lightwave Technology*, 30(5):727–733, 2012. 7

[20] John D'Ambrosia, David Law, and Mark Nowell. 40 Gigabit Ethernet and 100 Gigabit Ethernet technology overview. *Ethernet Alliance White Paper*, 2008. 126

[21] Ning Deng, Qingsong Xue, Mo Li, Andrew Lord, Peter Willis, Shiyi Cao, and Zhiyong Feng. Network modelling and techno-economic analysis of optical burst ring for metropolitan applications. In *European Conference and Exhibition on Optical Communication*, pages P5–07. Optical Society of America, 2012. v, 1

[22] Lars Dittmann, Chris Develder, Dominique Chiaroni, Fabio Neri, Franco Callegati, W Koerber, Alexandros Stavdas, Monique Renaud, Albert Rafel, Josep Solé-Pareta, Walter Cerroni, Nelly Leligou, Lars Dembeck, Brian B Mortensen, Mario Pickavet, Nicolas Le Sauze, Mike Mahony, Bela Berde, and G Eilenberger. The European IST project DAVID: a viable approach toward optical packet switching. *Selected Areas in Communications, IEEE Journal on*, 21(7):1026–1040, 2003. v, xix, 10

[23] Marcus Duelk and Martin Zirngibl. 100 Gigabit Ethernet-applications, features, challenges. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–5. IEEE, 2006. 125

[24] António Eira, João Pedro, and João Pires. Cost-optimized dimensioning of translucent WDM networks with mixed-line-rate spectrum-flexible channels. In *High Performance Switching and Routing (HPSR), 2012 IEEE 13th International Conference on*, pages 185–190. IEEE, 2012. 14

[25] Tarek S El-Bawab. *Optical switching*. Springer Science & Business Media, 2008. 7, 8

[26] Fiber24. url: http://shop.fiber24.net/index.php/en/SFP-Modules-Transceiver-Modules/c-SFP-PLUS-MODULE. accessed on: June. 2015. 47

[27] Greg N Frederickson. A single source shortest path algorithm for a planar distributed network. In *STACS 85*, pages 143–150. Springer, 1985. 62

[28] Andrea Fumagalli, Isabella Cerutti, and Marco Tacca. Optimal design of survivable mesh networks based on line switched WDM self-healing rings. *IEEE/ACM Transactions on Networking (TON)*, 11(3):501–512, 2003. 20

[29] Paul Ghobril, Clara Zaiter, and Esther Le Rouzic. Rearrangement: From wavelength routed to sliced-spectrum optical networks. In *Transparent Optical Networks (ICTON), 2012 14th International Conference on*, pages 1–4. IEEE, 2012. 14

[30] Annie Gravey, Bogdan Uscumlic, Yvan Pointurier, Ion Popescu, Philippe Gravey, and Luay Alahdab. Efficient resource allocation in time-domain wavelength interleaved networks. In *Optical Network Design and Modeling, 2014 International Conference on*, pages 293–298. IEEE, 2014. 34, 35

[31] Chamara Gunaratne, Ken Christensen, and Stephen W Suen. Ngl02-2: Ethernet adaptive link rate (alr): Analysis of a buffer threshold policy. In *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*, pages 1–6. IEEE, 2006. 132

[32] Mark Gustlin, Gary Nicholl, and Oded Trainin. 100GE and 40GE PCS and MLD proposal. *IEEE P802. 3ba, IEEE, Jan*, 2008. 128

[33] Pierre Hansen, Martine Labbé, and David Schindl. Set covering and packing formulations of graph coloring: algorithms and first polyhedral results. *Discrete Optimization*, 6(2):135–147, 2009. 15

[34] IBM. ILOG CPLEX: High-performance mathematical programming engine. url: http://www.ibm.com/. accessed on: May. 2015. 41, 60, 85, 99

[35] Cisco Visual Networking Index. Forecast and methodology, 2012-2017. *White paper*, 29 May 2013. v, 1

[36] Cisco Visual Networking Index. Global mobile data traffic forecast update, 2012–2017. *White paper*, 29 May 2013. 96

[37] INTUNE Network. url: http:/www.intunenetworks.com. accessed on: May. 2015. 72

[38] Osamu Ishida. 40/100GbE technologies and related activities of IEEE standardization. In *Optical Fiber Communication Conference*. Optical Society of America, 2009. 126

[39] ITU-T. Draft revised G.694.1 version 1.3, Unpublished ITU-T Study Group 15, Question 6. *International Telecommunication Union-Telecommunication*. 7

[40] ITU-T. Terms and definitions for sub-lambda photonically switched networks. *International Telecommunication Union-Telecommunication*, Geneva. 2012. 1, 8

[41] Na Jihoon, Gopinath Mudhana, Seon Young Ryu, and Byeong Ha Lee. Implementation of an all-fiber variable optical delay line with a pair of linearly chirped fiber bragg gratings. *IEICE transactions on electronics*, 88(5):925–932, 2005. 52

[42] Masahiko Jinno, Bartlomiej Kozicki, Hidehiko Takara, Atsushi Watanabe, Yoshiaki Sone, Takafumi Tanaka, and Akira Hirano. Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network [Topics in Optical Communications]. *Communications Magazine, IEEE*, 48(8):138–145, 2010. 7

[43] Masahiko Jinno, Hidehiko Takara, Kazushige Yonenaga, and Akira Hirano. Virtualization in optical networks from network level to hardware level [invited]. *Journal of Optical Communications and Networking*, 5(10):A46–A56, 2013. 54

[44] Donald B Johnson. Finding all the elementary circuits of a directed graph. *SIAM Journal on Computing*, 4(1):77–84, 1975. 28

[45] Mahmoud I Kamel, Long Bao Le, and André Girard. LTE wireless network virtualization: Dynamic slicing via flexible scheduling. In *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*, pages 1–5. IEEE, 2014. 54

[46] A Kershenbaum and R Van Slyke. Computing minimum spanning trees efficiently. In *Proceedings of the ACM annual conference-Volume 1*, pages 518–527. ACM, 1972. 57, 62

[47] Junhyuk Kim, Chankyun Lee, and June-Koo Kevin Rhee. Traffic off-balancing algorithm for energy efficient networks. In *SPIE/OSA/IEEE Asia Communications and Photonics*, pages 83100C–83100C. International Society for Optics and Photonics, 2011. 128

[48] Miroslaw Klinkowski and Krzysztof Walkowiak. Routing and spectrum assignment in spectrum sliced elastic optical path network. *IEEE Communications Letters*, 15(8):884–886, 2011. 14

[49] Diego Kreutz, Fernando MV Ramos, Esteves P Verissimo, Esteve C Rothenberg, Siamak Azodolmolky, and Steve Uhlig. Software-defined networking: A comprehensive survey. *proceedings of the IEEE*, 103(1):14–76, 2015. 54

[50] Daewon Lee, Hanbyul Seo, Bruno Clerckx, Eric Hardouin, David Mazzarese, Satoshi Nagata, and Krishna Sayana. Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges. *Communications Magazine, IEEE*, 50(2):148–155, 2012. 96

[51] Martin Maier. *Optical switching networks*. Cambridge University Press, 2008. 8

[52] Takahiro Miyazaki, Ion Popescu, Mirai Chino, Xiaoyu Wang, Kunitaka Ashizawa, Satoru Okamoto, Malathi Veeraraghavan, and Naoaki Yamanaka. High speed 100GE adaptive link rate switching for energy consumption reduction. In *Optical Network Design and Modeling, 2015 International Conference on*. IEEE, 2015. 125

[53] Biswanath Mukherjee. *Optical WDM networks*. Springer Science & Business Media, 2006. 6

[54] Avishek Nag, Massimo Tornatore, and Biswanath Mukherjee. Optical network design with mixed line rates and multiple modulation formats. *Lightwave Technology, Journal of*, 28(4):466–475, 2010. 14

[55] Reza Nejabati, Eduard Escalona, Shuping Peng, and Dimitra Simeonidou. Optical network virtualization. In *Optical Network Design and Modeling (ONDM), 2011 15th International Conference on*, pages 1–5. IEEE, 2011. 54

[56] Juniper Networks. MPLS transport profile (MPLS-TP), a set of enhancements to the rich MPLS toolkit. *White paper*, 2011. 73

[57] NTT Electronics. NTT Electronics Ships Industry's First 20nm Low-Power Coherent DSP. 133

[58] Satoru Okamoto, Takahiro Miyazaki, Ion Popescu, Mirai Chino, Naoaki Yamanaka, Malathi Veeraraghavan, and Eiji Oki. Proposal of the dummy signal insertion method for realizing the adaptive link rate of the serial transmission type Flex Ethernet. In *IEICE General Conference, March 2015 (written in Japanese)*. 131

[59] Ankitkumar N Patel, Philip N Ji, Jason P Jue, and Ting Wang. Routing, wavelength assignment, and spectrum allocation algorithms in transparent flexible optical WDM networks. *Optical Switching and Networking*, 9(3):191–204, 2012. 14

[60] Oscar Escriba Pedrola, Alberto Casales Castro, Luis Velasco, Marc Ramirez Ruiz, Juan P Fernández-Palacios, and Davide Careglio. CAPEX study for a multilayer IP/MPLS-over-flexgrid optical network. *Journal of Optical Communications and Networking*, 4(8):639–650, 2012. 14

[61] Yvan Pointurier, Bogdan Uscumlic, Isabella Cerutti, Annie Gravey, and J-C Antona. Dimensioning and energy efficiency of multi-rate metro rings. *Lightwave Technology, Journal of*, 30(22):3552–3564, 2012. 72, 97

[62] Yvan Pointurier, Bogdan Uscumlic, Ion Popescu, Annie Gravey, Qing Wei, and Matthias Lott. A comparison of subwavelength optical switching networks for LTE mobile backhauling. In *Transparent Optical Networks (ICTON), 2014 16th International Conference on*, pages 1–4. IEEE, 2014. 97

[63] Christina Tanya Politi, Vasileios Anagnostopoulos, Christos Matrakidis, Alexandros Stavdas, Andrew Lord, Víctor López, and Juan Pedro Fernández-Palacios. Dynamic operation of flexi-grid OFDM-based networks. In *Optical Fiber Communication Conference*, pages OTh3B–2. Optical Society of America, 2012. 28, 29

[64] Ion Popescu, Isabella Cerutti, Nicola Sambo, and Piero Castoldi. On the optimal design of a spectrum-switched optical network with multiple modulation formats and rates. *Optical Communications and Networking, IEEE/OSA Journal of*, 5(11):1275–1284, 2013. 14

[65] Ion Popescu, Takahiro Miyazaki, Mirai Chino, Xiaoyu Wang, Satoru Okamoto, Annie Gravey, Philippe Gravey, Malathi Veeraraghavan, Maite Brandt-Pearce, and Naoaki Yamanaka. Application-centric energy-efficient ethernet with quality of service support. *Electronics Letters*, 51(15):1165–1167, 2015. 119

