Contribution to the Design of Optical-packet Based Metropolitan Area Networks

Lida Mollazadeh Sadeghioon

N° d'ordre: 2013telb0285

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Bretagne-Sud

Ecole Doctorale - sicma

CONTRIBUTION TO THE DESIGN OF OPTICAL-PACKET BASED METROPOLITAN AREA NETWORKS

Thèse de Doctorat

Mention : STIC

Présentée par **Lida Mollazadeh Sadeghioon**

Département : OPTIQUE et INFORMATIQUE

Directeur de thèse : Jean-Louis de Bougrenet de La Tocnaye

Soutenance le 23 October 2013

Jury :

Mme Lena Wosinska, Professeur, KTH Royal Institute of Technology (Rapporteur)
M. Guido MAIER, Assistant Professeur, Politecnico di Milano (Rapporteur)
M. Jean-Louis de Bougrenet de la Tocnaye, Professeur, Télécom Bretagne (Directeur de thèse)
M. Emmanuel Boutillon, Professeur, Lab-STICC - Centre de recherche / Université de Bretagne Sud (Examinateur)
M. Yvan Pointurier, Ingénieur de recherches, ALCATEL-LUCENT (Examinateur)
M. Bernard Cousin, Professeur, Laboratoire de Recherche IRISA/Université Rennes 1 (Examinateur)
Mme Paulette Gavignet, Ingénieur de Recherche en Réseaux Optiques, Orange Labs Networks (Invitée)
M. Maurice Gagnaire, Professeur, Télécom ParisTech (Examinateur)
M. Philippe Gravey, Directeur d'études, Télécom Bretagne (Invité)
Mme Annie Gravey, Directrice d'études, Télécom Bretagne (Examinatrice)

Acknowledgements

It is with immense gratitude that I wish to acknowledge, first and foremost the infinite support and continuous help of my two wonderful supervisors Prof. Annie Gravey and Mr. Philippe Gravey. It is not only with their profound insight, full competence and constant presence, but also with their endurance and immeasurable kindness all the way during my studies that this thesis became possible.
I am very thankful to Prof. Jean-Louis de Bourgrenet de la Tocnaye, for accepting to be the director of my thesis in an excellent environment at the Optique Department of Télécom Bretagne.
During my thesis I had the honor and the pleasure to receive numerous advices from several members of Optique and Informatique Departments, who I wish to express my gratitude to all and in particular to Mr. Michel Morvan with his enlightening analyses and intriguing course and conversations on optical telecommunication technologies.
I am very thankful to all administrative members of Télécom Bretagne who never stopped helping me from the beginning to the end specially Mme. Anne-Catherine CARIOU Mme. Jennifer Romer, Mme. Armelle Lannuzel and Mme. Viviane Guillerm.
I am indebted to my many colleagues and friends who supported me during the years of my studies scientifically and morally and I wish to thank them all: Ion, Bougdan, Barbara, Aurélie, HouBo, Nam, Vinicius, Kedar, Mervin, Bernard, Théo, Hani, Samir, Rabia, Yulia, Maina, Pascaline, Soraya, Hamid, Neda, Laura, Rouzbeh, Patrig and many more all around the globe that I have had the privilege to learn and receive their kindness and positive energy.
And last but not least I thank wholeheartedly my parents, my sister Ladan, and my brother Ali who held me strong with their abundance love throughout my life.

Brest, October 2013

Lida Sadeghioon

Abstract

This thesis aims at proposing a transparent optical architecture for metropolitan area networks. This multi-service architecture allows supporting both unicast and multicast traffic and offers a performance suitable for metro networks. The approach that we followed was based on an optical packet ring concept. The first contribution of this thesis is a complete MAC, which is based on labels allowing identifying the different flows in order to handle then with dedicated mechanisms. Then, we propose original fast protection mechanisms, both for unicast and multicast. The protection type can be specified for each flow. In the last part of the thesis, we consider architecture based on the combined use of optical circuit and packet switching, in order to provide several levels of granularity while offering end-to-end transparency.

Résumé

Le but de cette thèse est de proposer une architecture optique transparente pour les réseaux métropolitains. Cette architecture est multi-service, supporte unicast et multicast, et offre une performance en matière de protection compatible avec les besoins d'un réseau de transport. Les travaux s'appuient sur la notion d'anneau de paquets optiques. La première contribution de la thèse est une MAC complète, basée sur l'utilisation de labels permettant d'identifier les flots et de leur appliquer des traitements différenciés. Dans une seconde partie, des mécanismes originaux de protection sont proposés, à la fois pour l'unicast et le multicast. Le niveau de protection peut être spécifié pour chaque flot. Une dernière partie propose de considérer simultanément commutation de circuits et de paquets optiques afin de supporter plusieurs niveaux de granularité tout en offrant une transparence de bout en bout.

Résumé

Introduction

La plupart des travaux menés dans cette thèse se sont appuyés sur une architecture d'anneaux de paquets optiques à base de Multiplexeurs à Insertion/Extraction de Paquets Optiques. Ces dispositifs sont dans la suite de ce document désignés par leur acronyme anglais P-OADM et par extension on utilisera le terme d'anneau P-OADM. Un anneau P-OADM est bi-directionnel (i.e. constitué de deux fibres optiques où les signaux se propagent respectivement dans les sens positifs et négatifs). Les données numériques sont transportés dans des trames de longueur fixe, appelés slots. Dans chaque fibre, plusieurs canaux de données sont multiplexés en longueur d'onde, ainsi qu'un canal de contrôle. Les slots de tous ces canaux sont synchronisés. Chaque nœud du réseau dispose, pour chacune des directions, d'un émetteur accordable en longueur d'onde) et d'un groupe de récepteurs, dédiés chacun à une longueur d'onde spécifique, pour les canaux de données, et d'un émetteur/récepteur pour le canal de contrôle. Cette architecture offre ainsi une granularité sub-longueur d'onde.

L'insertion des paquets sur l'anneau se fait de manière opportuniste, dans les slots se trouvant vides au niveau du nœud (après réception des paquets destinés à celui-ci). On notera de plus que le réseau est supposé avoir été préalablement dimensionné à l'aide d'une matrice de trafic connue. Avec cette méthode, chaque nœud ne peut insérer qu'au plus un paquet à la fois sur l'anneau, s'il trouve un slot libre et par conséquent les paquets en transit sont toujours prioritaires. La performance de cet accès opportuniste a été étudiée dans le cadre de la thèse de Bogdan Uscumlic, dans le cas d'un anneau uni-directionnel [Bogdan Thesis]. Ces travaux ont montré que, dans un réseau convenablement dimensionné, il était possible de remplir efficacement chaque longueur d'onde, typiquement avec un taux de l'ordre de% 80. Dans ces conditions, nous avons décidé de prendre la technologie P-OADM comme point de départ de nous travaux de thèse.

MAC multi-Service pour un réseau métropolitain à base d'anneaux P-OADM

Formulation du problème

Dans l'architecture P-OADM, la fibre est un support partagé dont l'accès doit être régulé par un protocole MAC (medium access control).

Les travaux antérieurs sur les protocoles d'accès pour un réseau P-OADM ont porté sur la définition de mécanismes destinés à différencier plusieurs classes de services, ainsi dans [SWING] un mécanisme de réservation com-

plexe susceptibles de prendre en charge l'ensemble des classes de trafic a été proposé. En fait, nous n'avons pas connaissance de l'existence d'une MAC globale pour ce type de réseau procurant à un réseau P-OADM toutes les caractéristiques requises pour les futures générations de réseaux métro.

Cette section a donc pour objectif de définir un protocole MAC pour un réseau de paquets optiques à base d'anneaux P-OADM bidirectionnels.

La MAC d'un réseau métro moderne doit permettre de :

- insérer/extraire des flots correspondant à de multiples services, notamment unicast ou multicast
- distinguer plusieurs types de flots
- fournir des méthodes d'accès à différentes classes de services : garanti et best effort

Contribution

La couche MAC est divisée en sous-couches ayant chacune des fonctions spécifiques. Il est ainsi possible de transférer des informations en provenance de protocoles de couche supérieures, en assurant une transparence vis-à-vis de ces protocoles et en offrant différents services à ces protocoles clients.

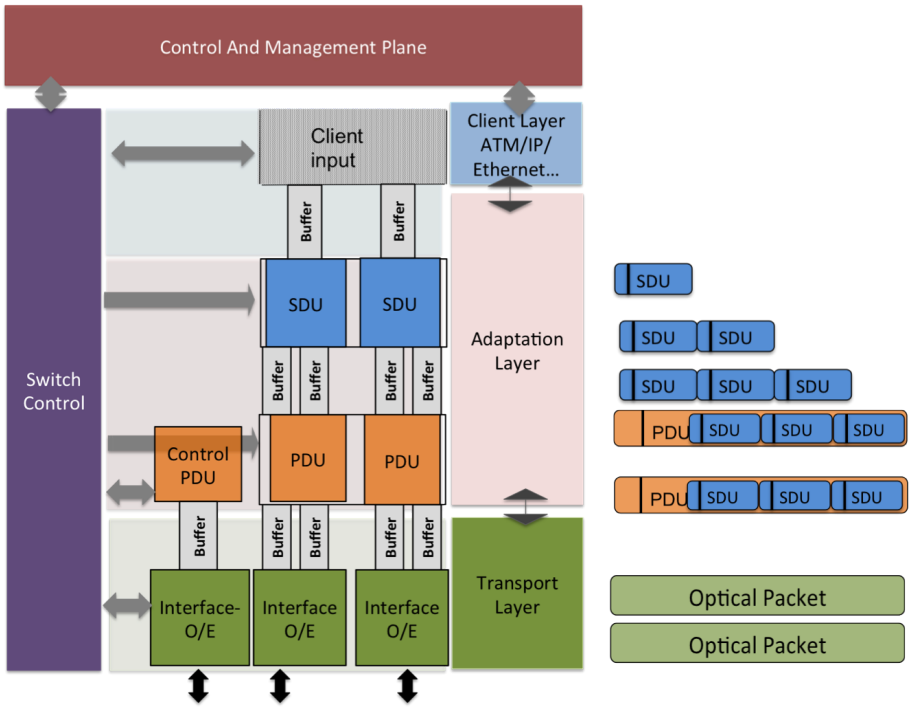

Figure 1: Structure de la MAC POADM et encapsulation des paquets

Nous avons décidé d'utiliser des labels dans le cadre de la MAC, ce qui nous a conduit à introduire un champ permettant d'insérer plusieurs labels

dans chacune des zones dédiées aux canaux de données des paquets de contrôle. Ces labels facilitent :

- l'identification de chaque flot selon sa source et sa destination ;
- la définition des différentes opérations à exécuter selon le service et le flot ;
- l'agrégation entre flux au niveau des nœuds d'interconnexion entre différents anneaux.

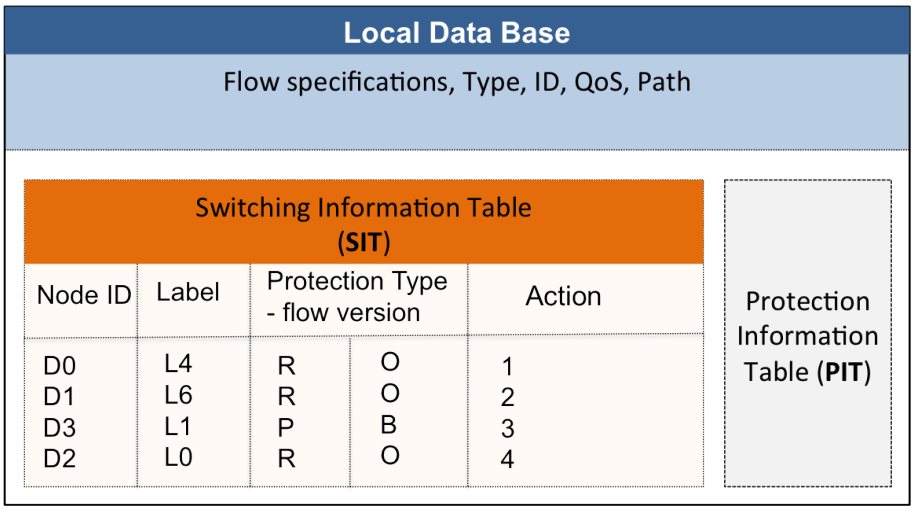

Figure 2: Table locale d'information d'un nœud d'un anneau P-OADM

Les informations relatives à la sous-couche d'adaptation sont transposées dans différents champs du paquet de contrôle : type de service, version du flot (original ou backup, pour la protection) et type de flot (par exemple unicast ou multicast).

Nous avons ainsi défini une MAC intégrant des fonctionnalités multiples. Les travaux décrits dans ce chapitre ont fait l'objet de plusieurs communications dans des conférences internationales : l'introduction de labels et la diffusion de flots multicasts ont été décrits dans un papier présenté à ONDM 2011 ; le support de plusieurs calasses de service grâce à un mécanisme de réservation a été décrit dans un papier à ETS 2011. Enfin, le transport de flots multicasts en utilisant les labels a été présenté à NGI 2012.

Benchmark La MAC multi-service que nous avons élaborée propose un ensemble de caractéristiques et de méthodes comparable avec celles qui sont accessibles en commutation de paquets électronique, comme le multicast et la commutation de labels.

Schémas de protection pour un réseau de paquets tout optique

Background
La résilience aux pannes est une des principales exigences que doivent satisfaire les réseaux de transport en général et particulièrement les réseaux

métropolitains. La résilience à une panne donnée est la capacité du réseau à continuer de fournir un service de transport après qu'une panne se soit produite. Ces pannes ont généralement une cause humaine. Il y a dans un réseau différentes causes de panne possibles, notamment une défaillance matérielle dans un nœud (par exemple une panne de transpondeur ou un problème logiciel entrainant une configuration erronée de l'équipement) ou dans une coupure de câble. Exceptionnellement, tout les équipements d'un nœud de réseau peuvent devenir indisponibles suite à une catastrophe naturelle, une coupure d'énergie ou un sabotage. Cependant, selon [Ramaswami], parmi toutes les causes de panne possible la coupure d'un câble est de loin le plus fréquent. Notre étude portera donc sur les remèdes à apporter en cas de coupure de fibre. Les technologies traditionnelles comme SONET/SDH procurent une disponibilité de 99.999% avec un temps de récupération du trafic inférieur à 50 ms, qui est devenu la référence pour un réseau de transport d'opérateur. Bien que les exigences en matière de récupération du trafic dépendent de la nature des données et des applications, la plupart des mécanismes de protection utilisés par les différentes technologies de réseau de transport (Resilient Packet Ring (RPR), anneau Ethernet, MPLS-TP) ont été conçus de sorte à satisfaire le critères des 50 ms maximum de temps de coupure.

Les anneaux RPR, à base de commutation de paquet électronique, ont été développés pour les réseaux métropolitains ou cœur. Ces anneaux à deux fibres disposent de deux mécanismes de protection : steering (redirection)ou wrapping (rebouclage). La figure () représente un anneau RPR à 8 nœuds. Le nœud N1 envoie des paquets à destination du nœud N3 via la fibre 0. Lorsqu'on utilise le mécanisme de redirection, tous les nœuds du réseau disposent d'une image de la topologie effective du réseau, qui est mise à jour par des messages de contrôle envoyés périodiquement ou en cas d'une modification de la topologie. Si une coupure de fibre intervient entre N2 et N3, celle-ci induit une modification de la topologie qui est notifiée par des messages de contrôle. N1 dispose d'une image de la nouvelle topologie, sur la base de laquelle il émet les paquets destinés à N3 sur la fibre 1. Lorsque la protection se fait grâce à un mécanisme de rebouclage, en cas de coupure entre N2 et N3, N1 continue à envoyer ses paquets sur la fibre 0 et N2 les aiguille de façon à les transférer sur la fibre 1, jusqu'à l'autre bord de la coupure, c'est-à-dire N3. Au niveau de N3, les paquets sont de nouveau aiguillés sur la fibre 0 et atteignent ensuite leur destination.

En résumé, chacune des deux méthodes précédentes comporte deux phases :

1. **Notification/ localisation de la panne**
2. **Sélection d'une procédure de protection**

Formulation du problème

Cette section est consacrée à l'étude de la résilience d'un anneau P-OADM et à l'analyse de différent schémas de protection. Pour cela, il est néces-

saire de définir des mécanismes pour chacune des deux phases rappelées ci-dessus.

Contributions En nous appuyant sur la MAC présentée dans la section précédente, nous avons étudié comment adapter les schémas de protection usuels 1+1 et 1:1 pour des trafics unicast ou multicast. Tout d'abord,nous avons suggéré une procédure d'identification et de localisation des pannes en introduisant dans le paquet de contrôle un champ de notification de pannes, de façon à diffuser ces notifications à l'ensemble des nœuds. De plus, nous introduisons une table dans chacun de ces nœuds qui permet de déterminer les flots affectés selon l'emplacement de la coupure.

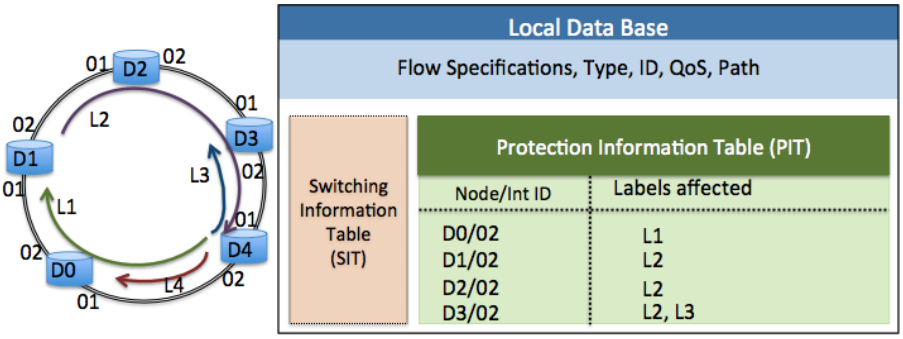

Figure 3: Contenu de la table d'information de protection au sein de la base de donnée locale du nœudD4.)

Sur cette base, nous avons étudié comment transposer les schémas classiques de protection 1+1 ou 1:1 au cas d'un anneau P-OADM bidirectionnel. Les schémas proposés sont dénommés respectivement "premium" et "regular". L'analyse des digrammes temporels des échanges de paquets (de contôle et de données) au sein de l'anneau a permis d'identifier et de quantifier les trois principales sources de dégradation de la performance : perte de paquets, arrivée en désordre et duplication. Comme attendu, la méthode "premium" n'engendre aucune perte et son coût est environ deux fois celui d'un anneau non protégé. Le prix de la méthode "regular" se situe entre les deux précédents, car elle permet un partage de certaines des ressources de protection. Panacher ces deux méthodes peut constituer un bon compromis pour un opérateur. Enfin, nous avons adapté ces deux précédentes à la protection du trafic multicast. Nous avons discuté deux variantes, utilisant soit l'approche usuelle d'extraction des paquets par la source du flot multicast soit une méthode originale où cette extraction se fait dans un nœud proche du point opposé à la source sur l'anneau. Cette dernière permet de supprimer pratiquement toutes les pertes de paquet. De façon générale, les études de performance ont mis en évidence que les pertes étaient toujours très faibles, avec un temps de coupure du trafic borné par le temps de parcours autour de l'anneau, soit au plus quelques millisecondes dans un réseau régional.

Benchmark Tous ces mécanismes que nous avons proposés offrent un temps

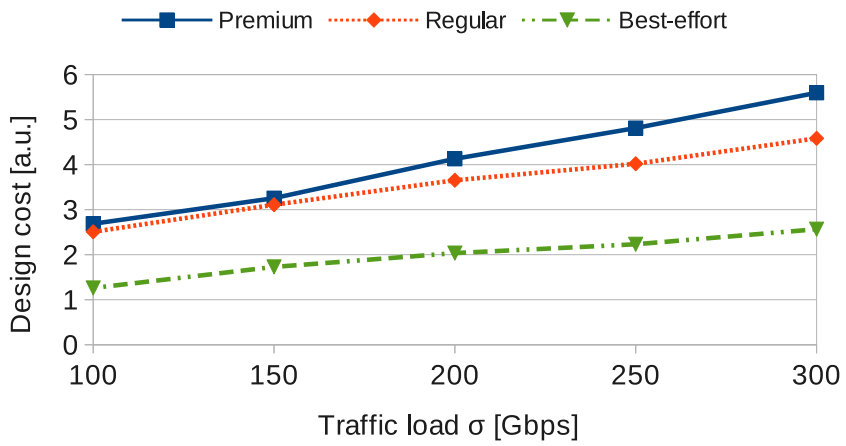

Figure 4: Coût d'un réseau dimensionné pour différents types de protection level et $\theta = 40\%$

de récupération pratiquement égal au temps de propagation à travers l'anneau. Ils offrent donc ainsi une performance compatible avec le délai standard de 50 ms. Il est important de noter que ces mécanismes opèrent au niveau des paquets et permettent d'offrir une protection propre à chaque flot. A notre connaissance, cette possibilité de fournir une protection rapide et adaptée à chaque flot constitue une caractéristique unique.