[66] Ion Popescu, Lida Sadeghioon, Annie Gravey, Philippe Gravey, and Michel Morvan. Synchronization of the time-domain wavelength interleaved networks. In *Transparent Optical Networks (ICTON), 2013 15th International Conference on*, pages 1–4. IEEE, 2013. 22

[67] Ion Popescu, Bogdan Uscumlic, Yvan Pointurier, Annie Gravey, Philippe Gravey, and Michel Morvan. A cost comparison of survivable subwavelength switching optical metro networks. In *Teletraffic Congress (ITC), 2014 26th International*, pages 1–9. IEEE, 2014. 74

[68] Ion Popescu, Bogdan Uscumlic, Yvan Pointurier, Annie Gravey, Philippe Gravey, and Michel Morvan. Cost of protection in time-domain wavelength interleaved networks. In *Networks and Optical Communications-(NOC), 2014 19th European Conference on*, pages 108–114. IEEE, 2014. 73

[69] Ion Popescu, Bogdan Uscumlic, Ahmed Triki, Yvan Pointurier, Annie Gravey, and Philippe Gravey. Scalable routing, scheduling and virtualization for TWIN optical burst switching networks. In *Networks and Optical Communications (NOC), 2015 20th European Conference on*. IEEE, 2015. 54

[70] Yang Qin, Daojun Xue, Chee Kheong Siew, Iraj Saniee, and Indra Widjaja. A topology based dynamic traffic scheduling in time-domain wavelength interleaved networks. In *Networks, 2004.(ICON 2004). Proceedings. 12th IEEE International Conference on*, volume 1, pages 142–146. IEEE, 2004. 16, 20

[71] S Ramamurthy, Laxman Sahasrabuddhe, and Biswanath Mukherjee. Survivable WDM mesh networks. *Journal of Lightwave Technology*, 21(4):870, 2003. 73

[72] ITU-T recommendation G.841. Types and characteristics of SDH network protection architectures. *International Telecommunication Union-Telecommunication*, July 1995. 77

[73] Pedro Reviriego, Bas Huiszoon, Victor López, RB Coenen, Jose Alberto Ramirez-Hernández, and Juan Antonio Maestro. Improving energy efficiency in IEEE 802.3 ba high-rate Ethernet optical links. *Selected Topics in Quantum Electronics, IEEE Journal of*, 17(2):419–427, 2011. 125, 128, 130, 133

[74] Philippe Robert and James Roberts. A flow-aware MAC protocol for a passive optical metropolitan area network. In *Teletraffic Congress (ITC), 2014 23rd International*, pages 166–173. IEEE, 2011. 35

[75] Kevin Ross, Nicholas Bambos, Krishnan Kumaran, Iraj Saniee, and Indra Widjaja. Dynamic scheduling of optical data bursts in time-domain wavelength interleaved networks. In *High Performance Interconnects, 2003. Proceedings. 11th Symposium on*, pages 108–113. IEEE, 2003. 16, 20, 34, 35, 36

[76] Kevin Ross, Nicholas Bambos, Krishnan Kumaran, Iraj Saniee, and Indra Widjaja. Scheduling bursts in time-domain wavelength interleaved networks. *Selected Areas in Communications, IEEE Journal on*, 21(9):1441–1451, 2003. v, vi, vii, 12, 15, 34

[77] Cristina Rottondi, Massimo Tornatore, Achille Pattavina, and Giancarlo Gavioli. Routing, modulation level, and spectrum assignment in optical metro ring networks using elastic transceivers. *Journal of Optical Communications and Networking*, 5(4):305–315, 2013. 14

[78] Lida Sadeghioon, Paulette Gavignet, Abdelfateh Triki, Jean-Luc Barbey, Esther Le Rouzic, Laurent Bramerie, Vincent Alaiwan, Eric Borgne, Christophe Betoule, Bernard Arzur, and Arnaud Carer. First experimental demonstration of real-time

orchestration in a Multi-head metro network. In *Transparent Optical Networks (ICTON), 2014 16th International Conference on*, pages 1–4. IEEE, 2014. 17

[79] Lida Sadeghioon, Annie Gravey, and Philippe Gravey. Rapid protection schemes in an all-optical packet metro ring. In *Networks and Optical Communications (NOC), 2012 17th European Conference on*, pages 1–6. IEEE, 2012. 26

[80] Lida Sadeghioon, Annie Gravey, Bogdan Uscumlic, Philippe Gravey, and Michel Morvan. Full featured and lightweight control for optical packet metro networks [Invited]. *Journal of Optical Communications and Networking*, 7(2):A235–A248, 2015. 17, 99

[81] Lida Sadeghioon, Ion Popescu, Bogdan Uscumlic, Philippe Gravey, and Annie Gravey. Virtual ring based protection for time-domain wavelength interleaved network. In *Network and Optical Communications (NOC), 2013 18th Conference on*, pages 121–128. IEEE, 2013. 21

[82] Lida Sadeghioon, Bogdan Uscumlic, Philippe Gravey, and Annie Gravey. Fully transparent design of a hybrid optical packet/circuit metropolitan area network. In *Optical Network Design and Modeling (ONDM), 2013 17th International Conference on*, pages 263–268. IEEE, 2013. 20

[83] Nicola Sambo, Piero Castoldi, Filippo Cugini, Giulio Bottari, and Paola Iovanna. Toward high-rate and flexible optical networks. *Communications Magazine, IEEE*, 50(5):66–72, 2012. 7, 14

[84] Iraj Saniee and Indra Widjaja. Design and performance of randomized schedules for time-domain wavelength interleaved networks. *Bell Labs Technical Journal*, 14(2):97–111, 2009. vi, 13, 15, 34

[85] IEEE standard 802.17. Resilient packet ring (RPR). *IEEE Computer Society*, 2004. 73

[86] IEEE standard 802.3az. Energy Efficient Etherne. *IEEE Computer Society*, 2010. 119

[87] IEEE standart P802.3ba. 40Gb/s and 100Gb/s Ethernet Task Force. *IEEE Computer Society*. 125

[88] Nobuo Tanba, Shinji Nishimura, Hiroshi Ichibangase, Takashiand Mizuochi, and Yoshinori Hibino. R&D on high-speed optical transport system technologies. (high-speed low-power-consumption optical transport technology for Ethernet. In *Ministry of Internal Affairs and Communication (written in Japanese)*. 132, 133

[89] The Network Simulator, ns-2. url: http://www.isi.edu/nsnam/ns/. accessed on: March. 2015. 41

[90] Ahmed Triki, Ramon Aparicio-Pardo, Paulette Gavignet, Esther Le Rouzic, Bernard Arzur, and Annie Gravey. Is it worth adapting sub-wavelength switching control plane to traffic variations? In *Optical Network Design and Modeling, 2014 International Conference on*, pages 186–191. IEEE, 2014. vi, vii, 15, 29, 122

[91] Ahmed Triki, Paulette Gavignet, Bernard Arzur, Esther Le Rouzic, and Annie Gravey. Bandwidth allocation schemes for a lossless optical burst switching. In *Optical Network Design and Modeling (ONDM), 2013 17th International Conference on*, pages 205–210. IEEE, 2013. vi, vii, 15, 16, 34, 35, 39

[92] Ahmed Triki, Paulette Gavignet, Bernard Arzur, Esther Le Rouzic, and Annie Gravey. Efficient control plane for passive optical burst switching network. In *Information Networking (ICOIN), 2013 International Conference on*, pages 535–540. IEEE, 2013. vi, 13, 15, 17, 20, 72

[93] Ahmed Triki, Annie Gravey, and Philippe Gravey. CAPEX and OPEX saving in SDN-compliant sub-wavelength switching solution. In *Photonics in Switching (PS), 2015 International Conference on*, pages 1–3. IEEE, 2015. 84

[94] Bogdan Uščumlić, Isabella Cerutti, Annie Gravey, Philippe Gravey, Dominique Barth, Michel Morvan, and Piero Castoldi. Optimal dimensioning of the WDM unidirectional ECOFRAME optical packet ring. *Photonic network communications*, 22(3):254–265, 2011. 17, 48, 61

[95] Bogdan Uscumlic, Annie Gravey, Isabella Cerutti, Philippe Gravey, and Michel Morvan. Stable optimal design of an optical packet ring with tunable transmitters and fixed receivers. In *Optical Network Design and Modeling (ONDM), 2013 17th International Conference on*, pages 82–87. IEEE, 2013. 72, 84

[96] Bogdan Uscumlic, Annie Gravey, Philippe Gravey, and Isabella Cerutti. Traffic grooming in WDM optical packet rings. In *Teletraffic Congress, 2009. ITC 21 2009. 21st International*, pages 1–8. IEEE, 2009. v, 1

[97] Bogdan Uščumlić, Annie Gravey, Philippe Gravey, Michel Morvan, and Isabella Cerutti. Network planning and traffic engineering in ECOFRAME optical packet ring. In *Telecommunications Forum (TELFOR), 2011 19th*, pages 808–815. IEEE, 2011. 17

[98] Bogdan Uscumlic, Yvan Pointurier, Annie Gravey, Philippe Gravey, and Luay Alahdab. Scheduling aware dimensioning for time-domain wavelength interleaved network. In *Computing, Networking and Communications (ICNC), 2014 International Conference on*, pages 484–490. IEEE, 2014. 34, 35, 72

[99] Bogdan Uscumlic, Lida Sadeghioon, Annie Gravey, and Philippe Gravey. The cost of traffic protection in bidirectional optical packet switching rings. In *Computers and Communications (ISCC), 2013 IEEE Symposium on*, pages 928–933. IEEE, 2013. 73

[100] Luis Velasco, Miroslaw Klinkowski, Marc Ruiz, and Jaume Comellas. Modeling the routing and spectrum allocation problem for flexgrid optical networks. *Photonic Network Communications*, 24(3):177–186, 2012. 14

[101] Yang Wang, Xiaojun Cao, Qian Hu, and Yi Pan. Towards elastic and fine-granular bandwidth allocation in spectrum-sliced optical networks. *Journal of Optical Communications and Networking*, 4(11):906–917, 2012. 14

[102] Qing Wei, Jamal Bazzi, Matthias Lott, and Yvan Pointurier. Multicast in mobile backhaul with optical packet ring. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2013 IEEE 9th International Conference on*, pages 181–188. IEEE, 2013. 97

[103] Indra Widjaja, Iraj Saniee, Randy Giles, and Debasis Mitra. Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network. *Communications Magazine, IEEE*, 41(5):S30–S36, 2003. v, vii, 1, 12, 16, 20, 72, 73

[104] Brian J Wilson, Ned G Stoffel, Jorge L Pastor, Mike J Post, Kevin H Liu, Tsanchi Li, Kenneth A Walsh, John Y Wei, and Yukun Tsai. Multiwavelength optical networking management and control. *Lightwave Technology, Journal of*, 18(12):2038–2057, 2000. 25

[105] Daojun Xue, Yang Qin, and Chee Kheong Siew. A slotted dynamic traffic scheduling algorithm in time-domain wavelength interleaved networks. In *Networks, 2005. Jointly held with the 2005 IEEE 7th Malaysia International Conference on Communication., 2005 13th IEEE International Conference on*, volume 2, pages 6–pp. IEEE, 2005. 16, 34

[106] Shun Yao, Biswanath Mukherjee, and Sudhir Dixit. Advances in photonic packet switching: an overview. *Communications Magazine, IEEE*, 38(2):84–94, 2000. 24

[107] SJ Yoo. Optical packet and burst switching technologies for the future photonic internet. *Lightwave Technology, Journal of*, 24(12):4468–4492, 2006. 8

[108] Hui Zang, Jason P Jue, and Biswanath Mukherjee. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. *Optical Networks Magazine*, 1(1):47–60, 2000. 13

[109] Georgios S Zervas, Joan Triay, Norberto Amaya, Yixuan Qin, Cristina Cervelló-Pastor, and Dimitra Simeonidou. Time Shared Optical Network (TSON): a novel metro architecture for flexible multi-granular services. *Optics express*, 19(26):B509–B514, 2011. 1

APPENDIX A

Application Centric Energy Efficient Ethernet

THIS Appendix proposes an energy saving strategy for application centric networks. A new architecture for Energy Efficient Ethernet (EEE) that meets the Quality of Service (QoS) requirements for different classes of service (CoS) is addressed.