Transparence optique dans des réseaux métropolitains de topologie complexe

Formulation du problème

Dans ce chapitre, on s'intéresse à la possibilité de bénéficier de la transparence optique dans un MAN de topologie arbitraire avec une granularité inférieure à la longueur d'onde. En effet, les topologies des MAN ont tendance à devenir plus complexes qu'un anneau ou des anneaux interconnectés. En principe la technologie OBS permet une transparence de bout-en-bout avec un granularité correspondant à un burst. Mais de manière générale, cette technologie ne fonctionne correctement que dans des réseaux faiblement chargés. Il y a cependant deux approches qui permettent un multiplexage optique efficace:

1. Les anneaux P-OADM, dès lors qu'ils ont été dimensionnés conformément à une matrice de trafic donnée. Cependant ces anneaux ne permettent pas de couvrir des réseaux très étendue et de plus leur interconnexion nécessite des conversions O/E/O.
2. Les réseaux du type TWIN dans lesquels l'ordonnancement se fait off-line et est optimisé pour une matrice de trafic donnée. Bien que cette technologie permette de concilier transparence de bout-en-bout et granularité inférieure à la longueur d'onde, le fait qu'elle attribue une longueur d'onde par destination pose un problème de passage à l'échelle.

En général, la granularité inférieure à la longueur d'onde n'est pas requise pour tout le trafic transporté par le réseau. Ainsi, le trafic échangé entre deux data centers peut utiliser toute la capacité d'une longueur d'onde. Par conséquent, il est souhaitable d'examiner comment des circuits et des paquets optiques peuvent être intégrés dans une architecture commune.

Contribution Dans une première approche, nous avons proposé une architecture hybride P-OADM/OCS. Les anneaux P-OADM servent à transporter la partie de trafic de granularité inférieure à la longueur d'onde et agrègent de façon efficace les différents flots de faible taille. Le réseau OCS met en œuvre - quand cela est nécessaire - à l'aide de ROADM "colorless et directionles" des circuits optiques dynamiques ayant pour granularité la longueur d'onde. Nous avons proposé une structure de nœud combinant les deux technologies. De plus, nous avons défini un algorithme qui permet de construire une topologie à base d'anneaux P-OADM multiples pour supporter la part du trafic transporté en mode paquet. Chaque flot est transporté sur un seul anneau, ce qui rend possible une transparence de bout-en-bout. Le choix des anneaux peut être optimisé en tenant compte des coûts respectifs des récepteurs nécessaires dans chaque nœud et des longueurs d'onde nombre de longueurs d'onde. Cette méthode a été appliquée à un réseau constitué de 7 nœuds pour différents scénarios de trafic. La part du trafic transportée en mode paquet reste assez faible mais le recours partiel à la technologie P-OADM permet une réduction significative du nombre total de transpondeurs.

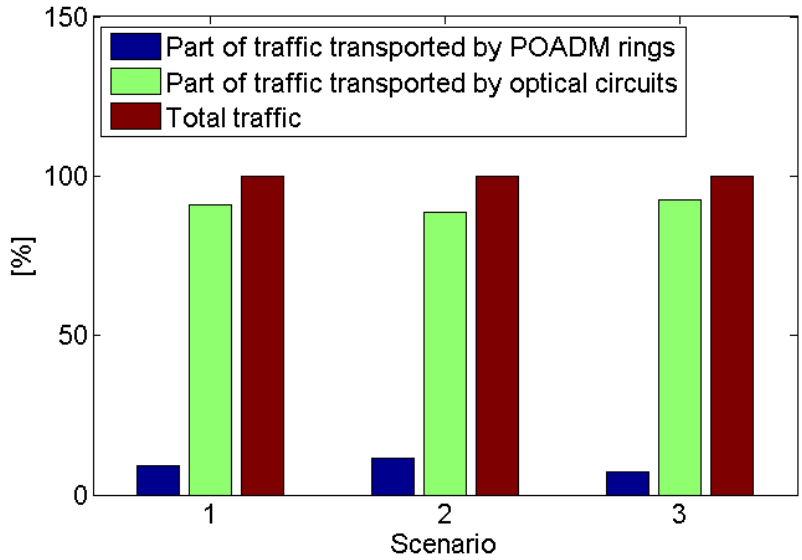

Figure 5: Participation of OPS and OCS traffic in the final solution

La seconde approche étudiée s'appuie sur une architecture originale de plan de commande d'un réseau du type TWIN, basée sur un ou plusieurs anneaux virtuels. Ces anneaux assurent la diffusion et la mise à jour régulière des tables utilisées pour ordonnancer les paquets émis par les

nœuds source. Les anneaux virtuels peuve,t également servir pour définir des chemins de secours à fin de protection du réseau TWIN.

Benchmark Les approches hybrides paquets/circuits optiques ont fait l(objet de beaucoup detravaux ces dernières années [1]. Nos travaux ont cependant permis d'identifier des méthodes pour déterminer un combinaison optimale entre OPS et OCS dans des réseaux hybrides.

Acronyms

ECOFRAME	Eléments de COnvergence pour les Futurs Réseaux d'Accès et MEtropolitains haut débit
MAN	Metropolitan Area Network
WAN	Wide Area Network
CO	Central Office
PoP	Points of Presence
TDM	Time-division multiplexin
SONET/SDH	Synchronous Optical Network/Synchronous Digital Hierarchy
PDH	Plesiochronous Digital Hierarchy
ATM	Asynchronous Transfer Mode
VLAN	Virtual Local Area Network
RPR	Resilient Packet Ring
MPLS	MultiProtocol Label Switching
MPLS-TP	MultiProtocol Label Switching Transport Profile
PBT	Provider Backbone Transport
IP	Internet Protocol
MAC	Medium Access Control
TDM	Time Division Multiplexing
WDM	Wavelength Division Multiplexing
FTTX	Fiber To The x
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
OCS	Optical Circuit Switching
OPS	Optical Packet Switching
OBS	Optical Burst Switching
FDL	Fiber Delay Line
ADM	Add/Drop Multiplexer
PSC	Passive Star Coupler
AWG	Arrayed-waveguide grating
OXC	Optical Cross-Connect

WSS	Wavelength Selective Switch
OADM	Optical Add/Drop Multiplexer
ROADM	Reconfigurable Optical Add/Drop Multiplexer
POADM	Packet Optical Add/Drop Multiplexer
SOA	Semiconductor Optical Amplifier
HORNET	Hybrid Opto-electronic Ring Network
DAVID	DAta and Voice Integration over DWDM
OPST	Optical Packet Switch and Transport
MAIN	Metro Architecture EnabliNg Subwavelength
RINGO	Ring Optical network
TWIN	Time-Domain Wavelength Interleaved Networking
OTN	Optical Transport Network
FIFO	First In First Out
LP	Linear Program
ILP	Integer Linear Program
MILP	Mixed Integer Linear Program
SDU	Service Data Unit
PDU	Protocol Data Unit
MR-N	Multi Shared Receivers per destination without flow splitting
ITU	International Telecommunication Union
IETF	Internet Engineering Task Force
CAPEX	Capital Expenditure
MuD	Multicast Drop off point
COHYB	Choosing Optical packet switching rings for Hybrid network
QoS	Quality of Service
SIT	Switching Information Table
PIT	Protection Information Table
PIT	Protection Information Table
CE	Control Entity
VTR	Virtual TWIN Ring

Contents

Introduction

Metropolitan area network (MAN) constitutes the middle segment in the generic three layer telecommunication network hierarchy, mostly referring to a dense metropolitan or regional zone and, it is the focus of this thesis. It has a frontier with the customer premise-based side (Access Network), connecting central offices (COs), and, on the other side, it is connected to the service provide backbone transport network (Core network) via Points of Presence (PoPs) Fig6. MAN thus mainly aggregate traffic generated in the access network up to the PoP.

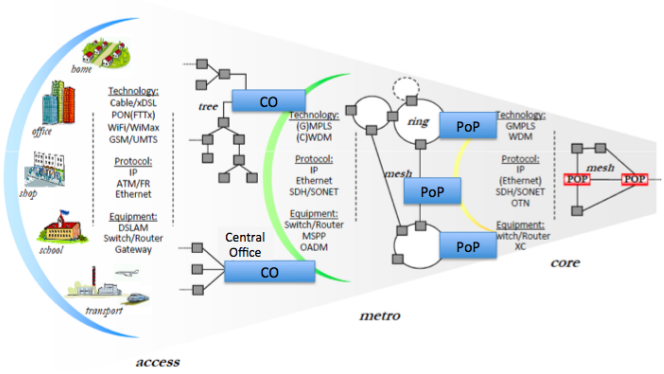

Figure 6: Telecommunication Network Hierarchy: Access, Metro, Core.

The term MAN in fact has its origin in data communication field, while terms such as aggregation or concentration network were used in legacy telephonic networks. However it is now widely applied also to operator networks, thus we will use the terms MAN or simply "metro network" through all this thesis. MAN experienced a fast change in late 1980s and early 1990s, with the introduction of optical fibre and SONET/SDH rings. SONET/SDH provided reliable time-division multiplexed (TDM) with in-service monitoring and less than 50 ms failure recovery time. Generally, the SONET/SDH rings had two or three levels with different spans: the edge ring where add-and-drop multiplexers (ADMs) aggregated the traffic electronically and hub the groomed traffic towards the COs, in the distance range of a few ten of kilometres. The second and the third levels are the Inter-Office rings that interconnect Edge rings via digital cross connects switches through hundreds of kilometres distance range. There were other technologies such as Fiber Distributed Data Interface (FDDI) and Asynchronous Transfer Mode (ATM) in MAN region

that merely disappeared and SONET/SDH remain the dominant technology in metro network.

At the end of the 1990s, wavelength division-multiplexing technology began to be used in SONET/SDH Inter-Office rings to offer higher capacity as feeder rings and allow physical dual fibre ring to carry multiple parallel SDH rings [4]. Still interconnected rings are preferred topology in metro area networks since ring topology offers more efficient fiber sharing and higher resiliency.

The high cost of O/E/O conversion at interconnecting points and the need to have an upgradable network to handle the increasing amount of traffic became a strong motivation for transparency study in the metro network. With emerging optical devices such as: Optical Cross Connect and Optical Add/Drop Multiplex (OADM) offering wavelength level reconfiguration flexibility, the ADMs were replaced in some metro networks, thus the SONET/SDH rings were expanded to WDM optical metro network [2]. In the same period, the local access network became engaged in a deep, still continuing, transformation with the increase of access rates through xDSL and FTTx and the development of mobile services.

Rapid growth of packet based services especially bursty Internet traffic and emerging new applications such as video and VOIP with different bandwidths and QoS requirements have changed the nature of expectations of metro networks. In spite of the introduction of new standards like LCAS (Link Capacity Adjustment Scheme) [6], SONET/ SDH circuit switched or wavelength switched WDM rings are no more capable of offering an efficient infrastructure in metro segment. Thus packet-based technologies became the focus of telecommunication industry standardisation forums and communities such as Internet Engineering Task Force (IETF) working and the Metro Ethernet Forum (MEF). All offering electronics level packet switching and grooming solutions to achieve higher bandwidth utilisation suitable for metro networks.

In the next generation metro network, any solution must fulfil the following requirements:

- Scalability: The capability to expand the network capacity in terms of bandwidth and node number.
- Multi-service: From different client layer protocol such as IP, Ethernet, to different service demands such as unicast, multicast in different class of services.
- Switching granularity: The nature of metro network traffic requires multiple level of granularity as oppose to a rigid circuit level switching to share the bandwidth and increase the throughput.
- Optical transparency: Reduce OEO conversion cost where is possible decrease the overall cost of the network and removes the unnecessary electronics processing bottlenecks.
- Network management and Simplicity: Extensive control and management facilities must be available in metro network in order to provision and operate the network. Mostly dynamic end to end service provisioning for emerging applications is required.

- **Availability and reliability: Metro network must be able to survive multiple failures. Protection in nodes and links must be available in order to provide the reliability necessary for carrier grade transport network.**

All these features need to be provided while operators are experiencing a continuous pressure on their capital and operational expenses (CAPEX and OPEX) as the general evolution of the business models leads to a steady decrease of the operator's revenue per transported bit.

In that context, transparency that is reducing the number of unnecessary (from the service point of view) optical to electronic conversions is a target that has been for long time identified by the research community.

Moreover, to address the above mentioned next generation metro network requirements, it would be attractive to combine the reduction of the optical to electronic conversions with the multiplexing granularity provided by packet switching by introducing "all optical packet switching" [5]. There are many experimental projects on optical packet metro rings [7] that a few of them are presented in the State-of-the Art chapter. The study in this thesis is based on one of the most advanced among these projects, called ECOFRAME. ECOFRAME was a French project, conducted during 2007-2010. This project introduced packet-OADM (POADM) in a WDM time slotted unidirectional ring [8]. A primary MAC was developed during the project with a single class guaranteed traffic. Issues such as network modelling and simulation for performance evaluations, unidirectional ring dimensioning were addressed at Telecom Bretagne [1] [3]. At the beginning of this work, our intention is to extend the work carried in ECOFRAME in order to improve the applicability (moving from single fibre unprotected ring to a dual fibre protected one) and scalability (moving to multiple interconnected rings) of the POADM technology and integrating metro services such as Multicast and QoS. Moreover we aim to investigate solutions to achieve end to end optical transparency and multiple level granularity, considering the trade-off between packet and circuit optical switching. A general overview of the thesis structure is in the following:

The first chapter of this thesis provides a brief state-of-the art of optical switching techniques and optical transport technologies in MANs. The second chapter then presents the extension of the POADM MAC. In particular, different kinds of labels are introduced in the ring control channel and a reservation mechanism is applied to non guaranteed traffic. The third chapter focuses on the implementation of fast optical packet-level protection mechanisms in a POADM ring. Chapter 4 presents alternatives to a 100% POADM architecture. The first one combines optical packet- and circuit-switched techniques. The second one is another packet-based technology named TWIN (for Time Wavelength Interleaved Network), which is regaining a lot of attractiveness as power consumption becomes an increasing OPEX factor, since it relies on passive core nodes. In both cases, topology design issues are addressed. Finally, the conclusion recalls the main results and discusses possible extensions or alternative to the network approaches that were studied in this work.

Bibliography

[1] **B. Uscumlic, "Optical Architecture and Traffic Engineering in Optical Metropolitan Networks", PhD Thesis, Télécom Bretagne, Brest, 2010.**

[2] **P. Toliver, R. Runser, J. Young, J. Jackel, "Experimental Field Trial of Waveband Switching and Transmission in a Transparent Reconfigurable Optical Network", OFC'03, pp. 783-784, Atlanta, GA, February 2003.**

[3] **B. Uscumlic, A. Gravey, P. Gravey, I. Cerutti, "Traffic Grooming Issues in WDM Optical Packet Rings", ITC'21, Paris, 15-17 September 2009.**

[4] **W. T. Anderson, "The MONET Project Final Report", OFC'00 Baltimore, MD, pp. 148-149, 2000.**

[5] **Shun Yao, S.J.B. Yoo, B. Mukherjee, S. Dixit, "All-optical packet switching for metropolitan area networks: opportunities and challenges", IEEE Communications Magazine,Vol.39, Issue 3, pp.142-148, March 2001.**

[6] **ITU-T Rec. G.7042, "Link Capacity Adjustment Scheme(LCAS) for Virtual Concatenation", Octber 2001.**

[7] **M. Herzog, M. Maier, M. Reisslein, "Metropolitan area packet-switched WDM networks: A survey on ring systems",IEEE Communications Surveys Tutorials,Vol. 6, Number 2, pp.2-20, 2004.**

[8] **D. Chiaroni, "Optical Packet Add/Drop Multiplexers for packet ring networks", Proc.35th European Conf. on. Opt. Commun. (ECOC), Brussels, Belgium, Sept.2008.**

Chapter 1

State of the art

Contents

Introduction

This chapter presents a brief background review of optical switching and transport technologies in metropolitan area networks with more focus on packet based paradigms.
The first part provides a short survey of optical switching technologies in terms of transparency and granularity along with the enabling optical components used in these technologies.
In the second part a few current and experimental solutions including the Packet-OADM ring architecture that is the basis of chapter 3 and 4 of this study are presented. There are plenty of proposed solutions under various research projects in this domain, however we focus on the most relevant and similar proposals to Packet-OADM based ring technology.

I Optical Switching Technologies

I.1 Switching Transparency and Granularity

The optical networks are the preferred medium for the fixed data transport network since fiber offers a tremendous amount of bandwidth with minimum attenuation and longer reach. In general there are two major type of optical networks:

Opaque optical network In opaque optical networks all the point to point links are fibers, however optical-electronic-optical conversion is performed at each node for the processing and all other functions such as multiplexing and switching like SONET/SDH networks.

Optically transparent network Optically transparent networks bypass the transit optical data channel/signal optically, thus from source to destination signal stays in optical domain. In addition multiplexing and add/ drop functions are all done all optically.

There is another type that is called translucent networks. It is a compromise between the opaque and transparent networks. In this type of network traffic continues on the optical bypass till it needs to be regenerated (it reaches its optical reach) then O/E/O conversion and regeneration is done to . Thus translucent networks is a combination of all optical and opto-electrical nodes[1]. In addition to the degree of the transparency in the optical networks specially in metropolitan area scope there is the switching granularity degree that is another defining factor in the networks.

In terms of granularity there are two major methods in optical switching and the third solution that is a combination between the first two:

Optical Circuit Switching(OCS)

In OCS optical switching is done at the light path level. The light path definition relies on different switching technologies where each of which presents a constant or variable level of grooming in fixed containers. These containers can be as coarse as waveband or spectrum slice including a few number of wavelengths to a single wavelength in a fixed grid ITU-based network or very narrow spectrum in a flex-grid or grid-less optical network as a light path in a WDM or OFDM based network. In a sub-wavelength granular lever in OCS a time slot in a WDM/TDM based network is also considered a light path or a circuit to be switched. Usually the duration of the established circuits are long and also the QoS provided by this method is high thanks to the stability of the multiplexing and switching methods [2].

Optical Packet/Burst Switching(OPS/OBS)

Optical packet switching:
OPS offers the finest sub-wavelength granularity. The main advantage of OPS is surely the statistical multiplexing that leads to higher throughout and maximising the network resource utilisation[2]. The ideal OPS would be relying on optical buffers and optical signal processing components such as optical gates in order to store the packet while processing the content of its header or for contention resolution purposes at input/output of the node. However despite of research communities continuous effort these components are not currently available out of the labs. Therefore there are different methods to mimic the OPS functionality by dividing the control channel and header information from the actual data packet and by means of partial transparency as explained earlier to process the header information in electronics domain and keep the data packet in optical domain. For implementation purposes there are intermediate solutions such as Optical burst Switching, and synchronous slotted or un slotted packet switching[3][4] .

Optical Burst Switching:
In OBS the burst is still defined as a sub-wavelength container that can have a variable or fixed size to carry more than one IP/ Ethernet Packet and the burst header information is carried in a separate channel that is transmitted in band and it is processed in electronically. Burst header specifies the destination and is sent in advanced before the data burst to set up the path. Depending on the OBS protocol the signalling for light path setup can be hop by hop (e.g., JET-OBS) or end to end (e.g., wavelength-routed-OBS)[5][6]. The data burst is transmitted with an offset delay after the control header is sent to the network. However the throughput of asynchronous network is considerably low with the blocking probability that is not desirable in a data transport network therefore the synchronous method is more considered to be an alternative for pure OPS [7].

Hybrid Optical circuit and packet Switching:
The hybrid switch definition here points to multi-granular optical switch-

ing solution that combines OCS and OPS switching technologies in order to benefit from both granularity and transparency. As a result traffic can be switch at fiber, waveband, wavelength and sub-wavelength level throughout the network depending on the hybrid switch architecture. There are increasing amount of researches on Hybrid optical switching techniques that combine slow and fast optical switching elements in an architecture to provide a transparent yet flexible solution for transport network [8] [9][10] .

I.2 Key Components in Optical Transport Networks

There are several key components that are deployed as the building blocks of optical switching structure in the various WDM Metro network architectures. In this part we concentrate on the advantages of a few of these components, in addition to their applications in the evolution of node structure in the next generation optical Metro networks.