A.1 Introduction

Nowadays, network components are designed and installed to handle peak load and to provide resilience, both of which require over provisioning. This mode of operation results in high energy consumption [17]. Energy saving strategies should be able to adapt link transmissions to the actual traffic patterns.

In Ethernet technology, in the absence of data an auxiliary signal called IDLE is sent. Therefore, the Ethernet Interface Card is active at all times leading to high energy consumption, independent of the traffic load. To reduce energy consumption, the Energy Efficient Ethernet (EEE) standard IEEE 802.3az introduced the concept of Low Power Idle (LPI) mode [86]. The standard proposes the use of "Active" and "Quit" modes to conserve power during periods of inactivity. LPI defines Tw (4.48 μs) as the time required for activating the Interface Card, Ts (2.88 μs) as the time to transition the Interface Card to sleep mode, Tq (39.68 μs) as large periods over when no signal is transmitted, and Tr (1.28 μs) as the time required to transmit a short signal to refresh the Interface Card state to maintain alignment.

The efficiency of EEE strongly depends on the transition times Tw and Ts during which the power consumption is 100%. To reduce the number of transitions, a first in first out (FIFO) queue in the Interface Card can accumulate multiple packets before transmitting them on to the link, a process called *packet coalescing* [15]. The results show that with coalescing, energy efficiency is close to the ideal case (i.e., the case when Tw, Ts, and Tr are equal to zero). However, coalescing increases packet delay, which leads to reduced Quality of Service (QoS) on connections with multiple EEE nodes. In this Appendix we propose an alternative to the EEE coalescing scheme that yields similar energy efficiency but reduces the packet delays to guarantee QoS for various Classes of Service (CoS).

To the best of our knowledge, there is no prior work on Ethernet energy reduction that considers different CoS. Therefore, in this Appendix, a novel application centric scheme with QoS support [65] is proposed for EEE.

A.2 Application centric EEE (ACEEE)

In this design multiple logical queues with different thresholds are used in the coalescing operation to handle the QoS needs of different CoS. Packet coalescing can be based on a maximum packet count (COUNT) or a coalescing timer, t_c, that tracks the time since the arrival of the first packet, or whichever comes first. Fig. A.1 illustrates the ACEEE design [14]. Each CoS has its own distinct COUNT and coalescing timer values.

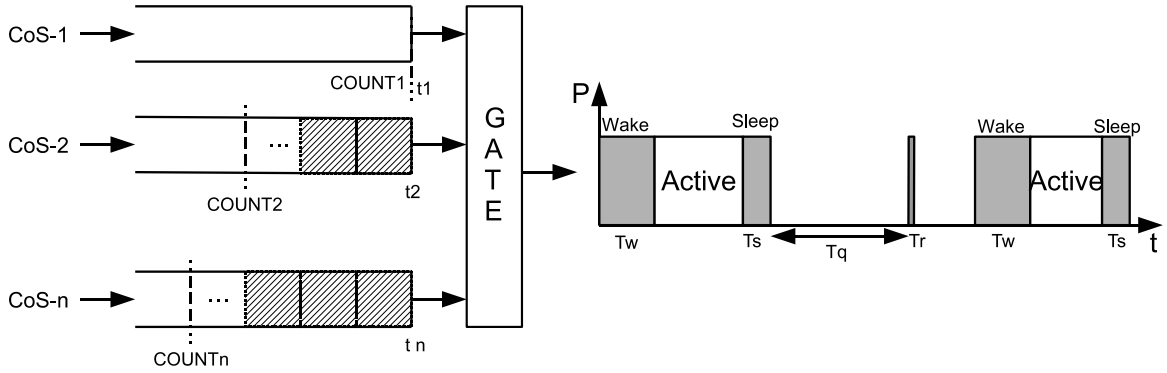


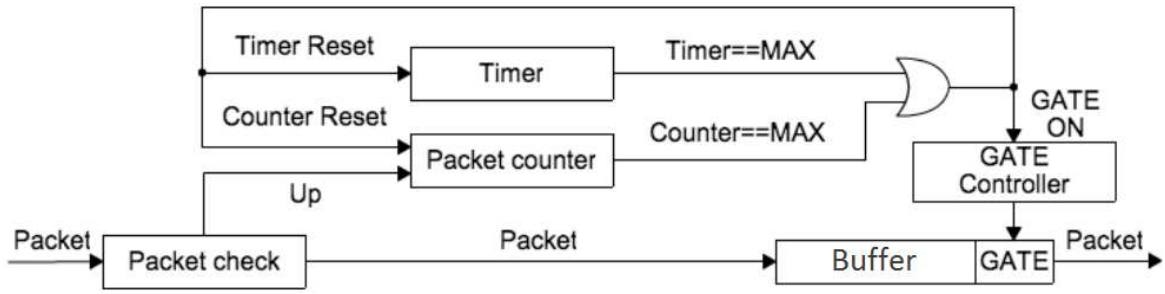
Figure A.1 — On the left, ACEEE; on the right, LPI in EEE [15].

The arrival of the first packet in an empty logical queue resets the COUNT variable and starts the coalescing timer for the CoS queue to which the packet belongs. The GATE is enabled when either the maximum COUNT of any queue is reached or the coalescing timer of any queue expires. As result all packets from all the queues will be transmitted. Furthermore, all packets arriving when the link transmitter is in the active mode will be sent as well.

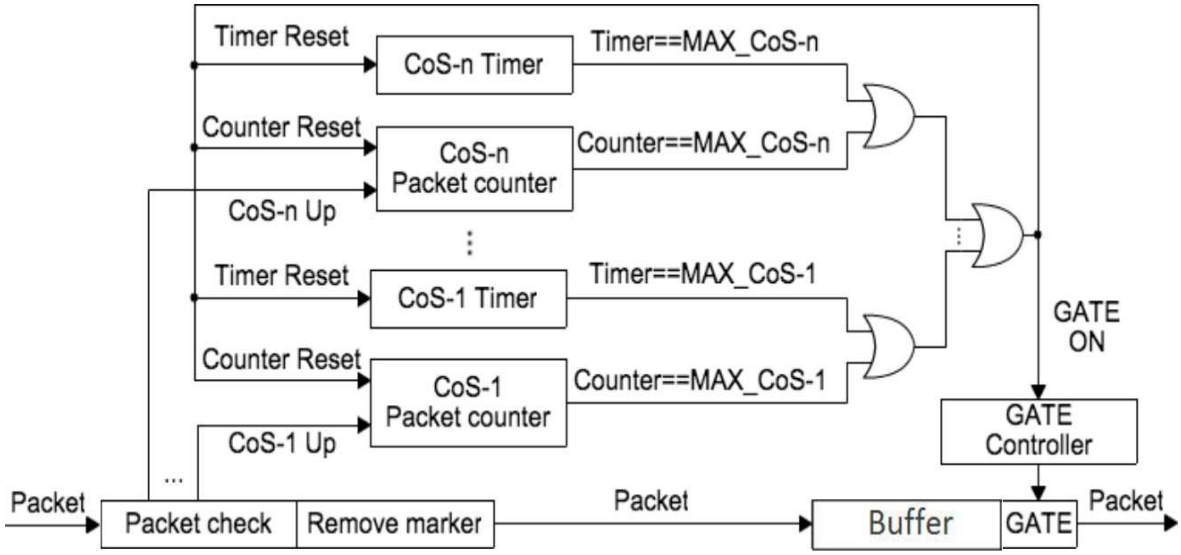
The above procedure reduces the total energy consumption considering multiple CoS by eliminating unnecessary transition times Tw and Ts.

A.3 Hardware implementation

Fig. A.2(a) shows a simplified diagram of the conventional LPI with coalescing. Initially, a coalescing timer and a packet counter are set to 0. An Ethernet frame is passed from the Medium Access Control (MAC) device to a Packet check function. For each received packet, the Packet check function, sends an "Up" signal to the Packet counter and the packet itself to a FIFO Buffer. The Timer is incremented periodically regardless of whether or not packets arrived. If the value of the Packet counter reaches the maximum COUNT value, or if the value of the Timer reaches the maximum Timer value, then the GATE ON signal is sent to a GATE Controller. The latter enables the GATE in front of the Buffer, causing all collected packets to be transmitted. The GATE ON signal also serves as a Counter Reset signal and a Timer Reset signal to reset the Packet counter and the Timer, respectively. When all packets in the Buffer



(a) Conventional LPI with coalescing



(b) ACEEE

Figure A.2 — Hardware implementation of an Ethernet Interface Card.

are fully transmitted the state of the GATE in front of the Buffer is changed to the OFF mode.

The ACEEE design requires upper layers support for packet CoS identification. Packet classification into different CoS classes can be based on the Type of Service (ToS) field in the IP header, or other fields such as TCP/UDP (Transmission Control Protocol/User Datagram Protocol) port numbers. Since the ACEEE design is only for the Ethernet Interface Card, we assume that the higher layers that perform packet classification pass down the Ethernet frame with an additional marker. This marker is used in the Packet check function shown in Fig. A.2(b), and then removed.

Fig. A.2(b) shows a simplified diagram of the proposed ACEEE design, which differs from the conventional design as follows. The ACEEE Packet check function reads the marker indicating the CoS type of each incoming packet. A Packet counter and a Timer are maintained for each CoS. For each received packet, depending on the marker, the Packet check function, sends an "Up" signal to the corresponding Packet counter and the packet itself to the FIFO Buffer (e.g., if the Packet check function based on the marker identifies a CoS-1 packet, then the CoS-1 counter is incremented and the packet itself is sent to the Buffer). CoS Timers are incremented periodically

regardless of whether or not packets arrive. If the value of any Packet counter reaches the maximum COUNT value, or if the value of any Timer reaches the maximum Timer value, then the GATE ON signal is sent to the GATE Controller. The latter enables the GATE in front of the Buffer, causing all collected packets to be transmitted. As in the conventional-LPI-with-coalescing design, the GATE ON signal also serves as a Counter Reset signal and a Timer Reset signal for all CoS Packet counters and Timers, respectively. Therefore, an ACEEE Interface Card can be implemented with minor changes to the conventional LPI with coalescing Interface Card.