Passive Star Coupler(PSC)

In general coupler is a passive multipurpose optical component that has variety of applications such as tap off small portion of optical signal for control and monitoring purpose, a passive star NXN broadcast structure, and a Multiplexer/Demultiplexer in more complex optical components such as optical switch architecture. The basic coupler functionality is combining/splitting light into/out of the fiber. A generalised NXN PSC with equal optical power splitting ratio α, is such that $\alpha = 1/N$, Hence, the input optical signal power is slitted at each output port according to $P_{out} = Pin/N$. Depending on the place and the role that PSC is deployed, the splitting ratio α may differ [11].

Arrayed-waveguide grating(AWG)

An AWG is a key passive optical component that is deployed in the optical node structures e.g. as an NXM wavelengths multiplexer/demultiplexer. The fundamental operating building blocks in an AWG are $NX1$ or $1XN$ coupler and arrayed wave-guide. The arrayed-waveguide functionality is like a prism that creates wavelength dependant phase shifts on the several copies of the same optical signals coming through the $NX1$ coupler, then after the phase shift the different wavelengths at the output pass through another coupler and are accessed from different output ports. The AWG is also used as static wavelength shift router and wavelength cross connect [12].

Optical Cross Connect(OXC)

OXCs are advanced optical network elements that are capable of switching light paths, adding and dropping locally generated and terminated client layer traffic, and configuring optical network topologies. They are particularly useful for mesh topologies or to interconnect WDM ring in metro networks.

As a reconfigurable device it is appropriate for protection, path reconfiguration. The granularity of the switching relies on the switching fabric,

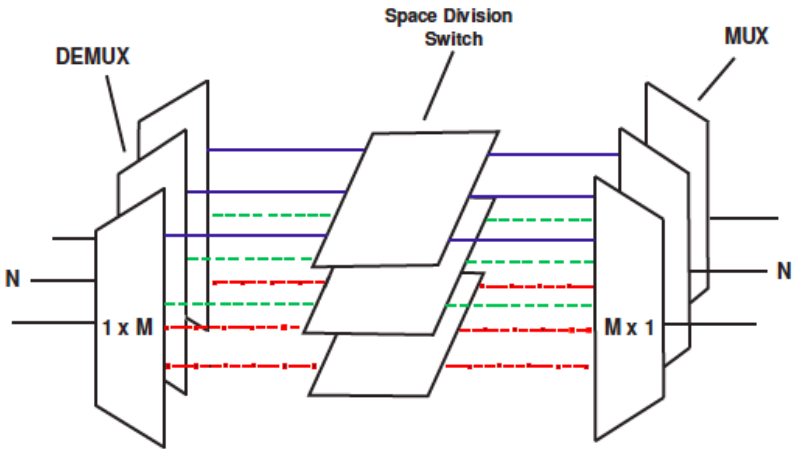

Figure 1.1: Optical Cross-Connect (OXC)(according to [2])

e.g. here there are three different categories of OXC with different granularities: Fiber Cross Connect, Waveband Cross Connect and Wavelength Selective Cross Connect. In general OXC consists of three major parts: Input/Output ports, switch fabric, and control unit. At network level between all OXC control unit there must be a communication and control to setup and tear down an end to end light path[2].

Wavelength Selective Switch (WSS)

Wavelength Selective Switch (WSS) is recently the core of current DWDM reconfigurable Agile Optical Networks. It can be dynamically reconfigured to route or block certain wavelengths, Fig1.2. The NxN wavelength selective switch can be made by deploying MUX/DEMUX to separate the wavelengths then an OXC or any space switch to alter/switch the path of the wavelength to different ports. The middle switch speed determines the reconfiguration time of the WSS. Currently the industrial WSS available are based on Micro ElectroMechanical Systems (MEMS) [13] and liquid Crystal On Silicon (LCOS) technologies [14], where both do not represent less than msec switching time.

Semiconductor Optical Amplifier(SOA)

SOAs are very attractive optical components due to their linear characteristics, their large nonlinear coefficients, the compact size since that they can be monolithically integrated with other optical components. They are becoming major role players in ultra-high bit-rate optical networks as space switches, signal regenerators and wavelength convectors.

As active space switches they have less than 1 ns switching time with 45dB extinction ratio [15].

The high speed response of the SOA switch and its implementation in different architecture has made it possible to consider it as sub-wavelength granular switch for OPS or OBS switching node structure.

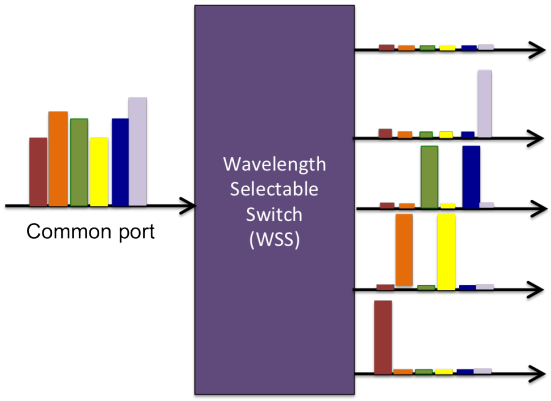

Figure 1.2: Wavelength Selective Switch functionality

Reconfigurable-Optical Add Drop Multiplex (ROADM)

The structure and functionality of Reconfigurable-Optical Add Drop Multiplex (ROADM) is a crucial component in evolving of optical transport network, since it provides optical add/drop and pass-through functions dynamically without the need to terminate the line, like it was necessary before. Early designs of optical add/drop multiplex were fixed OADM. Then by increase of the traffic it became necessary to add reconfigurability to add/drop function, and now multi-degree direction-less colour-less design shown in Fig1.3 is needed in order to add/drop and pass-through traffic from every direction. This way shared mesh restoration is possible, therefore the number of wavelength needed for network protection drops drastically thus more resources are available and this enables resource hungry applications[16]. The modular structure of the

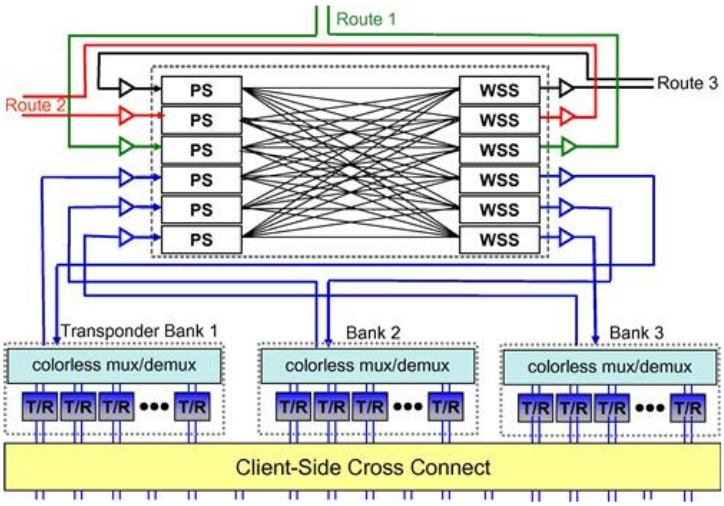

Figure 1.3: Architecture of a Colourless, Directional-less with a client side Cross Connect(according to [16])

Fast Tunable Laser Source

Fast Tunable Laser source is an optical source that can be controlled

and tuned to switch from one wavelength to another in less than a few hundreds of nano seconds for a channel spacing as low as 12.5 or 50 GHz. It has gained a lot of attention specially in packet switched optical networks as an optical tunable transmitter, since it brings fast switching more to the edge of the network [17].

II Transport Technologies in Metropolitan Area Networks

In the following there are examples of WDM metro network technologies with different granularity and transparency degrees.

Each of the examples benefits from one or combination of optical switching components mentioned above and they provide different degree of flexibilities due to their topology or switching architecture.

II.1 Optical Circuit Switched Networks

Optical circuit switching technologies mainly rely on slow optical switching components such as OXCs, and ROADMs. Therefore the channels are usually stablished for long time and there is no 'on the fly' reconfiguration. Nevertheless OCS provides steady connection with high QoS to the clients. Optical Transport Network (OTN) is ITU G.709 standard, defined as the key standard for optical circuits switched networks [18]. OTN has inherited many characteristics from the legacy SDH technologies such as: framing and scrambling, layers, automatic protection switch signalling. However it is based on both optical and electrical layer structures unlike SDH that fully relies on electrical layer services[19] Thus it offers a TDM-based method for aggregation and multiplexing low rate OTN signals into higher rate optical channels transparently.

OTN adds the overhead standard based Operations, Administration, Management and Provisioning (OAM-P) functionalities to the data payload and creates Optical Data Unit (ODU-n) at 1.25 Gb/s, 2.5 Gb/s, 10 Gb/s, 40 Gb/s, 100 Gb/s and ODU-flex. ODU-flex can accommodate client signal with any bitrate. Fig 1.4 illustrates OTN layers and (ODU-n) multiplexing hierarchy.

The OTN multiplexing bandwidth granularity is almost twice higher in order of magnitude than SONET/SDH, therefore it is more scalable to higher rates. In general, OTN provides a unified way as a digital wrapper technology for transport network to manage variety of client protocols and service infrastructures.

II.2 Electronics Packet Networks

This section presents current electronics packets switched technologies available in metropolitan area networks. Each benefits from statistical multiplexing, however O/E/O conversion is necessary at each node through the network.

RPR

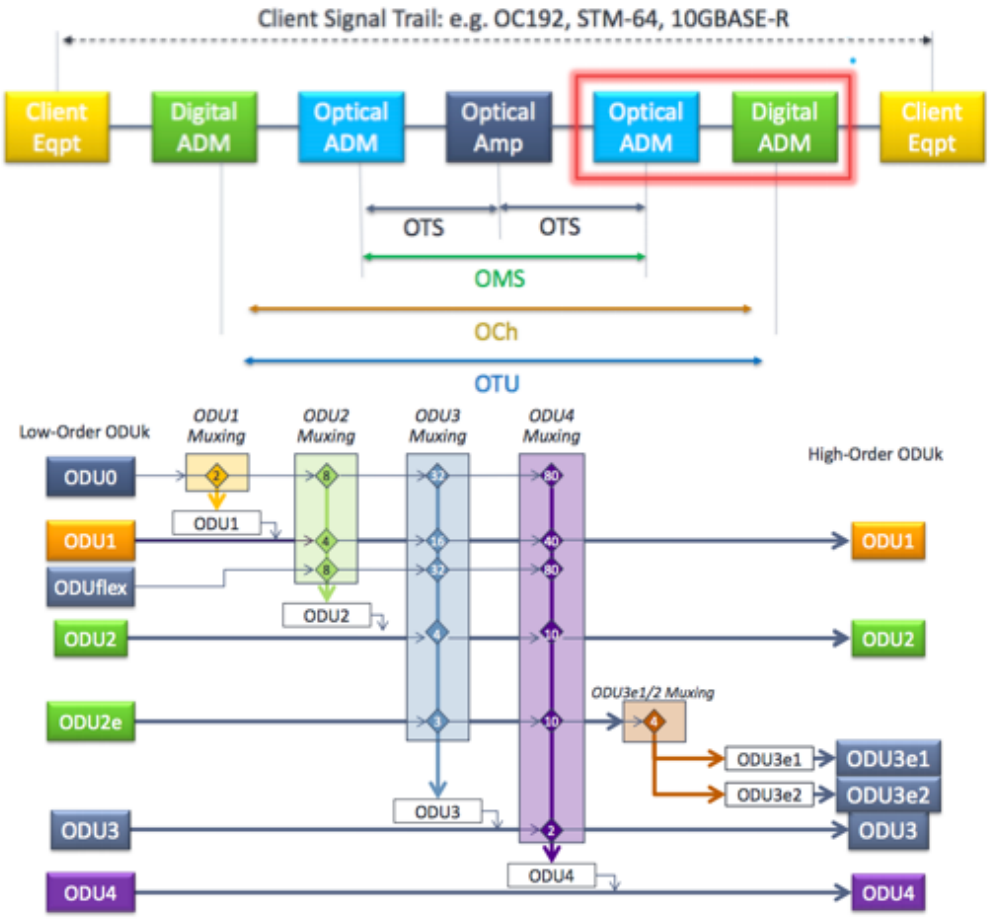

Figure 1.4: OTN Layers and ODU Multiplexing Hierarchy (according to [18])

The IEEE 802.17 Resilient Packet Ring (RPR) is a operational developed standard for bidirectional packet-switched metropolitan rings [20]. Both directions of the bidirectional (ringlets) are used to transport working traffic between nodes, in order to utilise the whole available bandwidth of the network. 1.5.

Here are the main features of RPR:

- **Efficient Unicast and Multicast transport system**
  - **Spatial reuse: The protocol supports destination packet removal so that a packet will not traverse all ring nodes and spatial reuse can he achieved.**
  - **Multicast: One Multicast packet is sent to the network and there is no need to replicate the multicast flows.**
- **Class-based service: It provides a three-level class-based traffic priority scheme with three level of latency-jitter. There are primary transit queue (PQT) for the class A with low jitter and secondary queue transit (SQT) for class B with predictable latency and jitter, no packet is discarded even if it belongs to best effort class.**

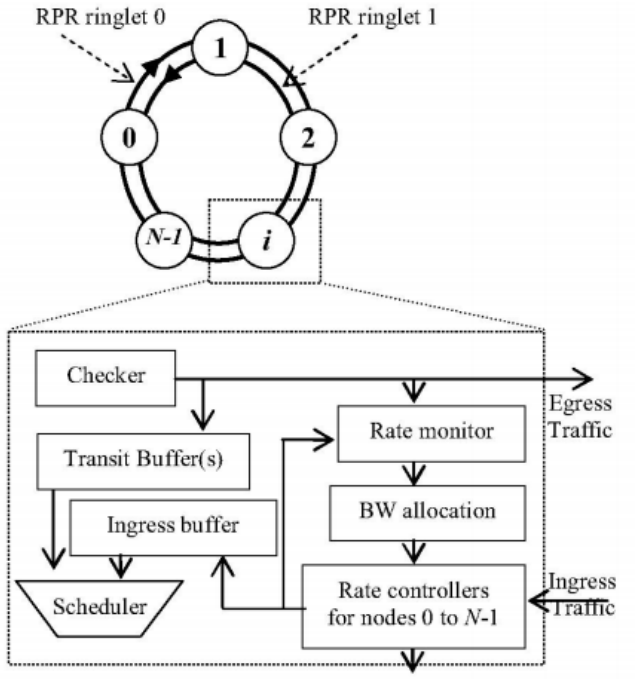

Figure 1.5: RPR ring and nodal architecture(according to[21])

- **Resilience: There are two main protection mechanism in RPR avoiding failed spans within 50 ms:**
  - **Wrapping Nodes at both end of the failure direct the packets away from the failure by wrapping traffic around to the other ringlet.**
  - **Steering The protection mechanism notifies all the nodes on the ring of the failure. All the nodes on the ring modify their topology accordingly to avoid the failure.**

RPR provides a three-level class-based traffic priority scheme. The objectives of this scheme are to let class A be a low-latency low-jitter class, class B be a class with predictable latency and jitter, and class C be a best effort transport class. It is worthwhile to note that the RPR ring does not discard frames to resolve congestion. Hence, when a frame has been added onto the ring, even if it is a class C frame, it will eventually arrive at its destination.

The protocol supports destination packet removal so that a packet does not traverse through all of the nodes in the ring thus spatial reuse can be achieved. However, allowing spatial reuse introduces a challenge to ensure fairness among different nodes competing for the ring bandwidth. Thus, a key performance objective of RPR is to simultaneously achieve high utilisation, spatial reuse, and fairness.

MPLS-TP

MPLS-TP is a transport profile of Multi-Protocol Label Switching that is a product of joint effort by IETF (working groups MPLS, CCAMP, and

PWE) and ITU-T (group SG15) [22]. It is a connection oriented technology defined for next generation converged packet transport networks. It supports large variety of services thus it needs to be client and physical layer agnostic.

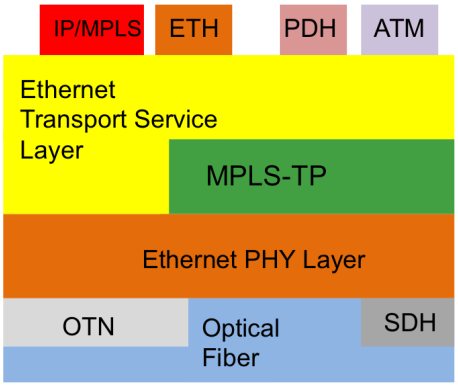

Figure 1.6: MPLS-TP position in Telecommunication Protocol Stack layers

Fig1.6 shows the position of MPLS-TP in telecommunication network protocol stack. The key roles defined in this technology are the implementation of OAM and resiliency features to ensure the capabilities needed for carrier-grade transport network.

Carrier-grade Ethernet

To enable Ethernet network to be a metro transport technology, there are several services and criteria to be considered such as: Support for TDM clients, Reliability, Scalability and providing QoS. A fully recursive, layered architecture known as M-in-M (MAC in MAC) or Provider Backbone Bridge is defined in 802.1ah. It relies on establishing point to point tunnels via Ethernet nodes, Provider Bridges (PB),that are placed at the edge of the network. PBs encapsulates 802.1q into 802.1ad frame, QinQ, while 802.1q carries the Vlan ID tag, that is related to customer information the 802.1ad tag holds the provider information.

Then MinM encapsulation is introduced to improve the scalability, regarding the MAC address learning issue since still at QinQ level the bridge should learn the customers MAC addresses.

Fig1.7 shows layer based frame encapsulation in PBB Frame.

There is a service Instance Tag, I-Tag that has the informations related to the logical service instance and B-Tag (Backbone-Tag) which together with the MAC address forms the provider tunnel. Then PBB adds another MAC header that is related to ingress PB and is known as Backbone-Source Address (B-SA). In this way the bridges in the core do not know about the customer information. At egress PB the Frame is decapsulated and the Ethernet packet is forwarded to the destination. Since customer and service information are bounded to the tunnel according to [24] it is possible to monitor the connectivity by Connectivity Fault Management

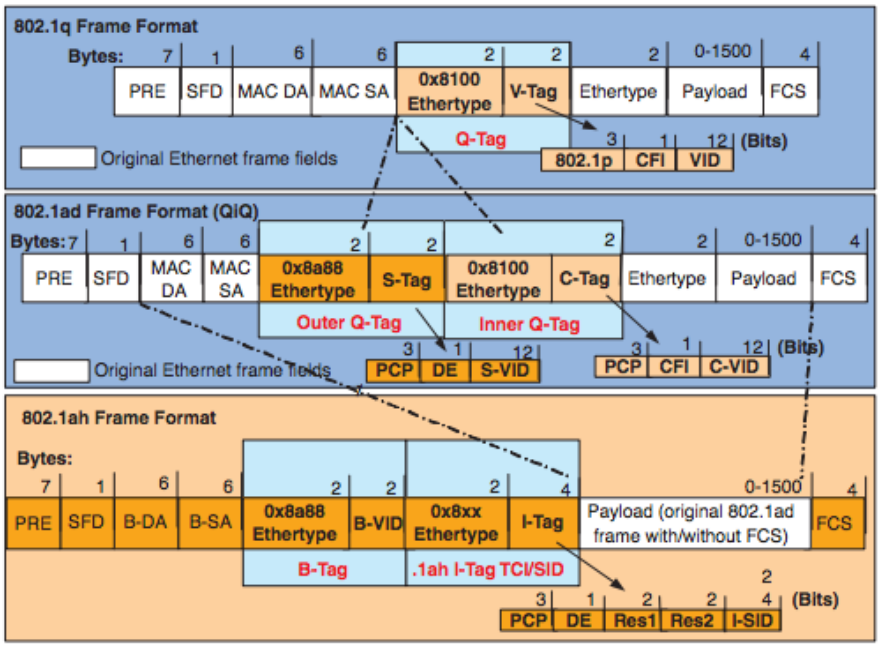

Figure 1.7: Frame formats for 802.1q, 802.1ad (PB), 802.1ah (PBB).(according to [23]

and Continuity Check Messages (CCM). In case of failure CCM permits the endpoints to trigger the backup tunnel.

II.3 Optical Packet Switched Networks

In line with the focus of this research on evolution of optical transport systems in metropolitan area networks, here in the following, there is a short review on a few number of optical-packet ring architectures proposed by several research groups in past few years.

RINGO

RINGO is the early version of the Italian project on optical time slotted packet switched WDM network, was based on unidirectional ring [25]. The first version consists of N nodes and per node one wavelength.
Each node is equipped with an array of tunable transmitter and one fixed–tuned receiver operating on a given wavelength that is allocated to the node.

Packets are transmitted in time slotted manner and all wavelengths are synchronised. To avoid collision the free time slot is sensed by the λ-monitoring method per each time slot. To access the time slots the priority is always with upstream transit packet and the packets to be sent are electronically stored in Virtual Output Queuing (VOQ) structure per destinations.
Later RingO evolved into a more advanced structure in a project in continuation of RingO called WONDER without the previous limitation $N = W$ [26]. The major difference of the latter structure is the separation between

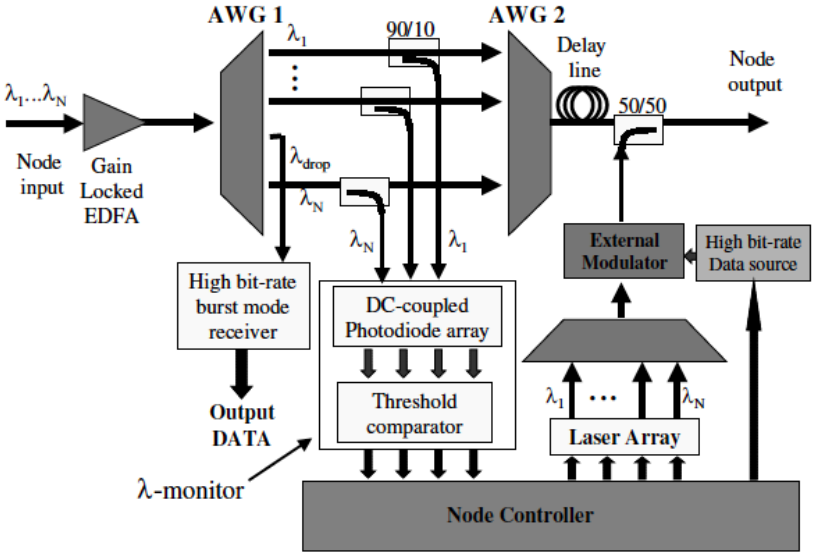

Figure 1.8: RINGO Node Architecture (according to [25])

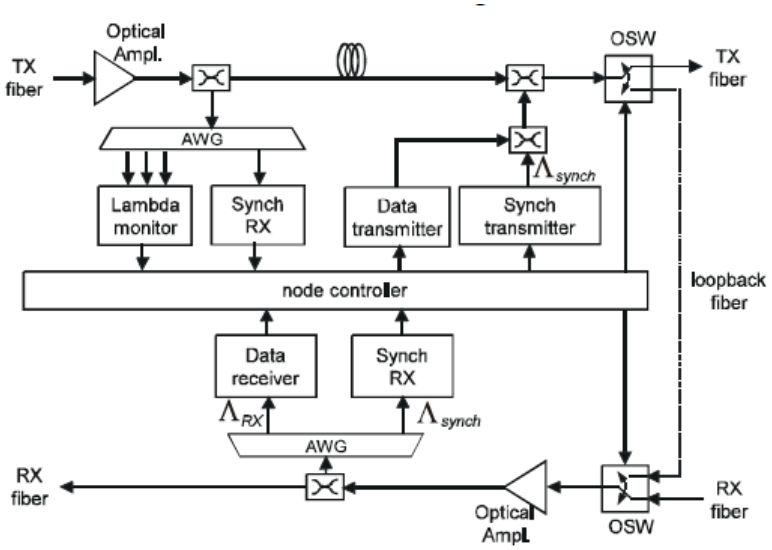

Figure 1.9: WONDER Node Architecture (according to [26])

resources dedicated to transmission from the reception resources, Fig 1.9. Fig1.10 shows that the packet transmissions is on one direction of the bidirectional ring and the receptions happen on the other direction. At some point the two direction of rings are folded, thus the actual topology forms a double bus. The drawback of this method is the loss of the space reuse that if it is possible it can increase significantly the throughput gain of the network.

This architecture offers a bidirectional ring. Therefore a single fault re-

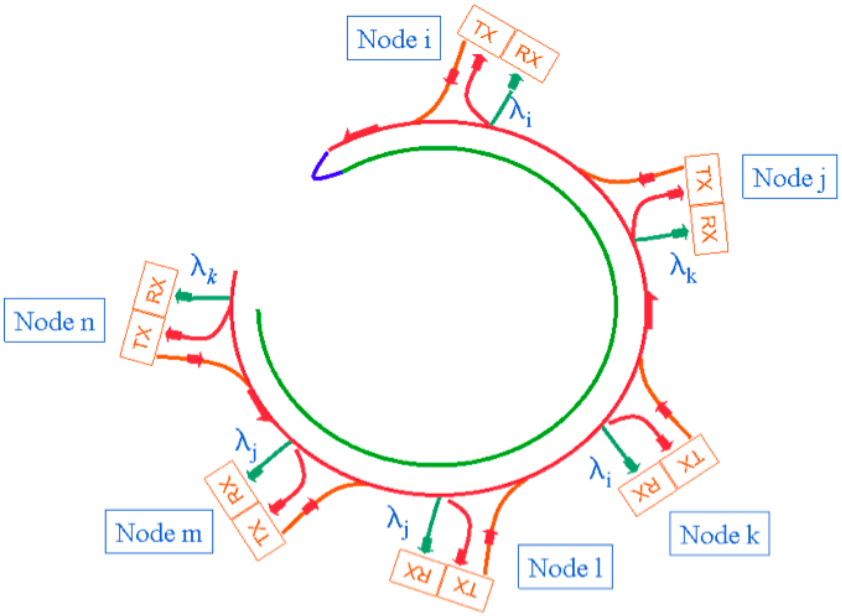

Figure 1.10: Wonder Network (according to [26])

covery is possible. This is simply done by selecting a replacing folding end after detecting the failure to close the loop of the counter rotating rings. The two directions are loops in a node called Master node in the ring. In case of failure, the loop is created in such a way to isolate the failed link and so there is a new master node for each remaining bus network Fig. 1.10.

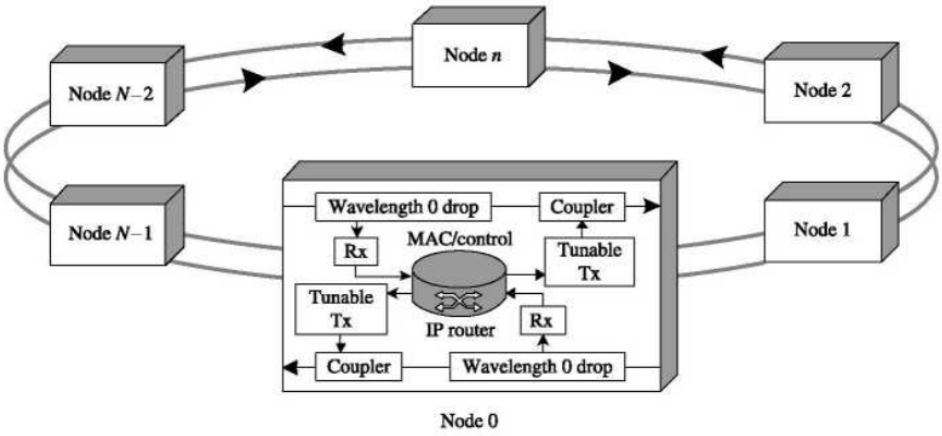

Figure 1.11: HORNET Network Architecture(according to [27])

HORNET

HORNET (Hybrid Optoelectronic Ring NETwork) is another time slotted WDM ring based metro network experimental architecture that evolved from a primarily unidirectional ring architecture with monitoring subcarrier multiplex (SCM) to sense the slot availability at each node to a bidirectional ring with a separate control channel[28]. The two ring carry W wavelengths for data and a separate wavelength as control channel. Each node is equipped with fast tunable transmitter and a fixed tuned receiver.

The MAC protocol for HORNET supports the variable-packet size with a *segmentation and assembly on demand* **access protocol [27]. In this method the transmission of the packet continues till it is complete and if the packet is larger than the time slot then it is marked incomplete and it is segmented and the transmission of the incomplete packet continues in the next empty slot.**
The bidirectional HORNET is fault tolerant against single node/link failure with a distributed protection method. When a failure happens, both end of the cut/link failure detect it. Both send control messages around the ring away from the cut notifying other nodes. Then others simply change the direction of their transmission, and when the failure is repaired, every transmission resumes the original direction.

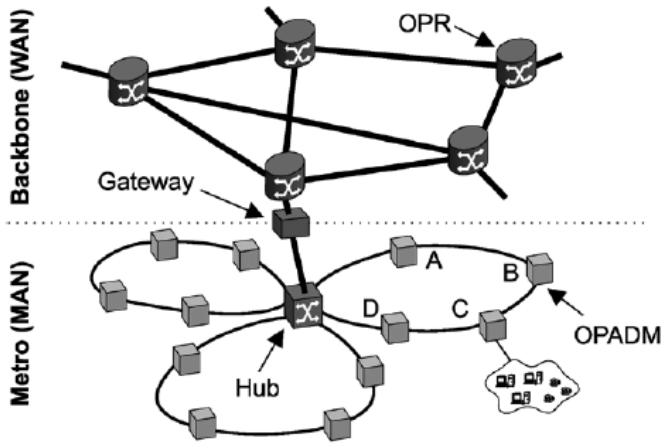

Figure 1.12: DAVID generic network(according to[29])

DAVID

The Data And Voice Integration on DWDM (DAVID) project [29] proposes a practical approach for OPS in Metropolitan area network via multiple physical rings interconnected through a so-called hub Fig 1.12. A ring will comprise one or more fibers, each operated in dense-wavelength-division- multiplexing (DWDM) regime. One wavelength is deployed as a dedicated control channel, while other wavelengths carry the actual data in the form of fixed-length packets. It is synchronous fixed time slotted

network. The ring architecture thus uses both wavelength- division multiple access (WDMA) and time-division multiple access (TDMA). The optical packet add/drop multiplexer (OPADM) ring node puts optical packets on the ring, using a MAC protocol to decide which time slot at which wavelength to use therefore contention is avoided and the need for buffering on the optical path within the MAN is eliminated.
At each node it is only possible to receive or transmit on one channel at the time.
The hub is buffer-less and it is connecting the interconnection point of multiple rings and provides access toward the wide-area network (WAN) through a gateway. This WAN connection from a logical point of view can be seen as an extra ring to and from which to switch traffic. The gateway will be responsible for solving contention between packet flows between MAN and WAN. The latter consists of optical packet routers (OPRs) interconnected in a meshed topology. In contrast to the MAN, an OPR in the WAN may exploit optical buffers in the form of fiber delay lines (FDLs) to aid in contention resolution.

Optical Light-trail

An optical light-trail is a unidirectional optical bus between a source and a destination at the two ends of the light-trail providing sub-wavelength granularity [30]. A wavelength is shared between all the nodes on the way form source till destination. This means that the intermediate nodes can transmit and share the bandwidth by adding/dropping traffic. At each intermediate node a portion of optical power is tapped off in order to detect the optical signal for routing, and add/drop purpose. If the traffic is destined to the node it is received otherwise it is passed through. In case the channel is detected empty available ready to be sent traffic is added to the light-trail.

Fig1.13 shows the node architecture and the network structure more in details. A signalling mechanism is used to ensure conflict-free communications between the nodes in the network.

Such a light-trail architecture is based on mature optical technologies. It improves wavelength utilisation without using optical buffers, fast optical switching and O-E-O conversions at the intermediate nodes. Due to the power loss caused by splitting at each hop, the length of a light-trail is limited and is expected to be not more than 5 hops.
Thus it is a technology considered more for WDM metro/access networks. The Extended version added other direction to the Light-trail to have a bidirectional optical bus. In the bidirectional light-trail all the nodes have the knowledge of their location in the network by set of set-up link acknowledge signal, thus the data transmission is based on the location of the source -destination is located on one of the directions.

OPST

Optical Packet Switch and Transport is a new networking platform introduced by Intune Networks [32]. The control plane runs internally inside a

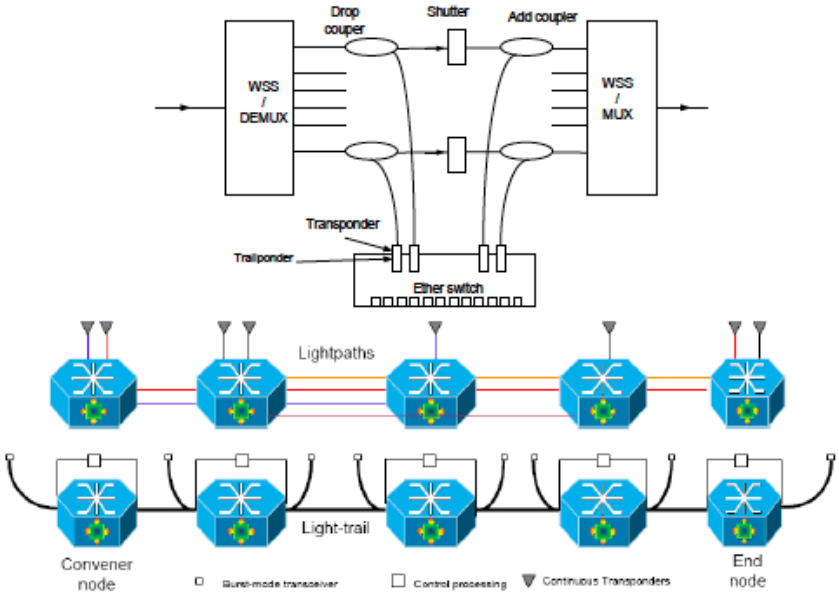

Figure 1.13: Light Trail Network Architecture (according to [31])

ring network, transforming the entire ring into a distributed switch that operates as a single new network element Fig 1.14. It consists of a dual ring with a MAC based on Carrier Sense Media Access with Collision Avoidance (CSMA-CA). Thus OPST employs an asynchronous access method that does not need a whole network synchronisation. Each node contain a fast tunable laser capable of tuning between wavelengths at nano-second speeds.

Each node also contains a fixed wavelength filter which is the wavelength-routed address of the ports of the system. Basic operation involves reading incoming packet addresses, translating them into wavelengths and queuing them in virtual output queues. A scheduler forms bursts from the queues of packets and modulates a burst onto a tunable laser transponder whose wavelength is rapidly tuned to the destination wavelength. Then the packet is sent out on the ring while a distributed scheduling system ensures fair access onto the ring. This is implemented using a wavelength selective switch typically used in ROADMs.

The filter is set once at the installation of the system. Each tunable transmitter can therefore simultaneously switch its packet flow to a destination as well as transmit to that destination by tuning to the target wavelength. When the transmitters are used on Optical Burst mode, virtual wavelength paths can be set up and pulled down in response to incoming packet flow requirements. The result is an ability to merge packet flows from different sources optically, so that they arrive multiplexed in time at the destination.

MAINS

MAINS (Metro Architecture EnabliNg Subwavelength) is a new network

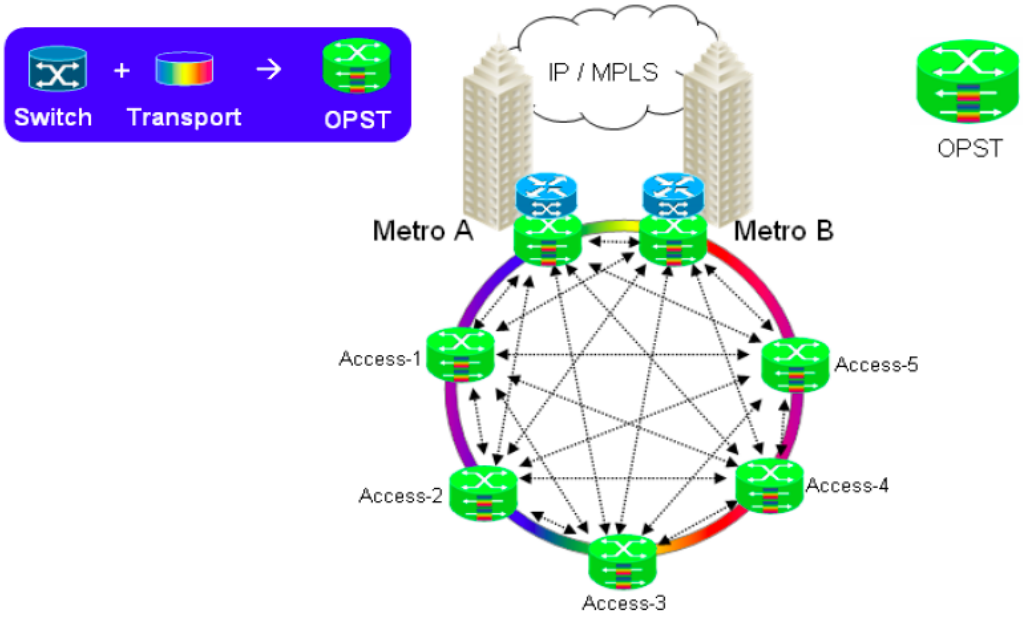

Figure 1.14: OPST Network Architecture (according to [32])

architecture, that is proposed and implemented within the European research project IST FP7. It targets the existing problems in IP network structure within metro network domain [33]. The MAINS is designed

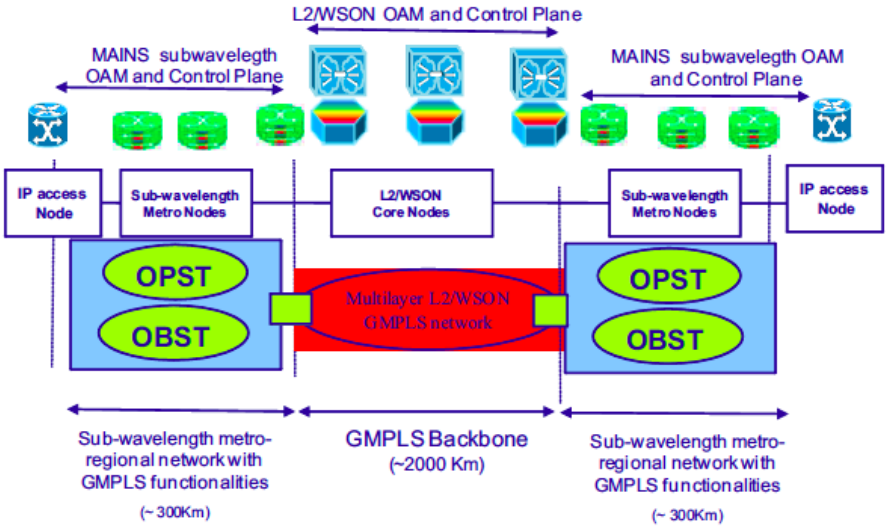

Figure 1.15: MAINS Architecture Concept (according to [33])

completely independent of the access and core technologies. It is relying on a superior and dynamic infrastructure based on optical sub-wavelength transport technologies with enhanced Control Plane capabilities allowing applications and network interworking.

The MAINS transport layer is based on OPST ring mentioned above and

OBST (Optical Burst Switched Transport) mesh sub-wavelength network solutions. Each of the switching technologies have their own internal scheduling, routing and provisioning mechanisms. MAINS addresses the multi-domain interworking aspects of network service transparency across the OPST ring and OBST mesh network solutions Fig1.15.

Generalized Multi-Protocol Label Switching (GMPLS) [34] Control Plane is enhanced to integrate the end-to-end routing and provisioning of sub-wavelength multi-domain interworking between OPST transparent ring domain and OBST transparent mesh domain. In addition to internetwork operation that is a crucial requirement in order to be able to provide end to end network service provision.

Packet-Optical Add Drop Multiplexing Ring

Packet Add/Drop Multiplexing (P-OADM) based technology on WDM ring networks, was initially proposed and studied in [35].