A.4 Numerical results

Our simulations use real traffic traces obtained from the Center for Applied Internet Data Analysis (CAIDA) [10]. Ten files of one-minute traces collected by a network monitor named "Equinix-Chicago" from 13.00 to 13.10 on 6th July 2014, and ten-files of one minute traces collected by a network monitor named "Equinix-SanJose" from 13.00 to 13.10 on 6th July 2014 are used. Both monitors are installed on 10 Gb/s links. From these we compute the average packet duration, yielding an average link utilization of 15.4% and 37.4% for the Chicago and San Jose links, respectively.

In the ACEEE simulation, three classes of service [90], with the parameters described in Tab. A.1, are defined:

- CoS 1: real time and interactive traffic, very sensitive to packet delay.
- CoS 2: streaming and bulk data traffic, less sensitive to packet delay.
- CoS 3: best effort traffic, P2P, download.

Based on the traffic statistics collected by the Chicago passive network monitor [11] for one month period between 15th February 2015 to 15th March 2015, we computed the average CoS distribution. In our simulation the marker is randomly generated for each packet with this distribution to be 2.4%, 84.2%, and 13.4%, for CoS-1, CoS-2, and CoS- 3, respectively. Since the CoS are assigned randomly, we simulate the packets as being all the same length (equal to the average for the data). CoS-1 packets are not subject to coalescing because the packets of this class need to be served as they arrive. On the other hand CoS-2 and CoS-3 packets are subject to coalescing and hence incur additional delays.

Table A.1 — Candidate parameters for ACEEE.

CoS	COUNT	Timer	Applications	Max Delay
1	1 packets	0 μs	Games, Chat,	3 ms
2	10 packets	12 μs	News, E-mail, Streaming,	5 ms
3	100 packets	120 μs	P2P, Download,	10 ms

Fig. A.3 summarizes the simulation results for different link utilization versus power used. Five different Ethernet technologies are considered:

- NO EEE: Ethernet with constant power use, which would be 100% and is independent of link utilization.
- EEE: Ethernet with LPI support but without packet coalescing.
- Coalescing-1 (10pkt/12μs): Ethernet with coalescing, and without CoS support. COUNT is equal to 10 packets and coalescing timer is 12 μs.
- Coalescing-2 (100pkt/120μs): Ethernet with coalescing but without CoS support. COUNT is equal to 100 packets and coalescing timer is 120 μs.
- ACEEE with the three CoS defined above.

There is an average gain of 39.5% and 26.5% in energy use for ACEEE compared with NO EEE and EEE respectively. The energy efficiency of ACEEE is similar to Coalescing-1 (10pkt/12μs) but it less than Coalescing-2 (100pkt/120μs) by about 20%, due to the higher number of the Tw and Ts states triggered by CoS-1.

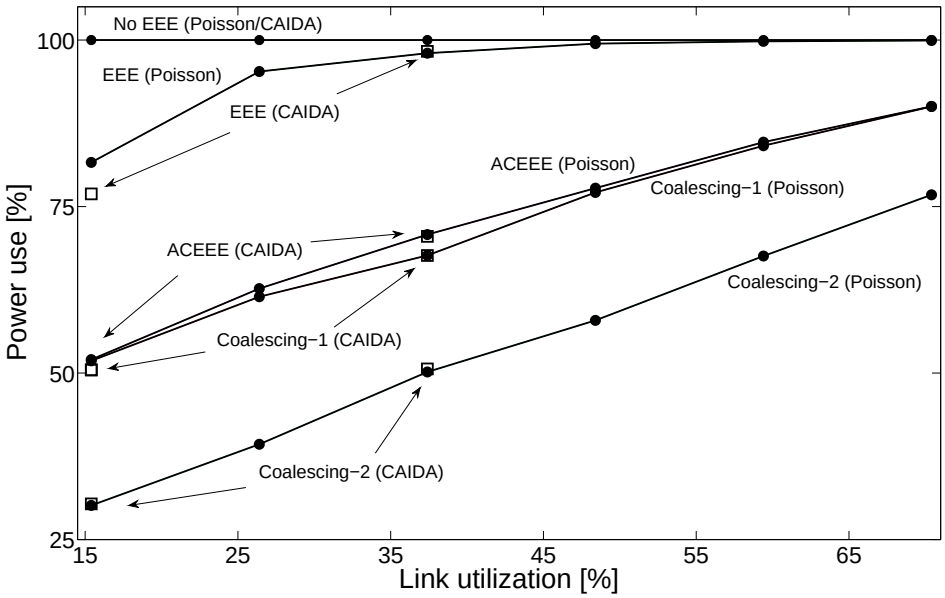


Figure A.3 — Energy consumption for different traffic types.

In order to determine if our results generalize, in Fig. A.3, we compare the Ethernet technologies in case of real CAIDA traffic and Poisson-based traffic with the same averaged packet duration and link utilization. We observe that the Ethernet technologies studied have the same energy use performance for both traffic types.

Tab. A.2 illustrates the packet delay versus link utilization, where the ACEEE packet delay for the delay sensitive class, CoS-1, is less compared to legacy Ethernet technologies. CoS-2 has similar performance compared to Coalescing-1. Out of the three classes, CoS-3 has the highest packet delay; this class is less sensitive to the packet delay. Note that, all CoS types have better performance compared to Coalescing-2, with a 20% increase in energy use.

Table A.2 — Packet delay for different traffic types.

Ethernet technologies	Packet delay [μs]			
	CAIDA traffic for different link utilization		Poisson-based traffic for different link utilization	
	15.4 [%]	37.4 [%]	15.4 [%]	37.4 [%]
NO EEE	0.72	0.85	0.82	0.86
EEE	4.51	4.36	4.61	4.34
Coalescing-1	10.32	8.71	10.65	8.81
Coalescing-2	66.53	62.79	70.3	67.7
ACEEE (CoS-1)	3.98	3.01	4.16	2.98
ACEEE (CoS-2)	10.02	8.26	10.32	8.22
ACEEE (CoS-3)	13.59	13.77	13.18	13.82

Finally, the ACEEE has better packet delay performance for the delay sensitive CoS (i.e., CoS-1 and CoS-2) compared to the legacy Ethernet technologies with the same parameters and without CoS support. Therefore, in case of a connection with multiple EEE nodes, the proposed ACEEE will meet the QoS requirements.

A.5 Conclusion

This Appendix proposed a new architecture for Energy Efficient Ethernet that supports multiple CoS. The ACEEE packet delay for the delay sensitive CoS was drastically reduced compared to the legacy Ethernet technologies. Simulation results showed that the proposed ACEEE has an average gain of 39.5% and 26.5% in energy use compared to NO EEE and EEE, respectively, for real CAIDA traffic and Poison-based traffic. In addition, ACEEE can be realized with insignificant hardware (and/or firmware) modification.

APPENDIX B

High Speed 100GE ALR Switching for Energy Reduction

THE energy efficiency is becoming a critical issue in high-speed networks. The 100 Gb/s Ethernet was standardized in 2010, and various papers have been published on the power consumption reduction methods using Adaptive Link Rate (ALR) technology. ALR reduces power consumption by adapting link speed to traffic demands. However, the rate switching time is a parameter that impacts the level of power reduction achievable with current ALR technologies. In this Appendix, we propose a new energy efficient architecture for 100 Gb/s Ethernet. The rate switching time is reduced using coherent lightwave transmission technologies. Based on the simulation results, we evaluate how the proposed architecture impacts energy efficiency, switching time and network performance.

B.1 Introduction

Due to the spread of devices connected to the Internet, the traffic demand increases dramatically year-by-year. To handle the explosive growth in network traffic demand, the 100 Gb/s Ethernet (100GE) standard was specified in 2010 [87]. While the transmission speed in Ethernet increased from 10 Mb/s to 100 Gb/s, the energy consumption increased correspondingly. Nowadays, the energy consumption has become a critical issue. With the evolution of optical transport technologies, the energy efficient solutions for the transport segment of the network have become available. However, network components are still designed and installed to handle peak load and link failures. Moreover, it is challenging to predict the required transport capacity [23]. Thus, typically the networks are over-provisioned, which results in higher energy consumption than required. For instance, the link utilization is only 30%-40% [17], in the core network . Hence, energy saving strategies, in which link rate is adapted to actual traffic variation, are under active research-and-development. There are several studies on power consumption reduction in 100GE [73], however, in these methods, the link rate switching time has a detrimental effect on packet delay.

In this Appendix, we propose an energy efficient scheme for 100GE with Adaptive Link Rate (ALR) technology [52], by reducing the rate switching time using coherent lightwave technologies (CLT).

The remainder of this Appendix is organized as follows. Section B.2 describes the legacy 100GE technology, with details on the architecture and the energy-reduction schemes. The Section B.3 introduces the proposed scheme, 100GE-ALR-CLT. The simulation results for the performance evaluation of the 100GE-ALR-CLT scheme are presented in Section B.4, and finally, Section B.5 concludes the Appendix.

B.2 Legacy 100GE technology

In this section, the architecture of a 100GE module is described, followed by an energy-reduction scheme for the legacy 100GE modules.

B.2.1 The architecture of a 100GE module

A general overview of the 100GE architecture is shown in Fig. B.1. The 100 Gb/s Physical Layer (PHY) implementation is generally referred to as 100GBASE-R [38]. The media-independent interface (MII) provides the logical interconnection between the medium access control (MAC) and PHY sublayers. The principles of operation, at the sending side, of the physical-coding sublayers (PCS), the physical medium attachment (PMA), and the physical-medium-dependent (PMD) sublayers, are described next.

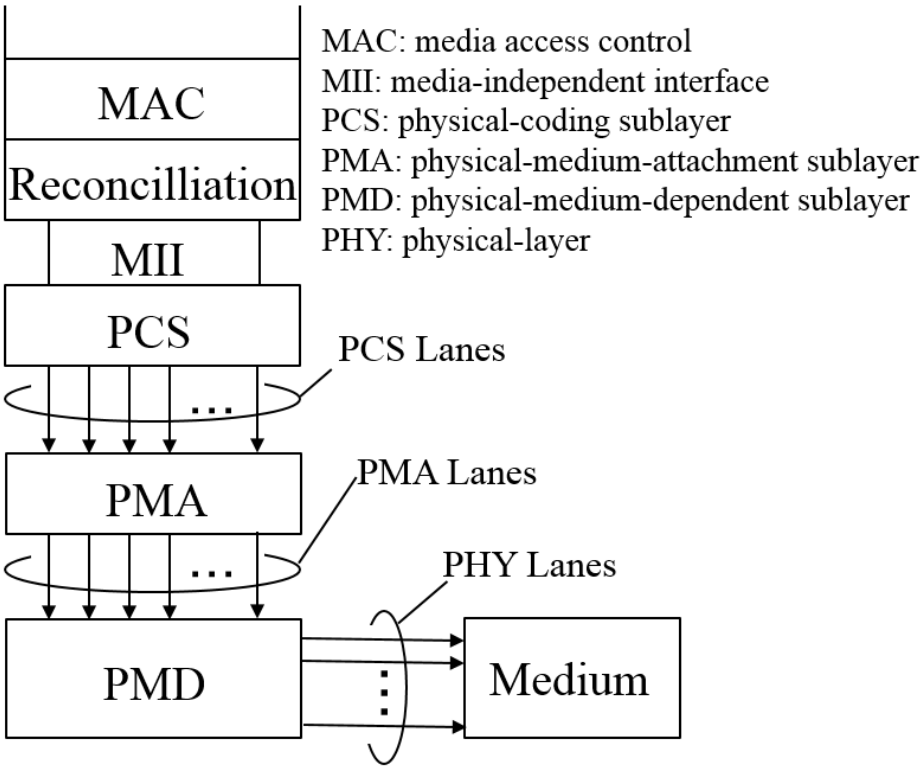


Figure B.1 — Generalized 100GE module architecture.

The PCS handles data sent by the MAC layer by encoding eight data octets into 66-bit 64-B/66-B blocks. These blocks are then distributed to the PMA components in round-robin fashion [20].

The PMA layer allows the PCS to be interconnected in a media independent way with a range of PHYs. The PMA has a flexible implementation, which allows different configurations of the PMD.

Finally, the PMD maps the incoming PMA data streams onto a matching number of PHY lanes.

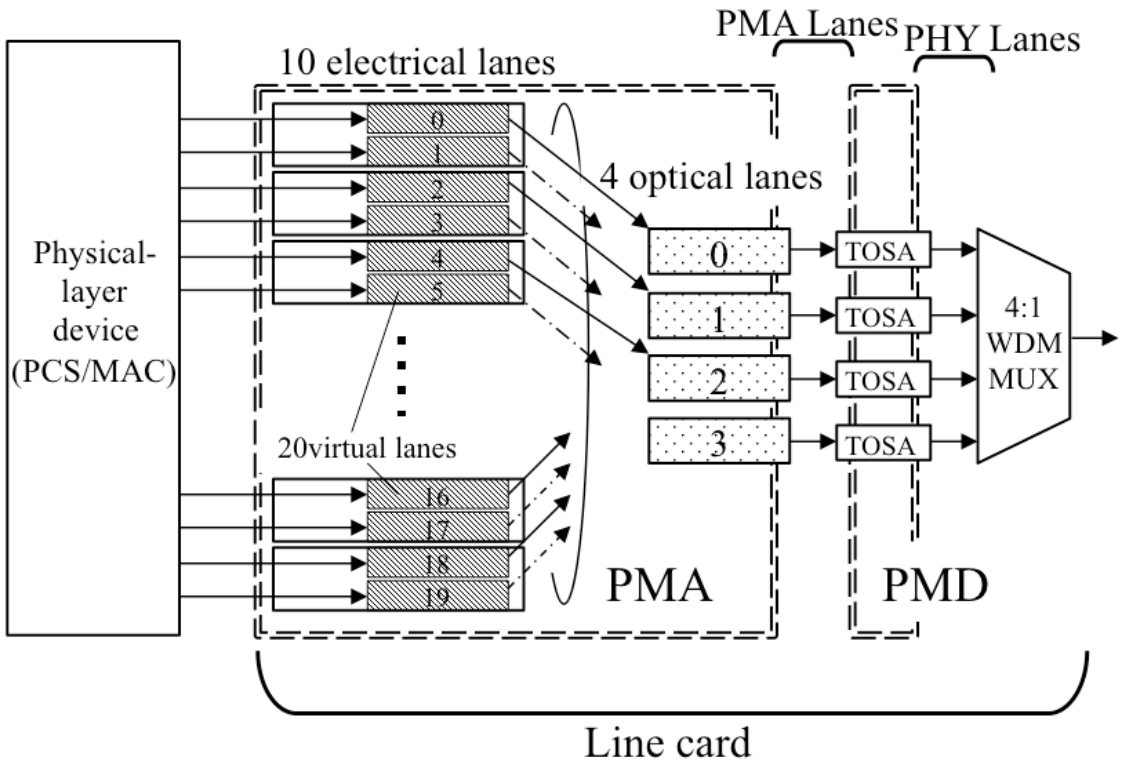


Figure B.2 — Detailed structure of the transmitter side in 100GE module.

Fig. B.2 shows the detailed structure of a 100GE module. The packets are sent from PCS/MAC to 20 virtual lanes. In return, these virtual lanes send data to 4 optical lanes which correspond to their electrical lanes. Finally, the 4 optical lanes generate optical signals from 4 optical modules, which are named Transmitter Optical SubAssembly (TOSA) in Fig. B.2.

In the following, the relation between Fig. B.2 and Fig. B.1, is described:

- The 20 virtual lanes in Fig. B.2 correspond to the PCS lanes in Fig. B.1.
- The 25G electrical interfaces between the 25Gx4 optical lanes and TOSAs in Fig. B.2 correspond to the PMA lanes shown in Fig. B.1.
- The TOSA modules, and 4:1 WDM MUX in Fig. B.2 correspond to the PMD in Fig. B.1.
- The 25G optical interfaces between TOSAs and 4:1 WDM MUX in Fig. B.2 correspond to PHY Lanes in Fig. B.1.

Fig. B.3 shows the packet flow from MLD to optical lanes. At the transmitter side, the MLD receives 8-octet packets from the PCS through 5 Gb/s interfaces and transfers

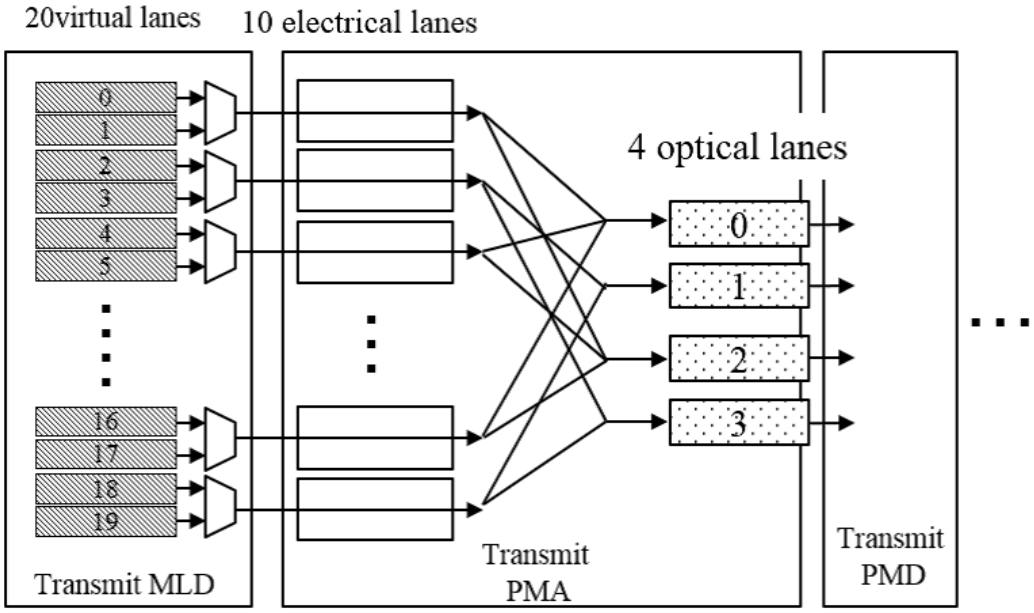


Figure B.3 — Packet flow from MLD to optical lanes.

these packets to the physical PMA. Between the MLD and PMA, 2-to-l multiplexers (MUXes) multiplex 2 virtual lanes (VLs) each to a 10 Gb/s interface. Between the PMA and PMD, the distribution of packets is done in round robin manner. The transmitter PMD and receiver PMD are connected by 25 Gb/s external interfaces (optical lanes). Using this approach, the 100GBASE-4 module can achieve a total of 100 Gb/s capacity between the transmitter and receiver line cards. In this architecture, the usage of 10 Gb/s interfaces and 25 Gb/s interfaces is determined by the distributed pattern of 5 Gb/s VLs (virtual lanes) [32], [47].

B.2.2 Energy reduction technologies in 100GE

The energy reduction in 100GE [73], [47] using ALR technology, is under active research. The power consumption is reduced in ALR technologies by decreasing link speed. For instance, in [47], the authors propose a method for efficient management of the PMA/PMD by using sleep/active states. In this method, not only the optical lanes can be turned off but also electrical lanes, too. Tab. B.1 lists the multiple VLCs (Virtual Lane Classes) and their corresponding virtual lanes number. For example, virtual lanes 0, 5, 8, 13, 16, correspond to VLC0, and are sent to the 0^{th} optical lane. Data bits from each VLC are sent to different optical lanes. When an optical lane is turned off, the corresponding electrical lanes, as specified in Tab. B.1, are turned off.

In [73], the energy reduction is realized by selectively activating some electrical lanes. However, turning off each lane changes the manner in which data are sent to the electrical lanes. Furthermore, the time to turn-off/turn-on the optical module is 2 ms. To summarize, in conventional 100GE modules, the electrical lanes, with their corresponding optical lanes, and TOSA module, is turned off simultaneously.

Table B.1 — VLC and its elements.

Virtual Lane Class	Virtual Lane Numbers	Optical Lane Number
VLC0	0, 5, 8, 13, 16	0
VLC1	2, 7, 10, 15, 18	1
VLC2	1, 4, 9, 12, 17	2
VLC3	3, 6, 11, 14, 19	3

B.3 Proposed method for high-speed link rate switching

In this section, a new method is proposed using 100GE adaptive link rate with coherent lightwave transmission technologies (100GE-ALR-CLT). Several mechanisms are proposed for link rate switching time reduction and for setting the link rate switching threshold policy in 100GE-ALR-CLT.

B.3.1 Switching time reduction

Fig. B.4 shows the architecture of 100GE-ALR-CLT. In legacy 100GE-ALR, each optical lane sends an optical signal from a TOSA module. Therefore, the link rate switching time to turn on/off an optical module is 2 ms. On the other hand, in 100GE-ALR-CLT, each optical lane sends data to a digital signal pocessor (DSP). Two optical modulators generate optical signals from a single laser diode (LD) using a dual polarization-quadrature phase shift keying (DP-QPSK) coding method.