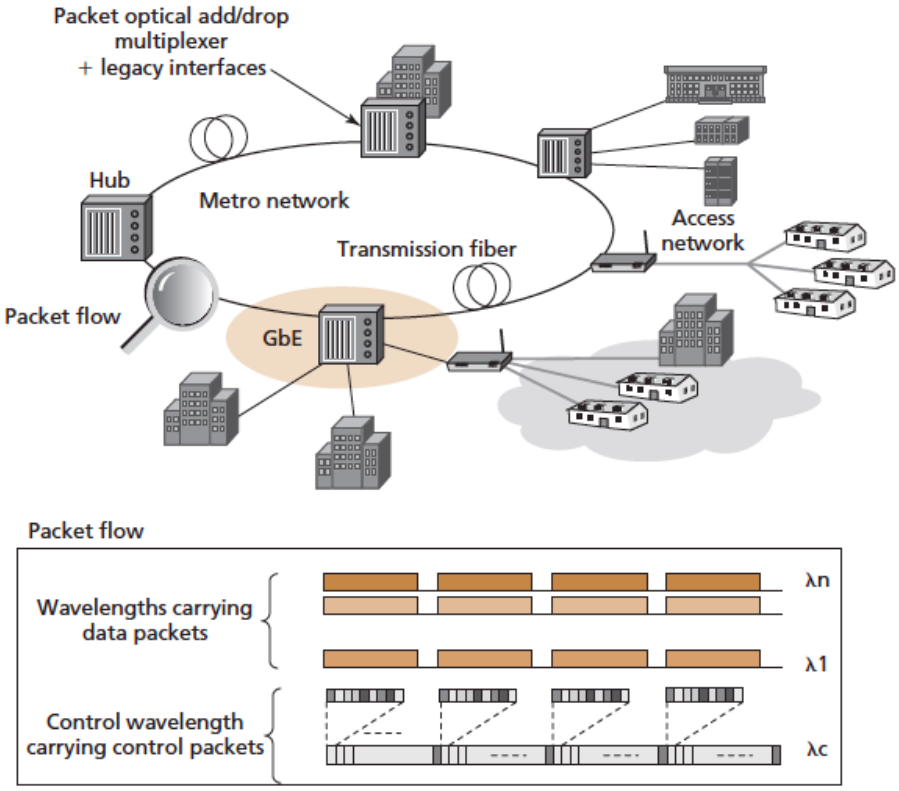

Figure 1.16: Metro Network based on Packet-OADM Ring Model(according to[35])

POADM based nodes are connected by time slotted, WDM, unidirectional ring with a single out-of-band control channel. The data channel is switched all optically. However the control channel packets experiences O/E/O conversion at each node in order to be processed, since control information relative to each data channel and to the global ring (e.g. OAM)

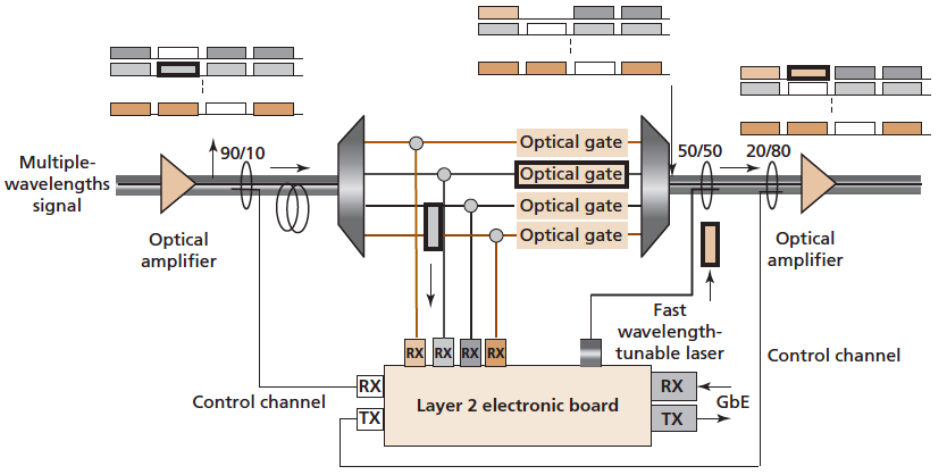

Figure 1.17: Packet Optical Add/Drop Multiplexer structure(according to[35])

is carried within the control channel packets. The data and control channel are synchronised and have fixed size.

Fig1.17 shows the node structure equipped with four key components:

- Tunable optical transmitter
- Optical multiplexer and demultiplexer
- SOAs as optical fast gates
- One or multiple receivers

The incoming WDM line is demultiplexed passing through the optical DEMUX.

The SOA gates 'ON' or ' OFF' states determine if the packets passthrough or are suppressed. The SOA gate states are reconfigured according to the content of control packet. Only the packets that are destined to the node are received and processed, and the rest cross the gates are then multiplexed into the exit fiber all optically. Each node can receive on one or several wavelength if it has one fixed wavelength receiver or an array of fixed wavelength receivers. Only one packet can be added by fast tunable transmitter at any wavelength if there is an empty time slot.

The POADM ring provides optical transparency and sub-wavelength granularity that makes it an attractive solution for metropolitan area network.

Time-Domain Wavelength Interleaved Networking

Time-Domain Wavelength Interleaved Networking (TWIN) is an optical transport network architecture that offers an all optical sub-wavelength granularity without a need for intermediate O/E/O conversion and buffering or optical fast switching [37]. Fig1.18 shows the TWIN network architecture.

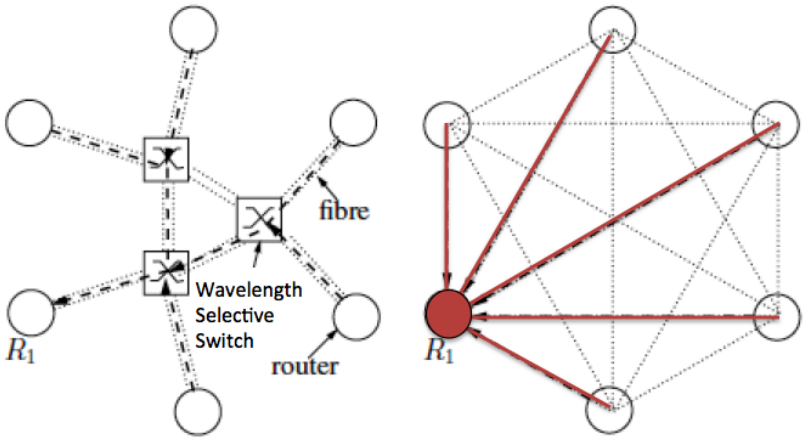

Figure 1.18: TWIN Network structure(according to[36])

Each node is equipped with a fast tunable laser and has a certain wavelength to receive traffic.
Therefore to support any arbitrary topology, wavelength selective switches and passive combiners are need at core of the network. As a result a tree like, multipoint to point overlay topology is formed per each destination. Then to avoid packet collision and keep the core passive, packet scheduling is vital at each source node, knowing that at each time a node as a source can only transmit on one wavelength [38] [39]. Since TWIN is transport network technology no node sends traffic before making sure that it is received. The traffic maybe blocked due to the lack of resources at source, however it should not be dropped in the middle or at destination. Thus there is a communication between source and destination to ask for grand and to send the traffic. Accordingly granting the permission to source nodes can be given in distributed or in centralised way. It is shown in [40] that centralised scheduling offers higher performance.

Conclusion

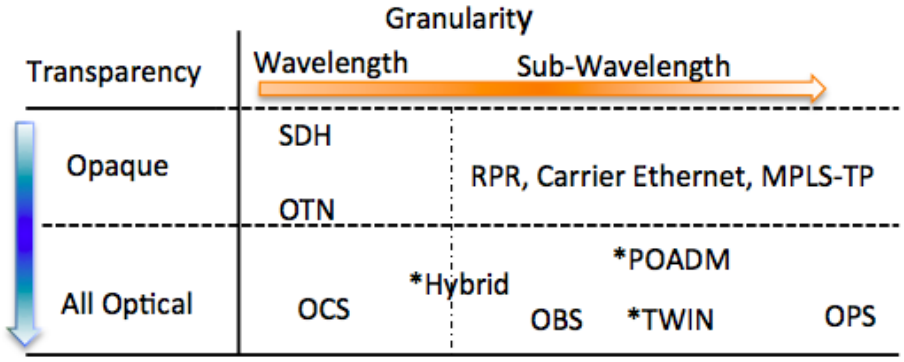

Figure 1.19: Granularity versus Transparency in Optical Transport Technologies

In this chapter we reviewed a few optical transport systems in the scope of metropolitan area networks with two focal points in perspective: granularity and optical transparency. Fig 1.19 offer a simple map which along its vertical axis moving down transparency degree grows, and on horizontal axis towards right we obtain higher degree of granularity. Therefore as it is shown in the Fig1.19 under the category of sub-wavelength granular transparent technologies the ideal OPS is the ultimate point in this two dimensional map. Then OBS technology may differ in the level of granularity depending on the implementation and burst assembly method, yet it still is as transparent as OPS. The POADM based technology provides full transparency within a single ring thus scaling the network may affect the transparency degree. It has been shown in [41] that it is scalable up to ten nodes. The granularity level is sub-wavelength yet larger than a single Ethernet packet. With the out of band control channel it provides all the necessary requirement of the metropolitan area network transport network. Therefore we chose this solution as the basis of the most of the work in this thesis from chapter II to chapter III and part of the chapter IV. Finally TWIN, that offers end to end transparency regardless of the scale of the network and maintain sub-wavelength granularity is another chose of this study in the second part of chapter IV.

Bibliography

[1] B. Ramamurthy, H. Feng, D. Datta, J.P. Heritage, and B. Mukherjee. Transparent vs. opaque vs. translucent wavelength-routed optical networks. In *Optical Fiber Communication Conference, 1999, and the International Conference on Integrated Optics and Optical Fiber Communication. OFC/IOOC '99. Technical Digest*, volume 1, pages 59–61 vol.1, 1999.

[2] Tarek S.El-Bawab. *Optical Switching*. Springer, 2006.

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

[4] Shun Yao, S.J.B. Yoo, B. Mukherjee, and S. Dixit. All-optical packet switching for metropolitan area networks: opportunities and challenges. *Communications Magazine, IEEE*, 39(3):142–148, 2001.

[5] Myungsik Yoo and Chunming Qiao. Just-enough-time (jet): a high speed protocol for bursty traffic in optical networks. In *Vertical-Cavity Lasers, Technologies for a Global Information Infrastructure, WDM Components Technology, Advanced Semiconductor Lasers and Applications, Gallium Nitride Materials, Processing, and Devi*, pages 26–27, 1997.

[6] Jian Wu Jin tong Lin Zhou Lan, Hong xiang Guo. Performance of distributed control protocol for wr-obs systems. In *Communications, 2004 and the 5th International Symposium on Multi-Dimensional Mobile Communications Proceedings. The 2004 Joint Conference of the 10th Asia-Pacific Conference on*, volume 2, pages 522–526 vol.2, 2004.

[7] C. Qiao M. Yoo. Optical burst switching (obs)–a new paradigm for an optical internet. *High Speed Networks, Journal of*, 8(1):69–84, 1999.

[8] G. Zervas, M. De Leenheer, L. Sadeghioon, Dimitris Klonidis, Y. Qin, R. Nejabati, D. Simeonidou, C. Develder, B. Dhoedt, and P. Demeester. Multi-granular optical cross-connect: Design, analysis, and demonstration. *Optical Communications and Networking, IEEE/OSA Journal of*, 1(1):69–84, 2009.

[9] A. Sahara, Y. Tsukishima, K. Shimano, M. Koga, K. Mori, Y. Sakai, Y. Ishii, and M. Kawai. Demonstration of connection-oriented optical burst switching network utilising plc and mems switches. *Electronics Letters*, 40(25):1597–1599, 2004.

[10] Qirui Huang, Yong-Kee Yeo, and Luying Zhou. Optical burst-over-circuit switching for multi-granularity traffic in data centers. In *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, pages 1–3, 2013.

[11] K. Okamoto, H. Okazaki, Y. Ohmori, and K. Kato. Fabrication of large scale integrated-optic n*n star couplers. *Photonics Technology Letters, IEEE*, 4(9):1032–1035, 1992.

[12] H. Takahashi, K. Oda, H. Toba, and Y. Inoue. Transmission characteristics of arrayed waveguide n times;n wavelength multiplexer. *Lightwave Technology, Journal of*, 13(3):447–455, 1995.

[13] M.C. Wu, O. Solgaard, and J.E. Ford. Optical mems for lightwave communication. *Lightwave Technology, Journal of*, 24(12):4433–4454, 2006.

[14] K. Sorimoto, H. Tsuda, H. Ishikawa, T. Hasama, H. Kawashima, K. Kintaka, M. Mori, and H. Uetsuka. A compact high-port-count wavelength selective switch using lcoss and a multi-stacked awg. In *IEEE Lasers and Electro-Optics Society, 2008. LEOS 2008. 21st Annual Meeting of the*, pages 376–377, 2008.

[15] G. van den Hoven. Semiconductor optical amplifiers for digital and analog communication. In *Optical Fiber Communication Conference and Exhibit, 1998. OFC '98., Technical Digest*, pages 40–41, 1998.

[16] Alan E.Willner Ivan P.Kaminow, Tingye Li. *Optical Fiber Telecommunications VIB: Systems and Networks*. Elsevier Science Ltd, 2013.

[17] P.M. Anandarajah, A. Kaszubowska, R. Maher, and L.P. Barry. Wavelength tunable lasers in future optical communication systems. In *Transparent Optical Networks, 2008. ICTON 2008. 10th Anniversary International Conference on*, volume 2, pages 109–109, 2008.

[18] Standardization trend and supporting technologies of optical transport network (otn). In *OptoeElectronics and Communications Conference (OECC), 2010 15th*, pages 1–28, 2010.

[19] Qiong Wang and Gao Ying. Otn for the future transmission network. In *Photonics and Optoelectronics (SOPO), 2012 Symposium on*, pages 1–4, 2012.

[20] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun. Ieee 802.17 resilient packet ring tutorial. *Communications Magazine, IEEE*, 42(3):112–118, 2004.

[21] Charlie B. Kawwas and M.R. Soleymani. Performance enhancements analysis in next generation resilient packet ring networks. In *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, pages 738–741, 2006.

[22] Y. Koike. Mpls-tp: Overview and status. In *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, pages 1–45, 2013.

[23] R.C. Sofia. A survey of advanced ethernet forwarding approaches. *Communications Surveys Tutorials, IEEE*, 11(1):92–115, 2009.

[24] IEEE. 802.1ag - connectivity fault management, draft 8. 2007.

[25] A. Carena, Vito De Feo, J.M. Finochietto, R. Gaudino, F. Neri, C. Piglione, and P. Poggiolini. Ringo: an experimental wdm optical packet network for metro applications. *Selected Areas in Communications, IEEE Journal on*, 22(8):1561–1571, 2004.

[26] S. Bregni, D. Carzaniga, R. Gaudino, and A. Pattavina. Slot synchronization of wdm packet-switched slotted rings: the wonder project. In *Communications, 2006. ICC '06. IEEE International Conference on*, volume 6, pages 2556–2561, 2006.

[27] S.M. Gemelos, I.M. White, D. Wonglumsom, K. Shrikhande, T. One, and L.G. Kazovsky. Wdm metropolitan area network based on csma/ca packet switching. *Photonics Technology Letters, IEEE*, 11(11):1512–1514, 1999.

[28] K.V. Shrikhande, I.M. White, D. Wonglumsom, S.M. Gemelos, M.S. Rogge, Y. Fukashiro, M. Avenarius, and Leonid G. Kazovsky. Hornet: a packet-over-wdm multiple access metropolitan area ring network. *Selected Areas in Communications, IEEE Journal on*, 18(10):2004–2016, 2000.

[29] L. Dittmann, C. Develder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Sole-Pareta, W. Cerroni, N. Leligou, L. Dembeck, B. Mortensen, M. Pickavet, N. Le Sauze, M. Mahony, B. 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.

[30] M.T. Frederick, N.A. VanderHorn, and A.K. Somani. Light trails: a subwavelength solution for optical networking. In *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, pages 175–179, 2004.

[31] A. Gumaste, J. Chandarana, P. Bafna, N. Ghani, and V. Sharma. On control plane for service provisioning in light-trail wdm optical ring networks. In *Communications, 2007. ICC '07. IEEE International Conference on*, pages 2442–2449, 2007.

[32] J. Dunne, T. Farrell, and J. Shields. Optical packet switch and transport: A new metro platform to reduce costs and power by 50performance levels. In *Transparent Optical Networks, 2009. ICTON '09. 11th International Conference on*, pages 1–5, 2009.

[33] J. Fernandez-Palacios, N. Gutierrez, G. Carrozzo, G. Bernini, J. Aracil, V. Lopez, G. Zervas, R. Nejabati, D. Simeonidou, M. Basham, and D. Christofi. Metro architectures enabling subwavelengths: Rationale and technical challenges. In *Future Network and Mobile Summit, 2010*, pages 1–8, 2010.

[34] E.Mannie. Generalized multi-protocol label switching (gmpls) architecture. *RFC 3945*, 2004.

[35] D. Chiaroni, G. Buforn, C. Simonneau, S. Etienne, and J-C Antona. Optical packet add/drop systems. In *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pages 1–3, 2010.

[36] P. Robert and J. Roberts. A flow-aware mac protocol for a passive optical metropolitan area network. In *Teletraffic Congress (ITC), 2011 23rd International*, pages 166–173, 2011.

[37] I. Widjaja, I. Saniee, R. Giles, and D. 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.

[38] K. Ross, N. Bambos, K. Kumaran, I. Saniee, and I. Widjaja. Scheduling bursts in time-domain wavelength interleaved networks. *Selected Areas in Communications, IEEE Journal on*, 21(9):1441–1451, 2003.

[39] 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.–, 2005.

[40] A. Triki, P. Gavignet, B. Arzur, E.L. Rouzic, and A. Gravey. Efficient control plane for passive optical burst switching network. In *Information Networking (ICOIN), 2013 International Conference on*, pages 535–540, 2013.

[41] D. Chiaroni, P. Guignard, P. Faccin, E. Grard, G. Tartarini, J. Gripp, S. Etienne, L. Guillo, A. Pizzinat, B. Charbonnier, V. Rodrigues, E. Ortego, and M. Popov. End-to-end integrated multiservice packet network. In *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2012 and the National Fiber Optic Engineers Conference*, pages 1–3, 2012.

Figures and tables

Figures

Chapter 2

Multi-Service Medium Access Control For Optical Packet Switched Metropolitan Area Network

Contents

Introduction

In this chapter, we introduce a Multi-Service MAC for an all optical packet switched metropolitan area network that is based on POADM technology. The contributions within this chapter are divided in five parts:
The first part features the underlaying principles of the optical time slotted packet switched network that is the basis of the study throughout this thesis.

The second part deals with the general access scheme of the MAC protocol suitable for the above mentioned network. The adaptation layer is defined in such a way that makes the network entirely transparent to the client layer protocols and technologies, yet it keeps the interoperability between access and metro network with multiple services. Then we have adapted the label based functionality in two levels in order to identify and integrate multiple services within the MAC layer. This novel approach provides the least control information processing at each node as well as deploying the conventional label usage.

In the third part we define the insertion-extraction methods and functionality of the defined labels for the proposed multi-service MAC protocol.

In the fourth part we propose an original reservation-based mechanism of best effort traffic insertion-extraction. Then we model and simulate the method for performance evaluation.

We concluded this chapter with the fifth part where we derive the essential content of control message per time slot for POADM based network. This control message with the defined features enables the POADM based networks to incorporate multiple services within the MAC layer without the need for OEO conversion in data plane.

I Packet OADMS for Metropolitan Area Network

I.1 Network Concept

As it is mentioned in chapter I, POADMS ring combines the advanced optical technology with the strength of electronics to address issues such as aggregation efficiency and sub-wavelength capacity. This is achieved by providing packet level granularity and deploying ultra-fast optical components, in addition to optical transparency using ROADMS technology.
The choice of these technologies provide flexibility and facilitates traffic management. The transient traffic remains in optical domain and there is no need to perform OEO conversion except at source and destination. Thus the node capacity is wholly decoupled from the network capacity.
We have extended the above mentioned technology that was originally defined as unidirectional ring to a bidirectional ring.

Thus each node per direction has one tunable transmitter, and an array of fixed WDM receivers. Moreover there is a separate out of band control channel

synchronized with the data channel per ring per direction. The data channel maintains the main rule as that the transit packet has always priority, it is never dropped.

II MAC Structure

The network under study uses fiber as a shared medium; therefore we need a medium access control(MAC) protocol to regulate the access to this shared capacity. Existing MACs for optical packet rings often rely on dedicating wavelengths either per destination or per source; in other cases they rely on specific hub nodes that control the access to the medium.
The Packet-OADM ring technology as the technology under study in this thesis shares the wavelengths between all nodes and does not rely on Hubs. Therefore there is a need to define a proper MAC layer considering the next generation metro network needs:

- Multiple levels of granularity
- Multi-protocol and multi-class
- Unicast and multicast
- Transport network level grade of service, e.g. low latency, 99.999 availability, etc.

II.1 Adaptation Layer

To provide a full client service transparency the MAC layer is divided into two sub-layers: adaptation, transport. Client data that is carried towards the edge nodes within different network and transmission protocols such as IP/ Ethernet, ATM, SONET/SDH, etc. has different bit-rates, QOS demands, addressing and encapsulation frame formats.