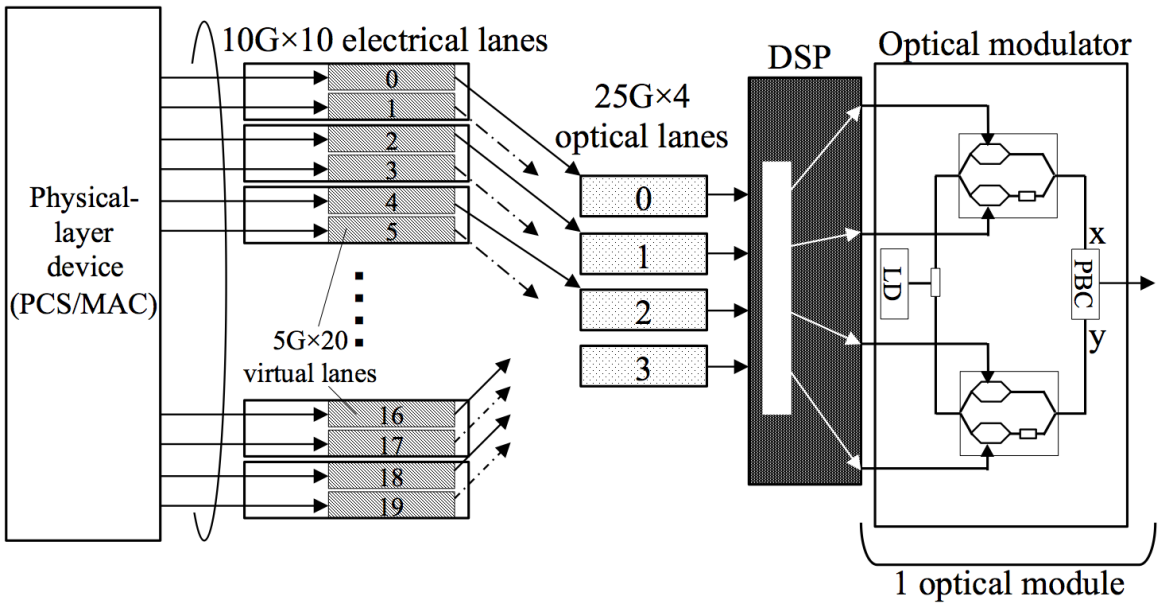


Figure B.4 — Optical module for 100GE with CLT technology.

The proposed method does not turn off an optical module; instead by using coherent lightwave transmission technology, one or more optical and electrical lanes are turned off as needed. To realize 100GE-ALR-CLT, we should consider the additional computation time and power consumption needed in the DSP.

However, the symbol switching and the modification of the subcarriers do not take longer than 1 μs. Consequently, the switching to a different link speed is realized in the worst case within 100 μs [73]. Therefore, the link rate switching time in 100GE-ALR-CLT is assumed to be 100 μs.

B.3.2 Dummy signal insertion in 100GE-ALR-CLT

To realize 100GE-ALR-CLT, the number of bits per symbol is decided according to the target link rate, for instance, 4 bits for 100 Gb/s and 3 bits for 75 Gb/s. However, in order to decrease the link rate, the transmission modulation format is changed from DP-QPSK to 8PSK, the optical signal is generated by one laser diode. Hence, the power consumption does not decrease. Furthermore, to change the transmission code, a negotiation between sender and receiver is needed, which adds additional rate switching delay.

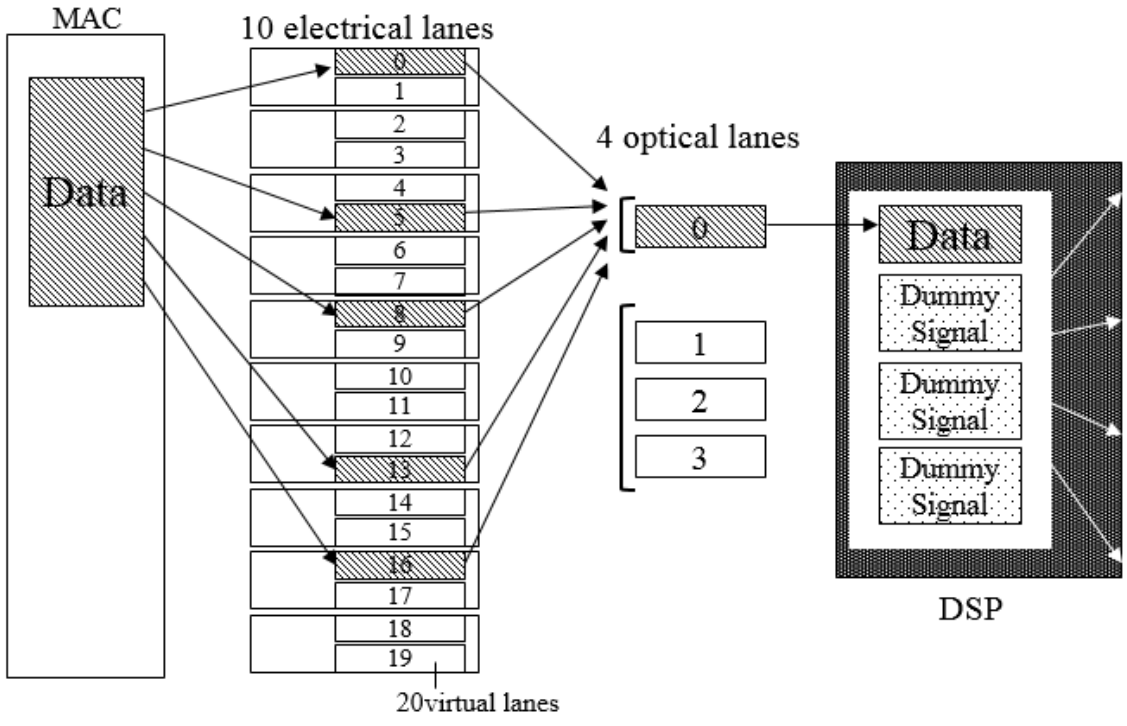


Figure B.5 — Example of the dummy signal insertion in 100GE-ALR-CLT.

To realize high-speed rate switching, a *Dummy* signal insertion is proposed. When turning off optical/electrical lanes at one of the DSP input ports, the DSP inserts dummy signals. The transmission code is kept, i.e., DP-QPSK. At receiver side, the DSP identifies the dummy signals. Using this technique, the high-speed switching is

realized, as result of two main reasons. Firstly, the time for amplifier tuning (to change gain control or output power) depends on the number of simultaneously lit wavelengths, and therefore, if a dummy signal is sent instead of a data signal under low loads, then the time to switch to a data signal as load increases/decreases is kept small. Secondly, the rate at which optical signal is sent stays unchanged, as result of the use of the dummy signal into a data signal, and thus, resynchronization is avoided.

Fig. B.5 shows the example of the dummy signals insertion in 100GE-ALR-CLT, three optical lanes and the corresponding electrical lanes are turned off. Firstly, the MAC decides how many optical lanes should be used, then, MAC sends data to electrical lanes 0, 5, 8, 13, and 16 since these electrical lanes correspond to optical lane 0. Then, the optical lane 0 sends the data signal to the DSP. In the meantime, the optical lanes 1, 2, 3 and their corresponding electrical lanes are turned off. Therefore, the DSP gets a data signal from only one optical lane. Accordingly, the DSP inserts dummy signals in place of data signals from optical lanes 1, 2, 3 since these optical lanes are turned off. The data and dummy signals are sent by the optical modulator, and the receiver-side DSP detects dummy signals. The dummy signal has a pre-determined pattern, which allows the receiver-side DSP to differentiate dummy signals from data signals.

B.3.3 Format of a dummy signal

As described in the previous section, the dummy signals are generated by DSP. The format of dummy signal was proposed in our prior work [58]. There are three requirements for the format of dummy signal. Firstly, it follows a fixed pattern in order to keep the DSP circuit design simple. Second, in order to achieve fast synchronization between the transmitter and the receiver, the dummy signal should be detected at high speeds. Finally, to avoid voltage imbalance problems between connected systems or components, the dummy signal has to be a DC-balanced signal.

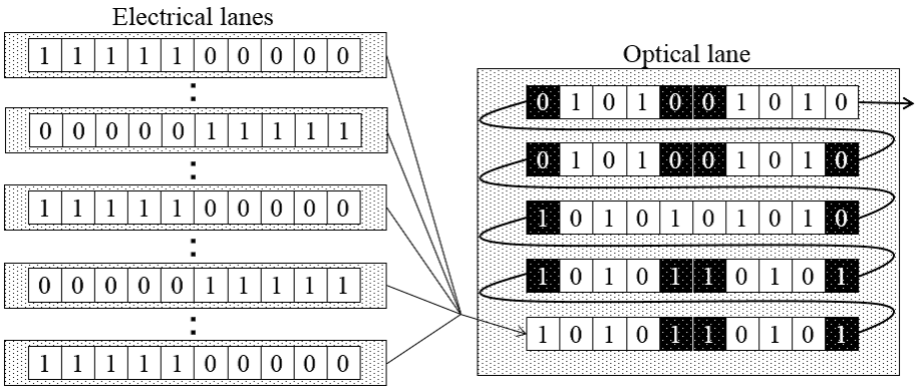


Figure B.6 — The format of dummy signal.

Fig. B.6 shows the format of a dummy signal. The optical lanes and corresponding 5 electrical lanes are not involved in the generation of the dummy signal (in fact, the optical and electrical lanes are turned off when there is no data signal and the dummy

signal is created by the DSP). However, the receiver-side DSP treats the dummy signal as if it was sent by the 5 electrical lanes and corresponding optical lanes. In Fig. B.6, the dummy signal format is shown.

In each virtual lane, the dummy signal is represented by "0000011111," and this signal is sent to an electrical lane. Data signals are scrambled to ensure that the pattern of the dummy signal does not appear within data signals. This allows the DSP of the receiver side to detect dummy signals very fast. To maintain DC-balance, each of the logical 0 and 1 bits is allowed to appear continuously for utmost five bits.

B.3.4 Threshold policy in 100GE-ALR-CLT

In 100GE-ALR-CLT, in order to change the link rate, a threshold policy is required [31]. Fig. B.7 illustrates the packet queue, if the buffer occupancy level is equal to/exceeds the threshold value α, then the link rate is increased. To switch to a lower link rate, a dual threshold policy is used for each bitrate. However, in this Appendix, for the sake of simplicity, three thresholds are used to select between 25, 50, 75, and 100 Gb/s.

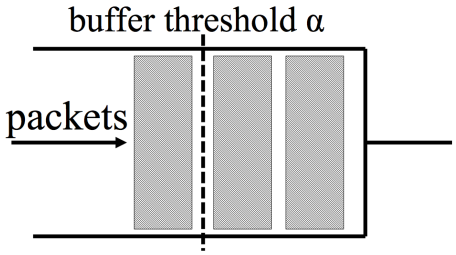


Figure B.7 — Threshold policy in ALR.

B.4 Performance evaluation

In this section, the performance evaluation methods and the obtained results are described. Firstly, the power consumption and packet delay for different thresholds of link rate change are presented. Next, energy consumption and packet delay under appropriate link rate change thresholds, selected from the first step, are characterized.

In the evaluation, CAIDA traces [9] are used. However, CAIDA traces provide only 10GE data. In our study, these 10GE traces are scaled to 100GE by dividing packet inter-arrival time by 10. One CAIDA trace file contains unsampled packet data for 1 minute of traffic. After rescaling, the data corresponds to a 6-second trace. In the simulation, we assumed that every packet was 1500 bytes, therefore, the link utilization is higher than in the original CAIDA trace.

In [88], the authors show the power consumption levels of each components of 100GE-ALR. The data per component power consumption is shown in Tab. B.2. SerDes circuit converts data between serial and parallel interface in each direction, and it

Table B.2 — Power consumption in 100GE.

Link speed	PCS	SerDes circuit	TOSA & ROSA
25 Gb/s	8.9 W	0.2 W	3.0 W
50 Gb/s	17.6 W	0.4 W	6.0 W
75 Gb/s	26.3 W	0.6 W	9.0 W
100 Gb/s	35 W	0.8 W	12.0 W

corresponds to PMA module shown in Fig. B.2. The power consumption in a DSP is 20 W, as reported by a manufactured device [57], which was released by a Japanese optical module vendor. The calculated energy consumption for each link speed using the conventional and proposed methods, are shown in Tab. B.3 [73], [88]. The power consumption of TOSA and ROSA, in conventional scheme, is replaced by the power consumption of DSP, in proposed scheme.

Tab. B.3 shows that the proposed scheme requires more energy than conventional scheme, because the DSP and optical modulator consumes more power in the 100GE-ALR-CLT design.

Table B.3 — Power consumption for each link speed.

Link speed	Conventional scheme	Proposed scheme
25 Gb/s	12.1 W	29.1 W
50 Gb/s	24.0 W	38 W
75 Gb/s	35.9 W	46.9 W
100 Gb/s	47.8 W	55.8 W
Switching time	2 ms	100 μs

Tab. B.3 also shows that the proposed scheme requires less switching time than conventional scheme. The reason why proposed scheme achieves high speed switching time is shown in previous section. Besides, the proposed scheme does not require wavelength division-multiplexing configuration, so it requires only 100 μs [73].

In the simulations presented next, we used the following threshold values: α, 2α, and 3α, to switch the link rate between 25, 50, 75, and 100 Gb/s. An extended set of results for different values of α, is proposed.

B.4.1 Impact of threshold policy on network performance

Figs. B.8 and B.9 show the power consumption and packet delays under different threshold settings. The value of link utilization in this traffic data is 38%. This value is representative, since average link utilization in core-network links is between 30%-40% [17].

The horizontal axis of two graphs shows the value of α. In Fig. B.8, for the value of α between 10 and 10000 packets, 100GE-ALR-CLT has lower power consumption than

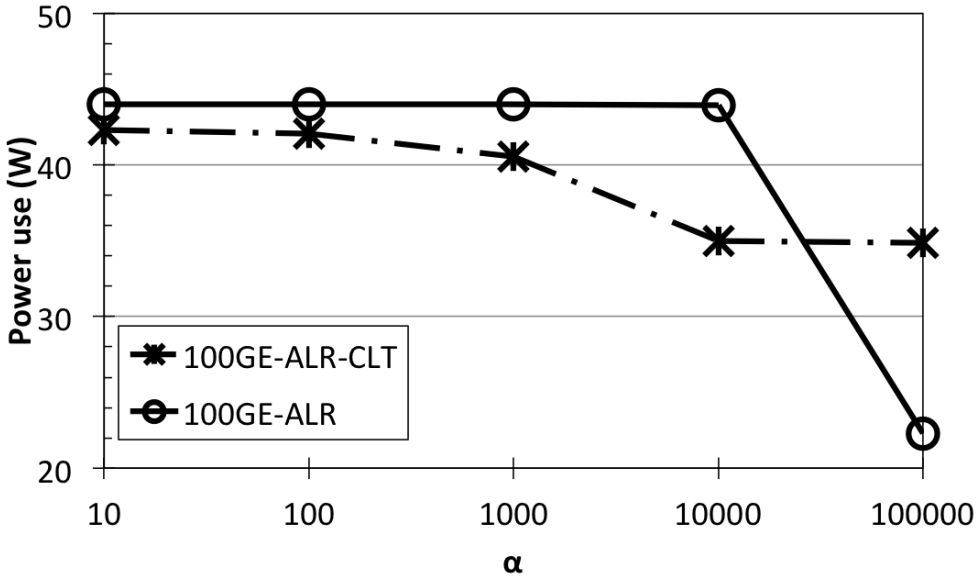


Figure B.8 — Power consumption under different threshold policies.

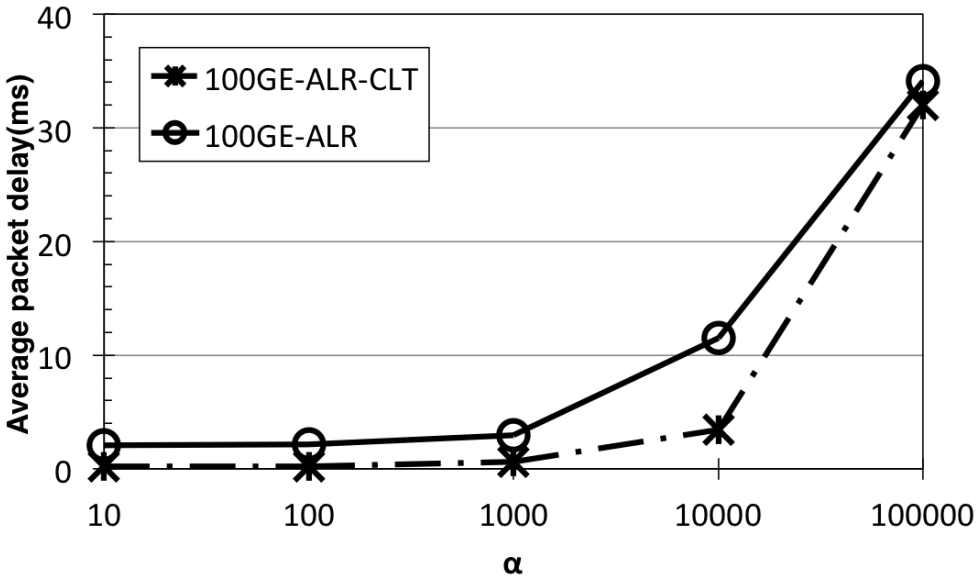


Figure B.9 — Average packet delay under different threshold policies.

100GE-ALR. However, when α is 100000 packets, 100GE-ALR-CLT has high power consumption. At this value of α almost all packets are sent at 25 Gb/s, which causes the power usage to be lower comparing to 100GE-ALR-CLT. On the other hand, when the value of α is 10, nearly all packets are sent at 100 Gb/s.

Fig. B.9 shows the packet delay in 100GE-ALR-CLT compared to 100GE-ALR. The packet delay of 100GE-ALR is higher even though, both modules have almost the same power consumption.

The two figures reveal that 100GE-ALR-CLT has a higher power efficiency then 100GE-ALR, considering the value of α equal to 10000 packets. Therefore, this value of α is used in following subsections.

B.4.2 Impact of link utilization on energy savings

Fig. B.10 shows the power consumption of four schemes: 100GE-ALR-CLT, 100GE-ALR (which is the conventional method), 100GE-CLT (with only 100 Gb/s link rate in 100GE-ALR-CLT), and 100GE (with only 100 Gb/s link rate in 100GE-ALR). It shows that for low link utilization, 100GE-ALR-CLT consumes less power than 100GE-ALR. There are two reasons why 100GE-ALR consumes more power: (i) 100GE-ALR consumes more energy during rate switching intervals, (ii) the switching intervals are longer than in our proposed scheme.

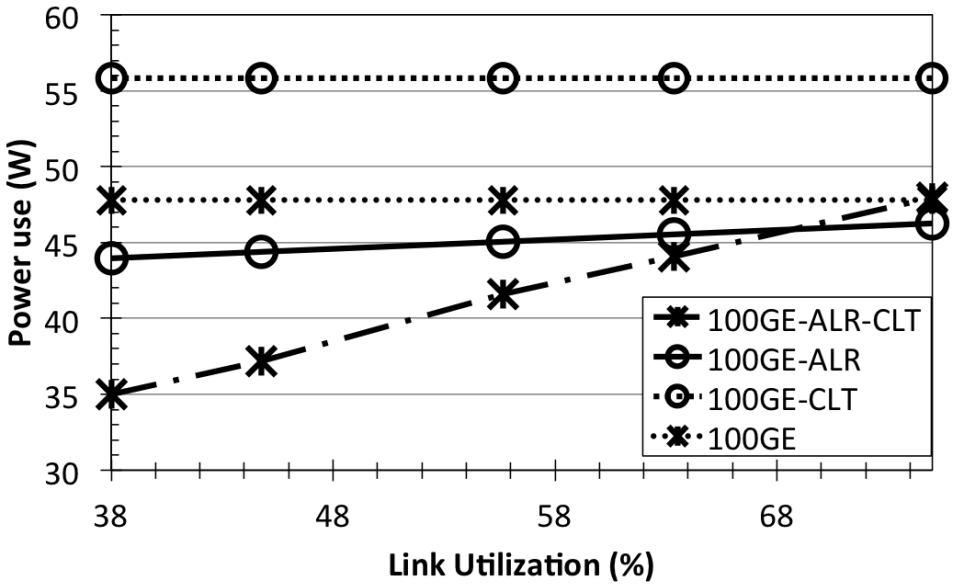


Figure B.10 — Power consumption in conventional and proposed schemes.

On the other hand, for high link utilization, 100GE-ALR-CLT consumes more power, but achieves higher energy efficiency comparing to 100GE-ALR. When link utilization is between 30%-40% [17], which represents the average link utilization in core networks [17], the 100GE-ALR-CLT outperforms the legacy 100GE-ALR. For a high link utilization, most of the time, both 100GE-ALR-CLT and 100GE-ALR transmit data at the 100 Gb/s link rate.

B.4.3 Impact of link utilization on packet delay

Fig. B.11 shows the average packet delay for the two schemes. The proposed scheme has lower packet delay compared to the conventional one. For 100GE-ALR module, the packet delay decreases as the link utilization increases.