Adaptation sub-layer is in charge of matching the cross protocol differences and transfer various aspects of service specifications from the client original protocol to the Optical packet transport MAC layer Fig2.1. Thus receiving and preparation tasks of the client packets for transmission as fixed optical packets on the packet OADM network in the adaptation sub-layer are presented in the following:

- Service Data Unit Client data packets that is arrived at the edge node is received by structure data unit and the addressing and control information and is extracted. Then the fixed size burst is formed by aggregating and encapsulating the client data at SDU level. All information extracted earlier from the original client packet such as QoS indicator, packet and flow id, segmentation and reassembly (SAR) information are included in the SDU level header along with scheduling, SDU burst length and SDU level label.
  The label carried by an SDU can implicitly identify the client layer, or e.g. the Virtual Private Network (VPN) to which the client packet belongs, or the multicast flow to which this packet belongs. Thus intermediate burst

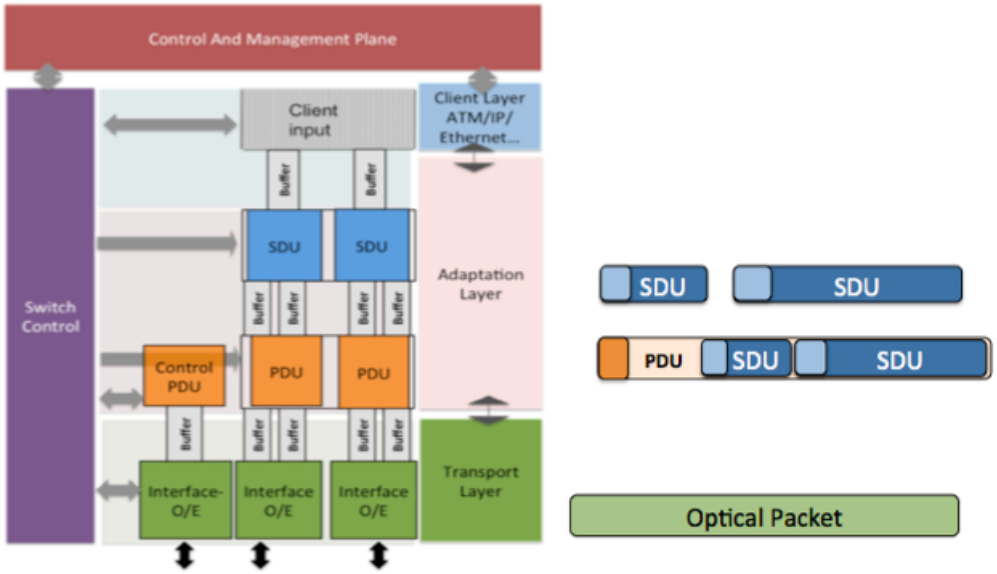

Figure 2.1: POADM MAC layer structure and Packet encapsulation

level grooming functionality is possible at transient HUB/ring interconnection point in inter-ring transient traffic flow case. This will be more detailed in the following sections in this chapter.

- **Protocol Data Unit The fixed size burst is then prepared to be converted to an optical burst at Protocol Data Unit section. At this level information regarding the physical layer, framing synchronizing, PDU Length, and Error correction are added to the PDU header. Then the burst is ready to be converted to optical burst in order to be transmitted.**

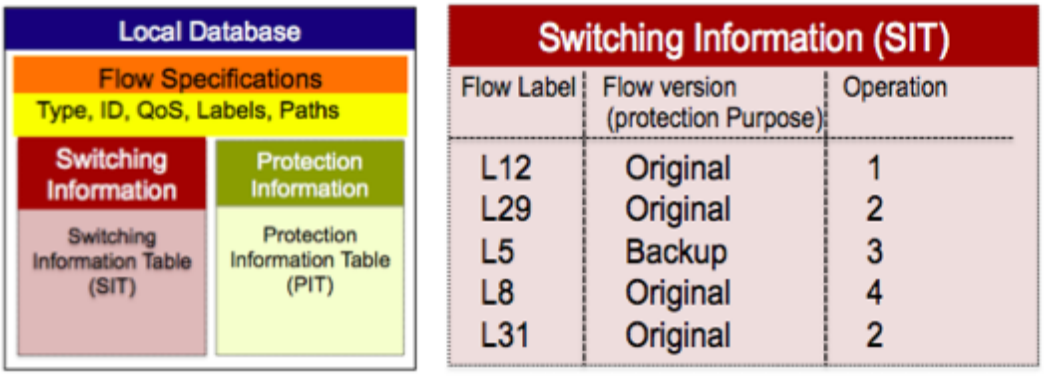

Figure 2.2: Local Information Data base in a Node within the POADM ring and detailed Switching Information Table

II.2 Label based Access for Multi-services Flow

Since all the nodes can potentially receive traffic on all the wavelengths, we need a mean to recognise each flow per time-slot and per wavelength via the information carried by the control packet.

We adopted a label switching concept, similar to the one used in MPLS, as it is the most current and efficient technology for packet switched networks[2], however we use label concept in a new two level approach.

A node identifies which PDUs to receive and extract using labels in the control packet characterising each PDU carried in a given time-slot. It is thus not necessary to carry an explicit destination address.

Moreover we have defined SDU level labels. This facilitates the packet grooming, meaning filling PDUs with different SDUs from different flows at the interconnection nodes between the rings. This way the labels facilitate inter-ring routing, and multi-cast services where O/E/O may be a necessity.

Fig.2.2 shows the node local database containing three main block of information crucial for operation: The flow specifications such as flow identifications, type, QoS and the path including all the nodes that the flow passes till it reaches the destination. The Switching Information Table, SIT that contains the mapping of the originating node ID,the label, the protection type and flow version that is a very important part of the information in case of failure, since it enables the node to react on the fly with minimum processing need and finally the action that the node does per time slot. The Protection Information Table contains details of the flow paths sorted per label and node ID and it is used in case of failure for locating the failure. This part and its functionality is explained more in the next chapter.

These tables are updated if there is a new entry or a change within the network by a central or distributed control/management plane.

There are four general types of actions that usually one of three occurs after reading the control packet per time slot.

1. Extract-Erase: The node receives the PDU and strips it off.
2. Extract-Pass: The node copies the PDU and does not erase it.
3. Erase: The node erases the PDU without receiving it.
4. Extract-Insert: The node extract the PDU and modifies the label stack (e.g pops the outer label), and retransmits the PDU into the network.

The first three actions occur when the station is either a source or a destination for the flow within a single ring. The last one occurs when there are two or more nodes interconnected we refer to the interconnecting node as Hub.

Ring interconnection relies on hubs (interconnecting nodes/ points) with packet OADM interfaces on two different optical packet rings. Inter-ring traffic corresponds to optical PDUs that a hub receives on one ring and inserts on the other. Here we consider packet level inter-connection as a hybrid optoelectronics structure for the Hub with shared buffers in both electronic and optics domain for synchronization purpose and if necessary bit rate, format, or wavelength conversion in addition to contention resolution. The hybrid all optically

circuit/packet ring inter-connection is introduced in chapter 4. When handling inter-ring traffic, a hub applies procedures corresponding to the PDU's service type:

- **Insertion policy (opportunistic for guaranteed traffic or reservation based for BE traffic).**
- **Erasure policy, which is dictated by traffic mode (unicast vs. multicast).**

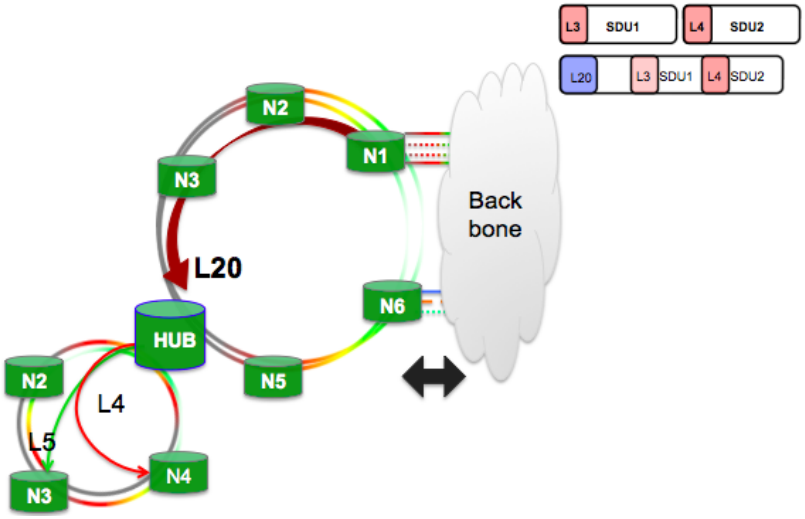

Figure 2.3: Inter-Ring traffic grooming benefiting SDU and PDU layers Labels

This is illustrated on a simple example in Fig.2.3. Two flows are sent from the backbone to respectively N3 and N4 located on the side ring. Regarding the size of the packets and that they have to pass the Hub node to reach to their destinations therefore they can N1 groomed them into one PDU with an outer or Level L20 as if their destination is Hub. The original destinations information is kept as SDU level labels, L3, L4. Thus they pass the nodes in the main ring till the Hub all optically. At Hub as they are extracted and reinserted into the side ring with their SDU labels now as their outer PDU level. This method creates a flexible grooming tunnel in a multi-ring topology.

III Unicast, Multicast MAC Operation

This section explains the insertion-extraction methods used to transmit-receive different types of flows and services between the nodes within the POADM ring. The methods vary depending on the flow types and their services. As mentioned earlier the information regarding the type of flow and service is mapped into the control packet fields, thus each node knows which operation to fetch according to the control packet contents per time slot. In the following subsections insertion/extraction operations of Unicast, and Multicast flows are explained in details.

III.1 Unicast

The focus here is on the insertion-extraction process of the guaranteed unicast traffic flow. This means that the network is dimensioned based on a given service level agreement profile for a given traffic matrix. In addition the unicast flows are considered to be original flows not backups. The insertion-extraction process of the backup unicast flows depends on the protection method in use, which is thoroughly explained in chapter III.

- Insertion
  As it is mentioned in chapter I, the insertion method proposed for unicast is opportunistic based on simple "empty slot" policy.

  A PDU carrying guaranteed traffic is transmitted as soon as an empty time slot on a suitable wavelength is available. The assumption here is that the original flow chooses the shortest distance direction on the bidirectional ring in normal operation mode,(e.i. no failure). Since the amount of guaranteed traffic is known, a given node cannot starve or even degrade the QoS offered to the other nodes as long as the transmitted traffic conforms to the traffic matrix (i.e. negotiated SLA).

- Extraction
  Extraction is performed by reading the top labels in the label stacks carried in the control packet. This allows a node to identify which of the following operations should take place:

  - Receive and Erase or a unicast PDU, if the node is destination the action is to receive and erase. This is the traditional “destination stripping” strategy that allows spatial reuse and thus increases network capacity.
  - Optical Pass The flow is on transit, therefore no action is necessary.

III.2 Multicast

Since multicast traffic is offered from one or multiple points and it is usually received at more than one destination, we use Multicast Service Point, MSP, as a general term for where the multicast traffic is inserted into the network and Multicast Drop off point, MuD for the point where we take off the multicast flow. Both of these points are assumed to be located on the POAMD rings Fig2.4. The multicast extraction method is a general *'Receive and Pass'* method, where the node just copies the flow and then the flow continues its path all optically within the ring.

In the following, we introduce *unidirectional* and *bidirectional* multicast services that are different in insertion and drop off methods:

Unidirectional Multicast Service:
The MSP inserts the multicast flow in one direction of the ring on a wavelength. The flow makes a full trip in the ring and goes back to the MSP, then source strip method is used to drop the flow off the ring, therefore in this case MSP and MuD are the same point on the ring that is also called source-stripped point.

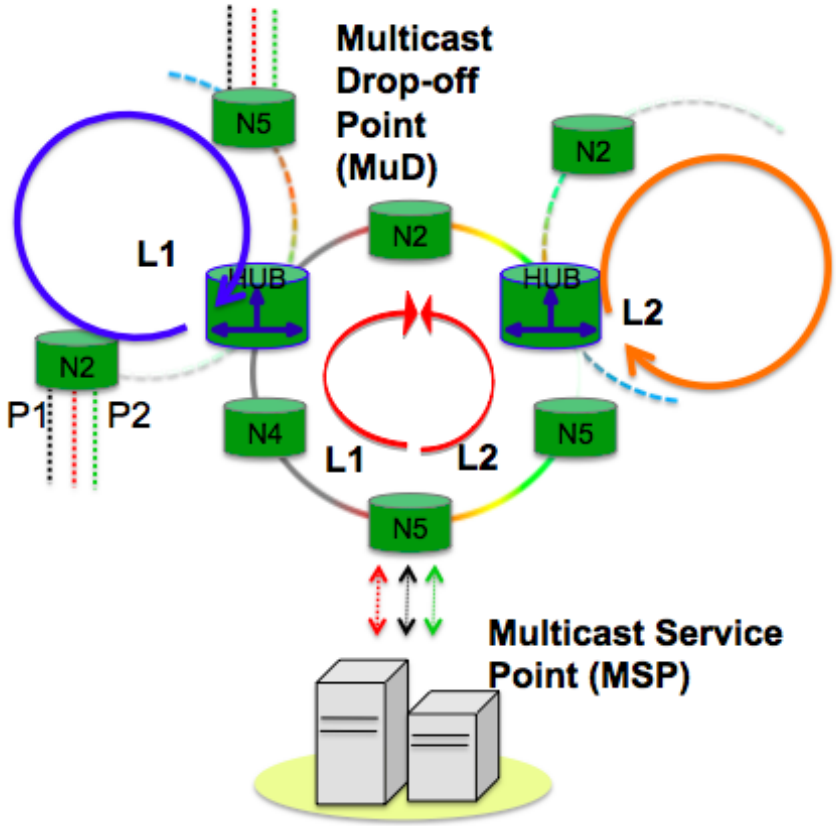

Figure 2.4: Multicast service on Packet-OADM multi-Ring deploying Multicast Drop off Point (MUD) method in the main ring and the source-stripped point in the side rings

Bidirectional Multicast Service:

The MSP inserts the multicast flow in both directions of the POADM ring with the same labels in the control packets associated to each direction. The drop off point in this approach is calculated to be in the middle of the ring where there is approximately equal distance on both directions from MSP to MuD. At Mud the the flows one flow is *'Received and Erased'* **the other is just** *'Erased/Dropped off'* **of the path.**

It is clear that the service time is half in the bidirectional approach comparing to the unidirectional one. However finding the exact middle point of a metro ring may not be practical in a real network. In chapter III we have studied other advantages and possible disadvantages in case of failure and protection procedures for the two methods.

IV Multiservice MAC

As mentioned earlier, unicast and multicast insertion-extraction methods were defined for guaranteed type of traffic on a dimensioned network for an static traffic matrix. Here we propose an approach to insert Best Effort (BE) traffic

while ensuring the following properties[4]:

- A transit packet has always the priority over the packet to be inserted at any priority level.
- In a given station, Guaranteed (G) packets always has the priority to be transmitted over BE packets.
- The QoS of the G traffic should still be within the design limits.

IV.1 Best Effort Traffic Access Method

The BE access method is reservation based: This means that to transmit a BE packet, the source has to make a reservation before hand. A simple analogy to make it clearer would be to have a waiting list and priority list, after the priority is served then if there is still capacity left the waiting list is considered to be served.
There is no guarantee to serve the waiting list. Here in this context reservation does not hold resources through out the path in the network. On the contrary it just makes a waiting list as mentioned above. In the control packet there is a field foreseen per time slot to carry the reservations. The BE reservation field carries the associated source-destination information to the reservation field. All the nodes in the network can add or drop reservations under the following conditions:

Add reservation:

- If there is BE traffic to be sent, the node can make one reservation per time slot, the reservation indicates the source and destination of the BE packet.

Drop reservation

- At any intermediate node the reservations can be dropped: To protection the G traffic or to enforce the fair share.
- The node that is the destination to BE reservation can drop the reservation if it considers that it can get congested.

The BE packet insertion operation is detailed in the following steps:

1. Receive any G packets on one or more channels and release corresponding slots.
2. Insert a G packet if possible: A G packet has always the priority and it is transmitted even if the slot is reserved as it was previously discussed.
3. If there is no G packet to be inserted, Insert a BE packet if possible. The BE packet can be inserted only on a slot that carries a reservation for its own flow (same source and same destination). If there is a reservation but there is no BE packet to insert, the reservation is just erased by the station.

IV.2 Performance Analyses

To evaluate the performance of the proposed reservation mechanism, we make an analytical evaluation for two scenarios and then compare with the results we have obtain via NS2 simulator. The simulations supports a reservation based insertion process for BE traffic while using a purely opportunistic insertion process for G traffic.

In the experiments reported below, we consider both the concentration scenario and the any-to-any scenario. The ring is constructed with six stations and three data channels. The G traffic arrival process is Bernoulli. We assume greedy BE sources, which means that each station always has a BE packet to transmit.

The QoS metric for G traffic is the sojourn time in the ingress which includes both the waiting time till a G packet arrives at the head of the queue, and its the insertion time on the ring. The QoS metric for BE traffic is the insertion time (since there is no actual queue for greedy traffic). We first assess the performance of the reservation based scheme for the concentration scenario.

- **Concentration Scenario In the "concentration scenario" we assume that there is a specific station (a Hub) which receives traffic from N_s other stations. The capacity of the egress link from the Hub is equal to N_w.**

  If there is no G traffic ($A = 0$), there is a single constraint that avoids saturation on the busiest link (i.e. the link to the HUB):

$$b_{iHub} \leq \frac{N_w}{N_s} \tag{2.1}$$

  In the general case for symmetric G traffic, let a_{iHub} be the amount of traffic sent by one station to the hub.

  The stability conditions are :

$$a_{iHub} + b_{iHub} \leq \frac{N_w}{N_s} \tag{2.2}$$

  This avoids the congestion of the most saturated link.

$$a_{iHub} \leq \beta[1 - ((N_s - 1)(a_{iHub} + b_{iHub})/N_w)^{N_w}] \tag{2.3}$$

  It ensures that G traffic to be inserted in the last station before the hub is still delivered an acceptable QoS in terms of insertion delay.

  Expression (2.3) is obtained by modelling the insertion queue in station N_s (the last station before the hub) with a Geo/Geo/1 queue where the ingress rate is a_{iHub} and the service time is geometric with parameter $[1 - ((N_s - 1)(a_{iHub} + b_{iHub})/N_w)^{N_w}]$.
  Indeed, the total rate of transit traffic in competition with the G traffic to be inserted in station N_s is $[(N_s - 1)(a_{iHub} + b_{iHub})]$. This rate is balanced on N_w data channels and the slots on all data channels are considered to be independent. Therefore, the probability that a G packet can be inserted in a given time slot is $[1 - ((N_s - 1)(a_{iHub} + b_{iHub})/N_w)^{N_w}]$. β is a parameter

(e.g. 0.9) chosen to limit an upper quantile of the waiting time in the above Geo/Geo/1 queue.

For the concentration case, the maximum amount of acceptable BE traffic is thus given by :

$$b_{iHub} \leq \frac{N_w}{N_s} - a_{iHub} \tag{2.4}$$

$$b_{iHub} \leq N_w(1 - a_{iHub}/\beta)^{1/N_w})/(N_s - 1) - a_{iHub} \tag{2.5}$$

Depending on the value for a_{iHub}, either condition (2.4) or condition (2.5) is the most constraining. It means that protecting G traffic may forbid using all spare resources on the ring.

Using relations (2.2) and (2.3) for $b_{iHub} = 0$ and $\beta = 0.9$, we observe that the system with only G traffic is unstable for $a_{iHub} > 0.55$. Therefore, relevant areas in Fig. 2.5 and Fig. 2.6 are on the left hand side, for a load offered per station (a_{iHub}) smaller than 0.55.

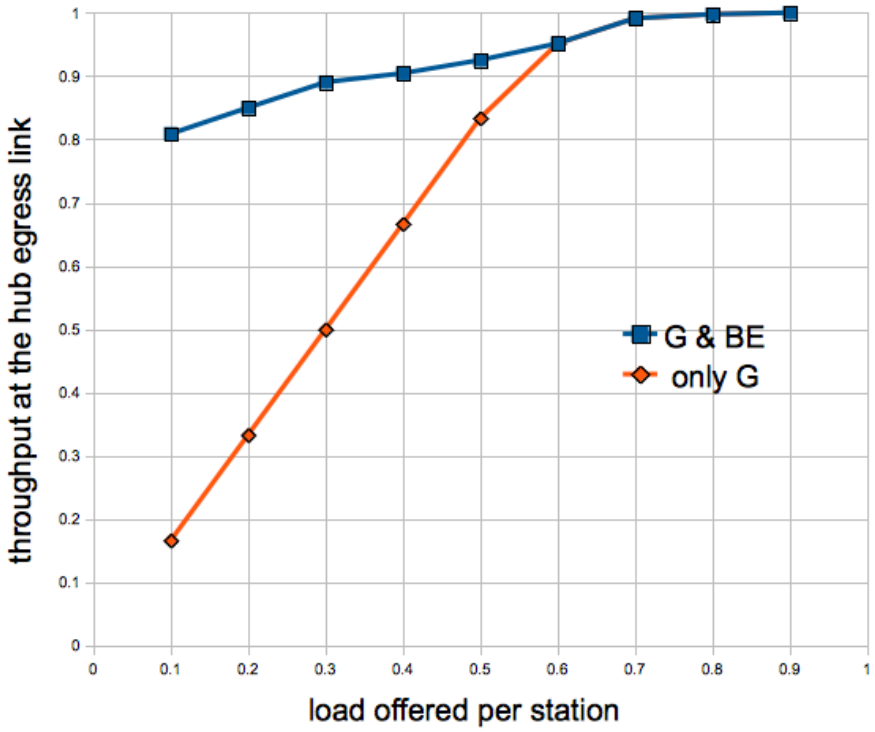

Figure 2.5: Concentration scenario. Hub Throughput versus offered G traffic per station. $N_s = 5$, $N_w = 3$.

Fig. 2.5 shows how the egress traffic from the hub varies versus offered G traffic per station comparing the system without and with BE traffic. It clearly shows that the reservation scheme succeeds indeed in using most of the spare resources to support BE traffic.

It is necessary to check that supporting BE traffic does not degrade the QoS offered to G traffic.

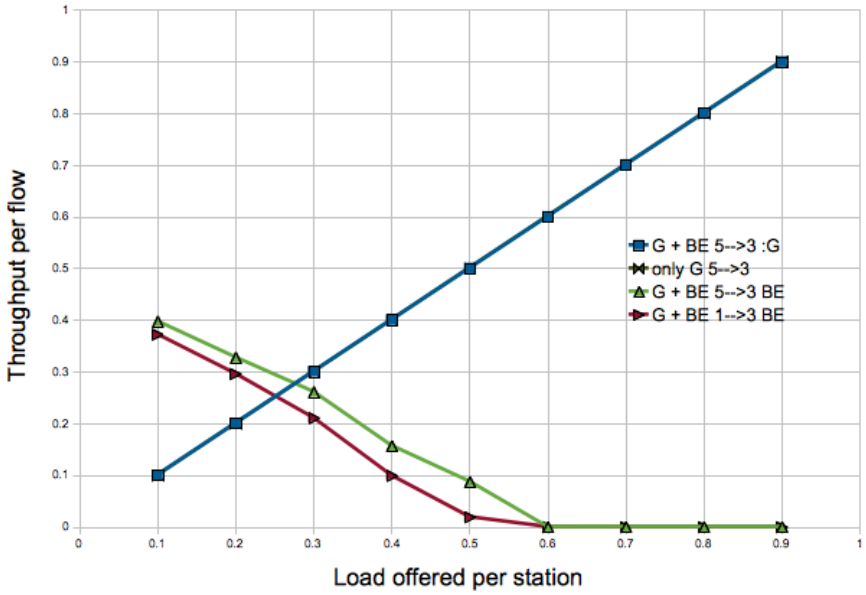

Figure 2.6: Concentration scenario. Achieved ingress throughput versus offered G traffic per station. $N_s = 5$, $N_w = 3$.

It is also important to verify if the reservation scheme fairly shares the spare resources between stations.
This is assessed in Fig. 2.6 which represents the throughput per station, for both G and BE traffic in the cases without and with BE traffic.

We first see on Fig. 2.6, that for the flow of traffic from station 5 to the Hub (station 3), G traffic is insensitive to BE traffic. We do not report what happens for the G traffic sent by the other stations, but the result is similar. These experiments clearly show that G traffic is protected by the reservation scheme. Let us now address the fairness issue for BE traffic. Fig. 2.6 shows the achieved BE throughputs for stations 5 and station 1. Since station 5 inserts its traffic before station 1 in the concentration scenario, it is to be expected that station 5 receives more opportunities for inserting its BE traffic unless the reservation scheme does enforce fairness. We can indeed see that station 5 is slightly favoured compared to station 1 in terms of BE traffic throughput, but fairness, although not perfectly enforced, is quite correct. Indeed, with greedy sources, the upstream stations could indeed starve the downstream stations unless the reservation scheme counterbalanced the topology advantage. Fig. 2.6 also shows how supported BE traffic decreases as G traffic increases.

- **Any-to-any Scenario Let us now assess the performance of the reservation scheme for the "any-to-any" scenario. In the "any-to-any" scenario, each station sends the same amount of traffic to each other station. The egress link from each station is of capacity 1. Let a_{a2a} be the amount of G traffic sent by one station to any other station; the amount of G traffic entering (and exiting) each station is thus $(N_s - 1)a_{a2a}$. Since the system**

is completely symmetric, the (single) link saturation condition reads:

$$(N_s - 1)(a_{a2a} + b_{a2a}) \leq 2N_w/N_s \tag{2.6}$$

Each ingress queue is modeled by a Geo/Geo/1 queue with parameters a_{a2a} and $[1 - ((N_s - 2)(N_s - 1)(a_{a2a} + b_{a2a})/2N_W)^{N_w}]$. The upper quantile of the insertion time is then limited by the following constraint :

$$(N_s - 1)a_{a2a} \leq \beta[1 - ((N_s - 2)(N_s - 1)(a_{a2a} + b_{a2a})/2N_W)^{N_w}] \tag{2.7}$$

As in the concentration case, it is straightforward to derive from these two conditions an upper bound for b_{a2a}. In this scenario also, depending on the value for a_{a2a}, it may be possible, or not, to fully use spare resources.

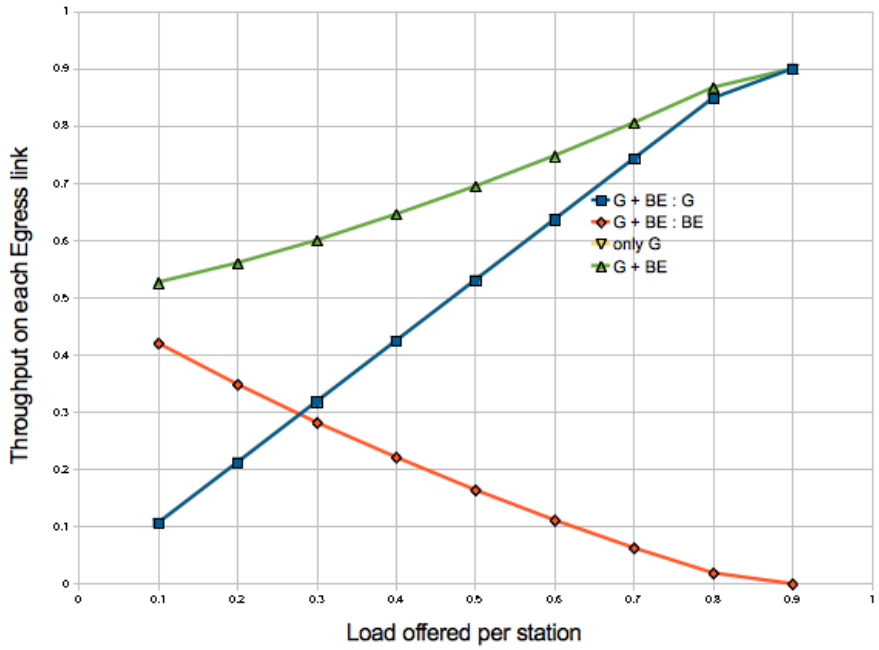

Figure 2.7: Any-to-any scenario. Achieved ingress throughput versus offered G traffic per station. $N_s = 6$, $N_w = 3$.

Using relations (2.6) and (2.7) for $b_{a2a} = 0$ and $\beta = 0.9$, we observe that the system with only G traffic is unstable for $(N_s - 1)a_{a2a} > 0.77$. Therefore, relevant areas in Fig. 2.7 and Fig. 2.8 are on the left hand side, for a load offered per station ($(N_s - 1)a_{a2a}$) smaller than 0.77.

Fig. 2.7 shows the achieved throughput in any station for both G and BE traffic, without and with BE traffic. As in the "concentration" case, we see that the G traffic throughput is not affected by the support of BE traffic thanks to the reservation scheme.
We can also check that spare resources are only partially used since in the area of interest (for the load offered per station smaller than 0.77), the total throughput varies between 0.51 and 0.85. This is not as efficient as

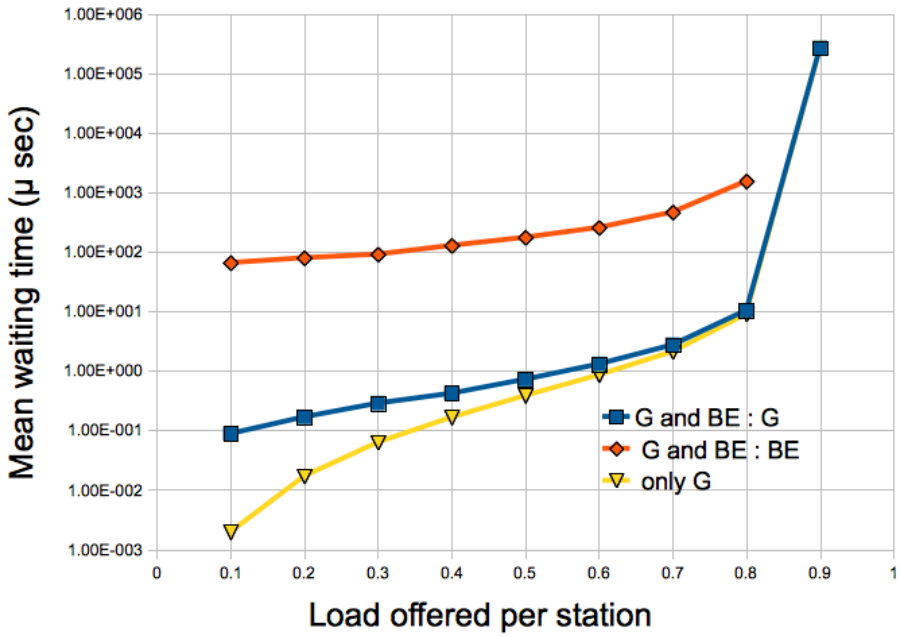

Figure 2.8: Any-to-any scenario. Mean insertion times versus offered G traffic per station. $N_s = 6$, $N_w = 3$.

in the previous case. Actually, the inefficiency especially at low G load is due to the fact that the reservation scheme precludes spatial reuse for BE traffic. This is why, at minimal G load, the reservation scheme only allows to use half of the ring capacity. This behaviour does not affect the "concentration" case which cannot take advantage of spatial reuse.

We assess the performance offered to both G and BE traffic in Fig. 2.8 which shows the sojourn time for G traffic and the insertion time for BE traffic. As predicted by our analytical model, we observe that the system becomes unstable when $(N_s - 1)a_{a2a}$ is close to 0.8. We also see that the sojourn time for G traffic is indeed impacted by BE traffic by comparing the mean waiting times for the cases without and with BE traffic. However, we see that the sojourn time degradation is very limited. We can also see that the insertion time for BE traffic is significantly larger than the sojourn time for G traffic. This was to be expected since the opportunistic insertion process is obviously more efficient and less constraining than the reservation based scheme used by BE traffic.

The results for any to any scenario are shown in Fig 2.7 and Fig 2.8.In this scenario all station behave similar and completely symmetrical ,thus the flow to one station is presented in the cases of and G+ BE flows.
As it is depicted in Fig 2.7 the links are never completely saturated and only in the sum flow of G+ BE the link utilisation saturation rate is faster. The waiting times experienced by customers in the any to any scenario before and after inserting Be traffic are shown in 2.8.Noticing the very little variation in the average waiting time for the two G and G+BE

curves confirms that the proposed reservation method method maintains the QOS for G flows while BE traffic insertion.

V Control information

Up to this section we have mentioned the critical information that each nodes has to have access to in order to properly treat each time slot. The structure and the content of the control packet determines the efficiency of control information processing and the switch reaction time. In the following we offer a general view of the control packet content that makes an integrated multi service, multi-functional MAC layer for suitable for POADM WDM metro rings.

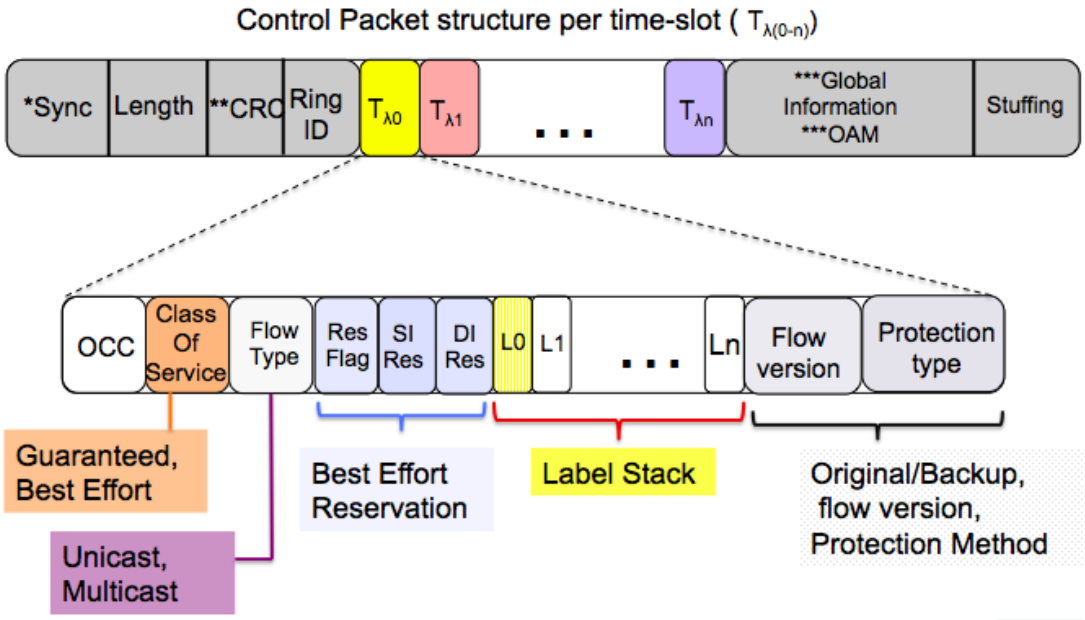

Figure 2.9: Control packet structure enabling Multi-service MAC layer

As mentioned in section II.1, part of the control information is contained in SDU and PDU headers, where the data traffic is still in electronics domain. The other part of the control/operational information is carried in the out of band control packet therefore these control information is split into in band and out of band, global and per time slot control information. Fig 2.9 shows the structure of the control packet. The global information such as physical link configurations, signalling information, OAM notification messages are carried in the global control fields:

- Framing and synchronising information.
- Length: control packet length
- Global information: includes global OAM notification messages.

- **Error correction information, which protects the complete CPDU contents.**

Some stuffing may be necessary to fill the control packet. It is easily discarded at the output station thanks to the Control Packet length information. In addition the information corresponding to the state of the data channel and more specifically on PDUs carried within in a given time slot are in the following:

- **Occupancy flag (OCC): Indicates if the time slot is free or occupied;**
- **Service type: Show to which class of service the PDU belongs .**
- **Flow types: It is the traffic type: unicast, or multicast.**
- **Reservation flag**
- **Source and Destination interface IDs relative to the reservation, if present.**
- **Label stack: To identify the flows and the relative action, the label stack application is mentioned in the section IV, Multi-Ring extension.**
- **Flow version: Indicates if the flow is original or a backup copy, the application is more explained in chapter III.**
- **Protection method: In case of failure it determines the method of extraction and behaviour of the switch during failure recovery time, detailed in chapter III.**

The maximum size for control channel packets considering three packet length and two bit rates. Thus for control channel with 2.5 Gbps bit-rate and the control packet length of 10 micro sec, the label stack depth can be at least equal to 5 for label size 20 bits. This allows covering all practical scenarios currently envisaged. This is also an upper bound on the label stack depth in operational MPLS backbone networks.

Conclusion

This chapter presented an extension on the architecture of perviously studied unidirectional time slotted POADM ring network into bidirectional multi-ring network.
The main features are kept intact as each node has an optical tuneable transmitter per direction and can transmit per direction one packet at a time slot. Moreover one out of band control channel exists per each direction of data channels. At the receiver side several wavelengths can be received on each direction and capacity at the receiver is limited by capacity of one wavelength. These assumptions made it possible to consider that each direction of the ring complies to the previous model and evaluations for unidirectional POADM ring. Then in the second section of the chapter we expanded the basic functional MAC layer that was only defined for a guaranteed unicast traffic. More detailed functional building blocks were defined in the general MAC structure sub-layers in order to build the SDUs and PDUs.
Then the concept of label switching was adapted for POADM ring as a mean to minimise the processes for flow and service identifications along with other necessary fields such as flow type, service type and control and management information. In addition to this we defined SDU and PDU level labels in order to facilitate packet grooming/un-grooming at the edge nodes.
Accordingly in the third section we proposed two new insertion-extraction methods for multicast flows were proposed. Noting that the insertion-extraction methods were all considered for original packets and we did not include the backup insertion-extraction methods since this subject is thoroughly studied in the following chapter.
Section IV focuses on ring interconnection methods. The two ring interconnection solutions use O/E/O conversion since connecting and synchronising all optical time slotted rings is a challenging task in practice due to the lack of optical buffers with practical capacity like minimum a few time slots.
In section V we propose a new reservation based access method for best effort traffic. Considering unknown amount of best effort traffic to be added to the guaranteed traffic, we utilise marking the time slots in order to use them in case there were no higher priority traffic. The proposed method then was evaluated for concentration-distribution and any to any traffic profile showing almost no disturbance on Guaranteed traffic QoS.

Finally we wrapped up this chapter by assembling the control packet necessary fields in order to provide a multi-service integrated MAC layer for POADM multi-ring metro network.

Bibliography

[1] **M. Herzog, M. Maier, and M. Reisslein. Metropolitan area packet-switched wdm networks: A survey on ring systems. *Communications Surveys Tutorials, IEEE*, 6(2):2–20, 2004.**

[2] **L. Sadeghioon, A. Gravey, and P. Gravey. A label based mac for ops multirings. In *Optical Network Design and Modeling (ONDM), 2011 15th International Conference on*, pages 1–6, 2011.**

[3] **Andersson L., Acreo AB, and Asati R. Multiprotocol label switching (mpls) label stack entry. *RFC5462*, 2009.**

[4] **L. Sadeghioon, A. Kabat, Darwish. S, and A. Gravey. An enhanced mac for supporting guaranteed and best effort traffic in a wdm optical packet ring. In *The first European Teletraffic Seminar*, pages 1–6, 2011.**

Figures and tables

Figures

Chapter 3

Protection Schemes for All Optical Packet switched network

Contents

Introduction

In the previous chapter, we introduced a multi service MAC layer protocol for an all optical POADM ring based technology suitable for regional/metropolitan area network. Here in this chapter we show how we benefit from the defined the MAC protocol with service integration in the control channel to provide a class-based all optical protection mechanism.
In the first part, we propose two protection approaches for unicast flows, by adapting the conventional 1:1 (Regular) and 1+1 (Premium) protection methods for a POADM based bidirectional single ring network. The two methods are analysed per time-slot by time-diagrams to find and measure the determining impairment factors such as packet loss, packet disorder, packet redundancy. The second part suggests two protection methods for multi-cast service in line with the extraction methods proposed for multicast flows earlier in the chapter II.
Ultimately this study leads to development of a general class-based protection method for POADM multi-ring metro network that also covers efficient methods for protecting services such as multicast. Thanks to per-slot control information availability, the single failure recovery time is kept in the range of a few milliseconds for regional/metro networks.

I Protection Mechanisms in Metropolitan Area Network Rings

Providing a survivable network against failures is one of the major requirements for transport networks and in particular for metropolitan area networks. A network with survivability is capable to continue offering service even after failure occurance.

Failures in the networks are generally produced by human errors. There are different possible sources of failures in a transport network, including hardware failure inside a node such as transmitter, or receiver failure, software problems that can be triggered by wrong configurations, in addition to links collapse that are usually caused by fiber cuts. Under very rare circumstances the whole central office can also fail due to natural disasters, power cut, or vandalism. However according to [Ramaswami], among all the above mentioned failure sources, the links failure caused by fiber cuts is the most common network failure incidence. Therefore we consider fiber cut as the failure cause throughout this study. The two fundamental Conventional technologies such as SONET/SDH offer 99.999% reliability with 50 ms recovery time that has set the standard in the carrier grade transport networks (actually the failure recovery time is 50 + 10 ms, as 10 ms is failure detection time). Although required failure recovery time entirely depends on the application and type of data being carried, most of the protection schemes on carrier grade technologies are designed to support the 50 ms or less, SONET/SDH standard.
There are plenty of works on protection schemes for electronic packets switched technologies in metropolitan area networks such as: Resilient Packet Ring (RPR), Ethernet Ring, MPLS-TP. They all comply to the conventional 50 ms standard.