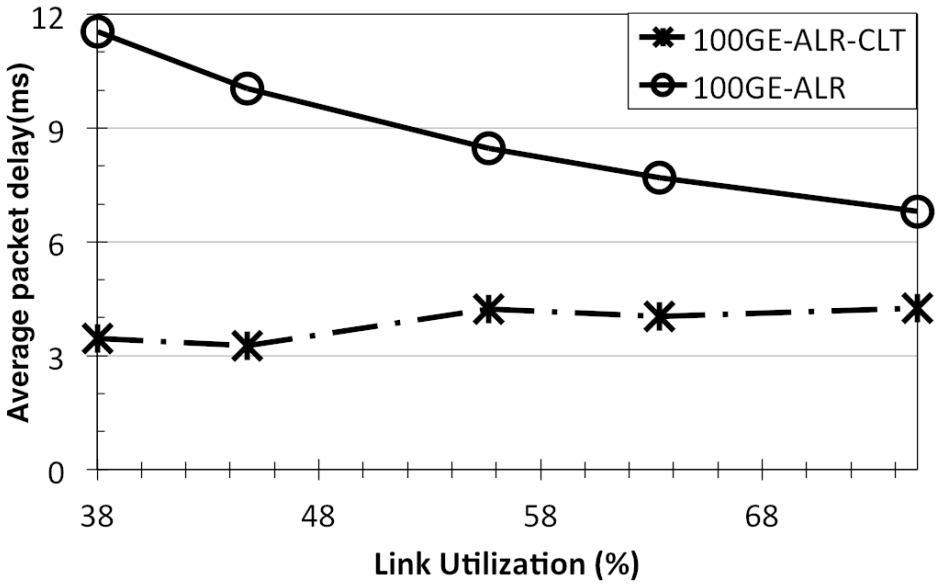


Figure B.11 — Average packet delay in conventional and proposed schemes.

The 100GE-ALR module uses a higher link rate compared to 100GE-ALR-CLT, because the 100GE-ALR cannot change its link rate rapidly, given its longer switching time. It also implies that the 100GE-ALR incurs shorter transmission delays, because it uses higher link rates. However, Fig. B.11 shows that the 100GE-ALR-CLT experiences lower packet delay for the considered link utilization.

B.5 Conclusion

In this Appendix, we proposed a new energy efficient architecture for 100 Gb/s Ethernet. The bitrate switching time was reduced using coherent lightwave transmission technologies. Our simulation results show that the proposed 100GE-ALR-CLT achieves 20.4% power reduction, and a 61.6% packet-delay reduction, for the link utilization between 30%-40% (which is the typical range for core-network links).

APPENDIX C

List of Personal Publications

C.1 International Peer-Reviewed Journals

EL2015: **Ion Popescu**, Takahiro Miyazaki, Mirai Chino, Xiaoyu Wang, Satoru Okamoto, Annie Gravey, Philippe Gravey, Malathi Veeraraghavan, Maite Brandt-Pearce and Naoaki Yamanaka, "Application-centric energy-efficient Ethernet with quality of service support", Electronics Letters, vol. 51, no. 15, p. 1165-1167, July 23, 2015.

JOCN2013: **Ion Popescu**, Isabella Cerutti, Nicola Sambo and Piero Castoldi, "On the optimal design of a spectrum-switched optical network with multiple modulation formats and rates", Journal of Optical Communications and Networking, vol. 5, no. 11, p. 1275-1284, October 31, 2013.

C.2 International Peer-Reviewed Conferences

GLOBECOM2015: Xiaoyu Wang, Malathi Veeraraghavan, Maite Brandt-Pearce, Takahiro Miyazaki, Naoaki Yamanaka, Satoru Okamoto and **Ion Popescu**, "A dynamic network design for high-speed enterprise access links", Global Communications Conference, GLOBECOM, 2015, San Diego, USA.

NOC2015: **Ion Popescu**, Bogdan Uscumlic, Ahmed Triki, Yvan Pointurier, Annie Gravey and Philippe Gravey, "Scalable solution for resource allocation and RWA in time-domain wavelength interleaved networks", 20th European Conference on Network and Optical Communications, NOC, 2015, London, United Kingdom.

ONDM2015: Takahiro Miyazaki, **Ion Popescu**, Mirai Chino, Xiaoyu Wang, Kunitaka Ashizawa, Satoru Okamoto, Malathi Veeraraghavan and Naoaki Yamanaka, "High speed 100GE adaptive link rate switching for energy consumption reduction", 19th International Conference on Optical Network Design and Modeling, ONDM, 2015, Pisa, Italy.

IEICE2015: Mirai Chino, **Ion Popescu**, Takahiro Miyazaki, Yoshihiro Isaji, Shoichi Nagata, Kunisada Ashizawa, Satoru Okamoto and Naoaki Yamanaka, "Application centric energy efficient Ethernet for minimum delay", *written in Japanese*, IEICE Technical Committee on Communication System, 2015, Fukuoka, Japan.

iPOP2015: Takahiro Miyazaki, **Ion Popescu**, Mirai Chino, Yoshihiro Isaji, Kunitaka Ashizawa, Xiaoyu Wang, Satoru Okamoto, Malathi Veeraghavan, and Naoaki Yamanaka, "Application-centric energy reduction scheme with dynamic lane assignment scheme in 100GE", 11th International Conference on IP + Optical Network, iPOP, 2015, Naha Okinawa, Japan.

iPOP2015: Xiaoyu Wang, Malathi Veeraghavan, Takahiro Miyazaki, **Ion Popescu**, Satoru Okamoto, and Naoaki Yamanaka, "Dynamic Layer-1 WAN Access Architecture for Large Enterprises", 11th International Conference on IP + Optical Network, iPOP, 2015, Naha Okinawa, Japan.

IEICE2015: Satoru Okamoto, Takahiro Miyazaki, **Ion Popescu**, Mirai Chino, Naoaki Yamanaka, Malathi Veeraraghavan, Eiji Oki, "Proposal of the dummy signal insertion method for realizing the adaptive link rate of the serial transmission type Flex Ethernet", *written in Japanese*, IEICE General Conference, 2015, Kyoto, Japan.

ITC2014: **Ion Popescu**, Bogdan Uscumlic, Yvan Pointurier, Annie Gravey, Philippe Gravey and Michel Morvan, "A cost comparison of survivable subwavelength switching optical metro networks", 26th International Teletraffic Congress, ITC, 2014, Karlskrona, Sweden.

ICTON2014: Yvan Pointurier, Bogdan Uscumlic, **Ion Popescu**, Annie Gravey, Qing Wei and Matthias Lott, "A comparison of subwavelength optical switching networks for LTE mobile backhauling", 16th International Conference on Transparent Optical Networks, ICTON, 2014, Graz, Austria.

NOC2014: **Ion Popescu**, Bogdan Uscumlic, Yvan Pointurier, Annie Gravey, Philippe Gravey and Michel Morvan, "Cost of protection in time-domain wavelength interleaved networks", 19th European Conference on Network and Optical Communications, NOC, 2014, Milan, Italy.

ONDM2014: Annie Gravey, Bogdan Uscumlic, Yvan Pointurier, **Ion Popescu**, Philippe Gravey and Luay Alahdab, "Efficient resource allocation in time-domain wavelength interleaved networks", 18th International Conference on Optical Network Design and Modeling, ONDM, 2014, Stockholm, Sweden.

NOC2013: Lida Sadeghioon, **Ion Popescu**, Bogdan Uscumlic, Philippe Gravey and Annie Gravey, "Virtual ring based protection for time-domain wavelength interleaved network", 18th European Conference on Network and Optical Communications, NOC-18, 2013, Graz, Austria.

ICTON2013: **Ion Popescu**, Lida Sadeghioon, Annie Gravey, Philippe Gravey and Michel Morvan, "Synchronization of the time-domain wavelength interleaved networks", 15th International Conference on Transparent Optical Networks, ICTON, 2013, Cartagena, Spain.

Résumé

Le trafic est en augmentation constante dans tous les segments du réseau de transport en raison de l'accroissement du nombre de terminaux connectés (Smartphone, tablettes, téléviseurs connectés) et de la diffusion de contenus en haute définition. Cette croissance en trafic devrait se poursuivre, et conduit à une utilisation importante de la bande passante. L'approche conventionnelle dans les réseaux de transport opaques consiste à agréger le trafic dans le domaine temporel à chaque nœud, afin d'optimiser l'utilisation de la bande passante au détriment de la consommation énergique et de la mise en œuvre de fonctionnalités électroniques coûteuses. Ainsi, les nouveaux réseaux de transport doivent supporter des interfaces à très haut débit et permettre un dimensionnement dynamique en utilisant les fibres existantes afin de répondre efficacement aux besoins des clients. Dans ce contexte, afin d'apporter la souplesse nécessaire dans la couche de transport, les technologies à base de commutation de paquets optiques peuvent être plus appropriées que la commutation de circuits opaques actuellement déployée.

La contribution de cette thèse fait partie du projet CELTIC-Plus Saser-SaveNet. Ce projet est consacré à l'évaluation des technologies différentes à base de paquets optiques et d'architectures pour les réseaux régionaux.

En cette thèse, nous abordons le problème de dimensionnement de TWIN (Time-Domain Wavelength Interleaved Network). Particulièrement, les coûts respectifs des différentes solutions à base de paquets optiques sont comparés. Deux scénarios sont considérés: les réseaux métropolitains protégés et LTE (Long Term Evolution) backhauling, avec l'accent mis sur les technologies TWIN et POADM (Packet Optical Add Drop Multiplexer). La commutation dans les deux technologies est réalisée en sous-longueur d'onde: TWIN opère sur une topologie maillée et POADM sur une topologie en anneau. Un unique modèle de dimensionnement de réseau, supportant la protection et le trafic de tipe multicast est proposé pour la première fois pour ces technologies.

Mots-clés : Réseaux optiques, Commutation de rafales optiques, TWIN, Protection

Abstract

Traffic in all segments of the transport network has been constantly growing in the past years as a result of increasing number of connected end-user terminals (smartphones, tablets, connected TVs) and the race towards higher definition contents. This growth is expected to continue, and it leads to significant bandwidth utilization. The conventional approach in opaque transport networks consists in aggregating/grooming traffic in the time domain at every node, in order to optimize bandwidth use at the expense of costly traffic grooming electronics and high power consumption. Thus, the new transport networks should support very high bitrate interfaces, dynamic bandwidth provisioning such that they meet customer's demands and use the existing fibers. In this context, optical packet-based technologies may be more appropriate than the currently deployed opaque circuit switching, in order to bring the required flexibility in the transport layer.

The contribution of this thesis is a part the CELTIC-Plus project SASER-SaveNet. The project is devoted to the evaluation of different optical packet-based technologies and architectures for regional networks.

Through this thesis, we address the problem of dimensioning in Time-Domain Wavelength Interleaved Network (TWIN). Furthermore, the cost of the optical transport solutions with sub-wavelength switching granularity is compared. Two scenarios are considered: survivable metro networks and Long Term Evolution (LTE) mobile backhauling, with the focus on TWIN and Packet Optical Add Drop Multiplexer (POADM) technologies. Both technologies apply optical sub-wavelength switching: TWIN on a physical mesh topology, and POADM on a physical ring topology. A single network dimensioning model, supporting protection and multicast traffic, is proposed for the first time for these technologies.

Keywords : Optical networks, Optical burst switching, TWIN, Protection

TELECOM Bretagne

n° d'ordre : 2015telb0365

Télécom Bretagne

Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3

Tél : + 33(0) 29 00 11 11 - Fax : + 33(0) 29 00 10 00