- **Resilient packet ring (RPR), defined in IEEE 802.17 is a shared dual ring infrastructure for packet transfer between nodes. RPR provides two protection mechanisms: Steering, Wrapping. Both mechanisms meet 50 ms failure recovery time, however they have differences in performance. Steering uses updating topology in case of failure and then the nodes switch on the protection path,(the alternative ring) that can result in large amount of packet loss. Whereas in wrapping the two ends of a failure loopback the line and the other nodes continue to send the packets on the same path and then they are looped back to the other end and redirected to destinations therefore less cost efficient in terms of use of resources.**
- **ITU G.8031 standard offers 1+1 and 1:1 protections both for unidirectional and bidirectional paths in Ethernet networks, where both ends of a failure will exchange failure notification messages and confirmation status to switch on protection paths. ITU G.8032 Ethernet Protection Ring, uses ring automatic protection switching (R-APS). To avoid loops one of the links on the ring is set to ring protection link, when a failure happens the working ports of the both failure end nodes are blocked and a notification failure message is sent to the network thus concerning nodes switch over on protection ring.**
- **Multiprotocol Label Switching - Transport Profile (MPLS-TP) as a connection oriented protocol to transport IP packet in transport network offers linear protection and ring protection. In the linear is the classic 1+1 and 1:1 mechanisms per label switched paths (LSPs) and relies on hello message for connection control. The ring protection RPR protection methods explained earlier per LSPs.**

To summarise each of the above mentioned methods involves two steps:

1. **Failure notification/ localization**
2. **Protection procedure selection**

We have addressed the survivability issues in a Packet-OADM ring by adapting those two steps in protection procedures in specific Packet-OADM ring technology and procedures. Therefore in the following we have to specify a global notification message and then to propose protection procedures. The performance of the proposed schemes needs to be evaluated.

II Resilient POADM bidirectional single ring

As we mentioned in section I there are different ways of protecting traffic that can be adapted to bidirectional POADM rings. Moreover since we have a label based access method to identify different flows, we can propose several protection classes[2]. Thus we consider three major traffic categories regarding different levels of availability:

- **Premium Traffic: Traffic loss is (nearly or totally) not tolerated, and the service must be constantly available.**

- **Regular Traffic: Certain level of traffic loss and limited service unavailability is tolerated.**

- **Unprotected: No protection mechanism is planned for this kind of traffic.**

The network operator can support one or several classes of protection. In the latter case the service provider has the choice to offer a range of options to the customers (the option could be selected e.g. in SLA).

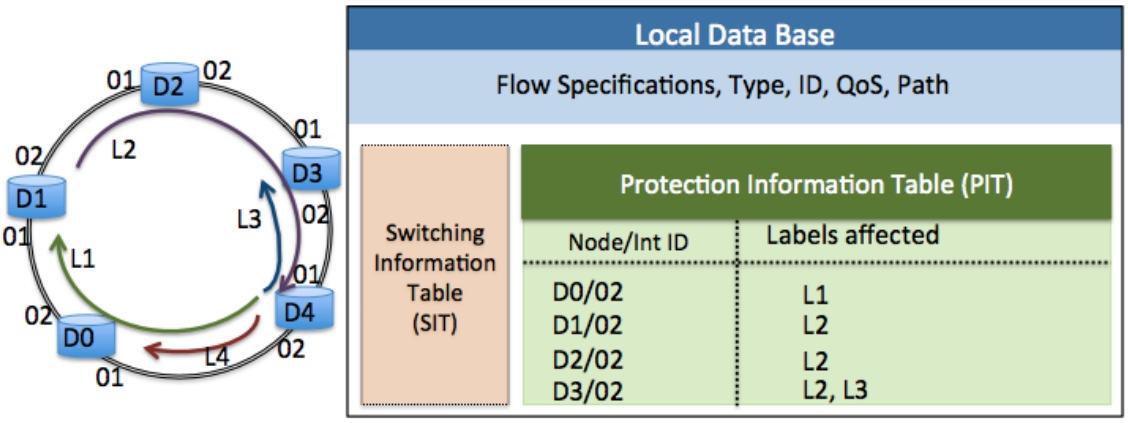

Figure 3.1: Protection Information Table content within the Switch local Information Database(node D4)

There are three main steps to provide protection of any class in the network:

- **Failure notification/Localization: Through control packet, general info section.**

- **Failure recovery: This depends on the protection method that is chosen.**

- **Restoration: This depends on the protection method**

There is a general failure localization method in order to trigger different types of protection methods in case of failure. In case of failure the two end point of the failure generate notification failure messages and send them towards the operational directions. The notification failure message carries node ID node, interface ID and other information such as time of the failure.

Fig 3.1 shows the protection information table (PIT) that is part of the nodes local information database. The table contains Node ID, Ring ID, and labels that are affected by the failure. Thus every node compares the notification failure message content with PIT and localize the failure and invokes the method per labels that are affected accordingly. PIT is generated when the switch is setup and configured for certain traffic matrix and it is updated if there is any change in the services it offers as source or destination.

II.1 Premium (1+1) protection

- **Nominal Operation Mode In the nominal operation mode (i.e. no failure), the original and backup copies of the same flow is sent on both directions of the ring. Both flows carry the same label in their control packet on their time-slot that identifies the source-destination pair. However the Flow-Version field in the control packet per each flow is differently set to `original` for the working path and `Backup` on the protection path. At the destination the original flow is received and stripped off where as the backup flow is discarded.**

- **Failure recovery process:**

  Reconfiguration of insertion process

  The affected source node stops sending the affected flow on the shortest path and marks the packets sent on the longest path as Flow-Version=`original`.

  Reconfiguration of extraction process the node stops dropping the packets marked as Flow-Version=`backup`. This is intended to avoid loosing packets but may lead to packet disordering.

- **Restoration The restoration procedure begins as soon as the broken link is repaired. The 2 nodes at both ends of the failure start receiving control packets again and thus send out the OAM RESTORE notification message in the control packets. Learning the end of a failure, the source nodes of affected flows reverse to their nominal behaviors, for all types of traffics. The sources resume sending affected flows in both directions and respectively reset the status of the Flow-Version field in the control message.**

II.2 Regular (1:1) protection

- **Nominal Operation Mode In the nominal operation mode (i.e. no failure), each Regular is received and stripped off at destination which enables spatial reuse.**

  Reconfiguration of insertion process: The source node of the affected flows stops sending the them on the shortest path and starts inserting the packets on the alternative direction.

  Reconfiguration of extraction process: the node has to receive packets on its other interface.

- **Restoration The source of Regular and Unprotected flows that were affected by the failure resumes sending packets on the shortest path towards the destination. the destination node of Regular and Unprotected packets receive packets on both interfaces, which may also lead to packet disordering.**

II.3 Performance of the Protection Schemes

There are some transient performance degradation that still affect Premium and Regular traffics in case of failure. By computing both recovery and restoration times, we have identified the potential degradations (packet loss, packet duplication and packet disorder) during these time periods.

In order to illustrate typical situations, consider a bidirectional ring with n nodes sending unicast traffic, and let T_R be the total cycle time (identical in both directions). Denote $T^s_{i,j}$ and $T^l_{i,j}$ be respectively the shortest and longest time distances between 2 nodes i and j.

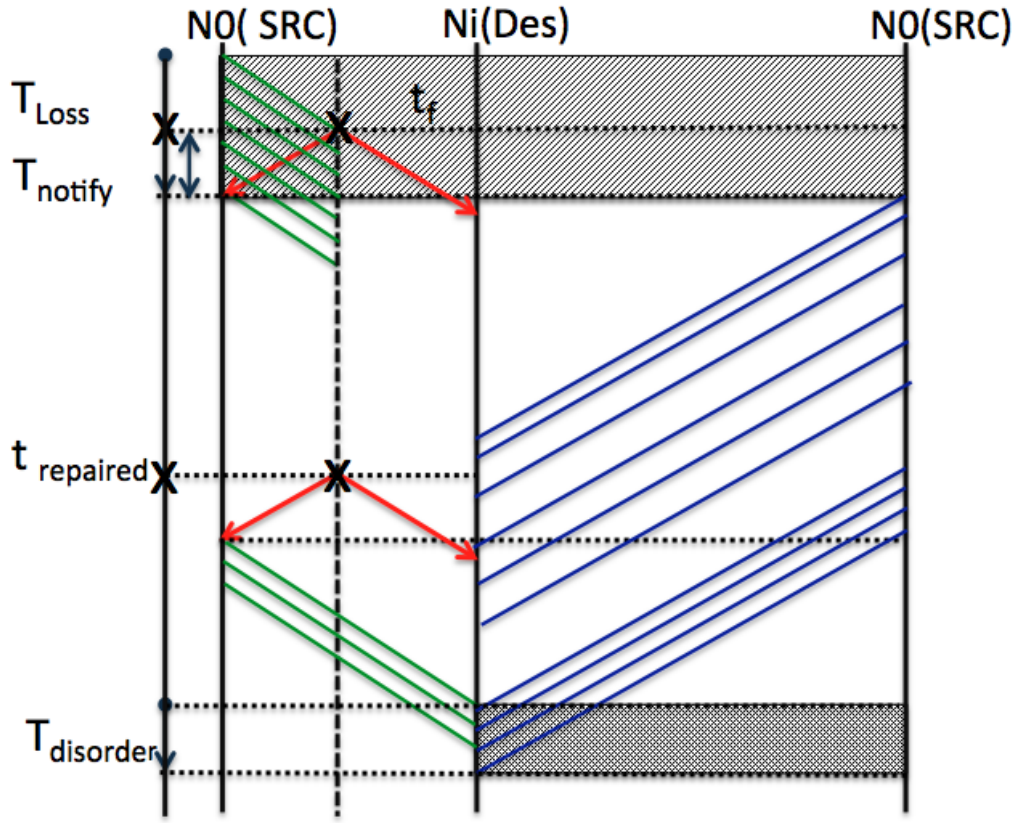

Figure 3.2: Time Diagram of 1:1 Regular protection method. Failure happens at t_f, $T_{notify} = T^s_{0,f}$.

- **Performance for Regular traffic**

  Assume that node N_0 sends a Regular traffic flow to node N_i $(i < n-1)$ on the shortest path, while a failure occurs at t_f between nodes N_f and N_{f+1}, $0 \leq f \leq i-1$. The top part of the time diagram in Fig.3.2 shows packets sent by N_0 before the source node is aware of the failure. Those packets are all lost. Let T_{loss} be the duration of the loss period. The worst case is when the failure occurs close to N_{f+1}. Assuming that the failure detection time is negligible comparing to point to point propagation time, N_0 learns about the failure at $T^s_{0,f+1}$ and we obtain :

$$T_{loss} \leq 2XT^s_{0,f+1} \tag{3.1}$$

which shows that in the worst case ($i = f + 1$ and $T^s_{0,f+1} = T^l_{0,f+1}$), T_{loss} is at most equal to T_R.

The middle part of the time diagram in Fig.3.2 shows the period of time during which the source, aware of the failure, sends the packets on the longest path. N_i may start receiving packets on the longest path at time $t_f + T^s_{0,f+1} + T^l_{0,i}$. Note that N_i is stopped from receiving packets from N_0 due to the failure during at most $T_{loss} + T^l_{0,i} - T^s_{0,i}$, which is also upper bounded by T_R (the worst case being when $i = f + 1$).

During restoration, some packets sent by N_0 on the longest path during the fault recovery period may arrive after packets sent on the shortest path by the source once it is aware of the restoration; this may yield packet disordering at N_i (but neither loss, nor duplication). This is shown on the bottom part of the time diagram in Fig.3.2. The period of time during which N_i may receive disordered packets ($T_{disorder}$) is equal to $T^l_{0,i} - T^s_{0,i}$. A solution to the disorder problem would involve numbering the packets, and buffering temporarily packets received during the restoration period. Another option is to let the upper layer deal with the disordering.

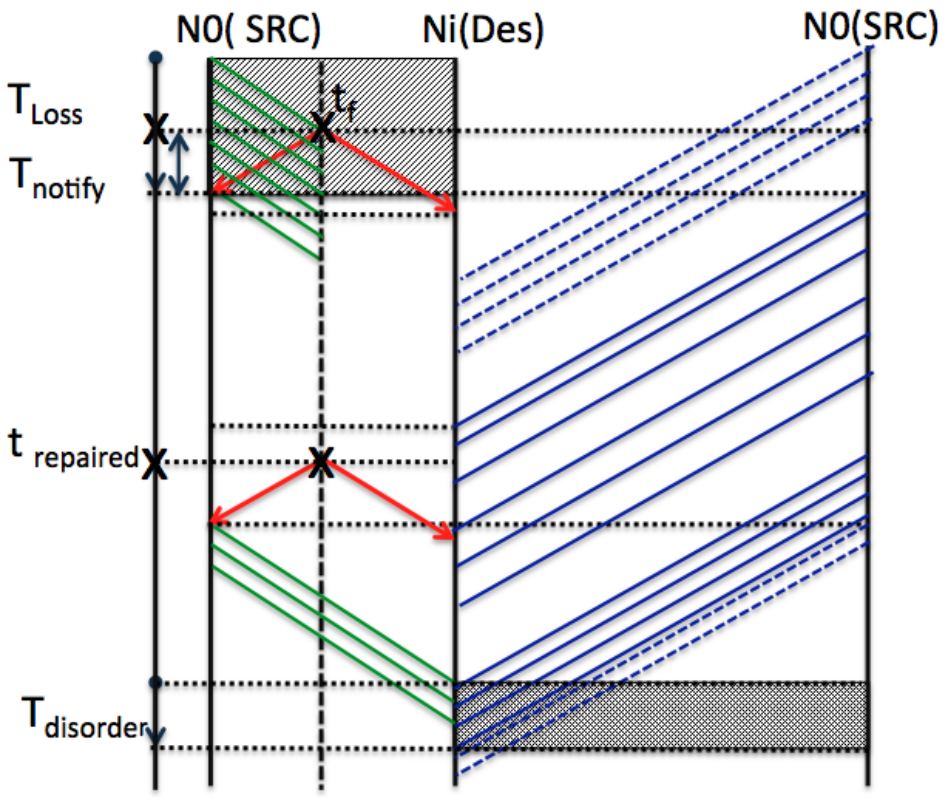

Figure 3.3: Time Diagram of 1+1 Premium protection method. N0 as Source node, Ni Destination node. t_f failure time and $T_{notify} = T^s_{0,f}$.

- **Performance for Premium traffic**

  We now assume that node N_0 sends a Premium traffic flow to node N_i in

both directions, while a failure occurs at t_f between nodes N_f and N_{f+1}, $0 \leq f \leq i-1$.

The top part of Fig.3.3 shows that the last instant of time when N_i, receives a packet from the shortest path is just before N_i learns of the failure. The destination node then starts receiving packets from the longest path, even when marked with Flow-Version=`backup`. Actually, the first packets had probably been received already on the shortest path! There is indeed a potential duplication period of duration $T^l_{0,i} - T^s_{0,i}$. Then, N_i starts receiving the packets sent on the shortest path by N_0 and lost due to the failure during the period of duration T_{loss} computed previously. However, in the Premium case, the packets are received on the longest path, although marked with Flow-Version=`backup`, and are thus not lost. After this second period, N_i continues receiving packets on the longest path, but marked with Flow-Version=`original`, as long as the failure is not repaired. Sending packets on both shortest and longest paths ensure that no Premium packets are lost although N_i receives some duplicate packets during this interval of time.

Restoration is somewhat similar for both Premium and Regular flows. Due to the difference between shortest and longest paths, some disorder can be observed during a period of time of duration $T^l_{0,i}-T^s_{0,i}$, when the destination node receives Premium packets marked with Flow-Version=`original` both on the shortest and longest paths. This period ends when N_i receives the first packet on the longest path marked with Flow-Version=`backup`.

As pointed out previously, duplication and disorder are handled with upper layer applications in the network.

II.4 The Cost of Traffic Protection in Bidirectional Optical Packet Switching Rings

The dimensioning problem in OPS rings usually consists in allocating the wavelengths and the equipment (i.e., transmitters and receivers) in the bidirectional ring in order to support the given traffic flows, with the objective of minimizing the CAPEX cost. The CAPEX cost of the ring is composed of two main costs: wavelength leasing cost per link l, (C_l), and cost of fixed single-wavelength receiver (C_r). 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. Transmitter cost is neglected as each node is equipped with a single tunable laser or even several of them in the case of high traffic demands [3].

The input data of the problem are the bidirectional ring topology and the set of traffic demands. The result of the design is the network configuration in terms of number of wavelengths needed in the network, the allocation of wavelengths to the traffic flows and distribution of fixed single-wavelength receivers at ring nodes. The costs are normalized and given in arbitrary units.

A part from wavelength assignment problem that need to be resolved in the bidirectional POADM ring, the two protection methods explained earlier are considered during formulation the problem.

II.4.1 Dimensioning solution for bidirectional POADM ring with protection

The problem of dimensioning of POADM ring is formalized as a multi-commodity flow problem with a 0-1 Integer Linear Programming (0-1 ILP) formulation, which is given next.

0-1 Integer Linear Program

The input traffic matrix of ordinary working mode T_o^D can contain flows of three types: for premium, regular and best-effort traffic, in one of two directions of communications (noted with D). It is supposed that between two stations (e.g. A and B) can exist only one single connection "A to B", with one of the three possible protection schemes. This means that connection B to A is also possible and it can benefit from another protection mode. Also, it is supposed that all the traffic matrices contains only the working flows that satisfy the "shortest-path" rule in each ring direction. The traffic matrices are given for both directions simultaneously.

From T_o^D, we derive traffic matrices T_l^D ($l \in \mathcal{L}$, where $\mathcal{L} = \{L_1, L_2, ..., L_N\}$ is the set of network links), which contain the working and protected traffic for the case of failure of the link l. In the case of failure, the traffic matrices will change because for regular traffic some new flows will be included. For premium traffic, the incident of link failure does not induce a traffic matrix change. Similarly, the best-effort traffic is unprotected, so the same flows for this traffic are given in all T_l^D and T_o^D matrices.

Given Parameters

- **$G(V, \mathcal{L})$: a non-directed graph representing the bidirectional ring, where V is the set of nodes, $\mathcal{L}$ is the set of bidirectional links;**
- **$N = |V|$: number of nodes in the ring;**
- **D: direction of transmission: $D = 1$ for clockwise direction, $D = 2$ for counterclockwise direction;**
- **T_o^D: traffic matrix in normal operation mode in direction D; in all traffic matrices, element, t_k, is the traffic rate (in bits/s) requested by the k-th flow;**
- **T_l^D: traffic matrix in protection mode (if link l has fallen) in direction D;**
- **T_e^D: any of matrices T_l^D or T_o^D;**
- **T_a^D: general traffic matrix, containing all possible flows in the ring, i.e. $T_a^D = T_{L_1}^D \cup ... \cup T_{L_N}^D \cup T_o^D$;**
- **$\mathcal{P}$: matrix of protection with elements $\mathcal{P}(k)$, such that: $\mathcal{P}(k) = 1$ for premium protection, $\mathcal{P}(k) = 2$ for regular protection, and $\mathcal{P}(k) = 0$ for all other flows k;**
- **B: wavelength capacity (in bits/s);**
- **C_r: cost of a receiver;**
- **C_l: cost of link $l \in \mathcal{L}$;**

- W: **maximum number of wavelengths per fiber;**
- π_k^D: **path of the k-th flow (i.e., set of links connecting node s to node d) in direction D; flow $\overline{k}$ is the protective traffic of the same size sent in the opposite direction, i.e. from d to s.**

Variables

- **binary $p_w^{k,D}$ indicates whether k-th flow is routed on wavelength w in direction D;**
- **binary r_w^i indicates whether node i requires a receiver on wavelength w in direction D;**
- **binary variable $x_w^{l,D}$ defining a "link-route", i.e. indicating whether wavelength w on link l is used to carry the traffic in direction D.**

0-1 ILP formulation

The objective function minimizes the overall ring cost, i.e., the cost of the receivers and the link-routes required in the ring:

$$Min\left(\sum_{D=1}^{2}\sum_{i\in V}\sum_{w=1}^{W} C_r \cdot r_w^{i,D} + \sum_{D=1}^{2}\sum_{w=1}^{W}\sum_{l\in\mathcal{L}} x_w^{l,D}\cdot C_l\right) \tag{3.2}$$

Constraint ensuring that one (and only one) wavelength is assigned to each traffic demand, in each direction (valid for all ring connections, i.e. the elements of T_a^D):

$$\sum_{w=1}^{W} p_w^{k,D} = 1, \quad \forall k : t_k \in T_a^D, \forall D \tag{3.3}$$

Constraint ensuring that, for each traffic matrix T_e^D, the traffic rate routed on wavelength w does not exceed the wavelength capacity:

$$\sum_{k:l\in\pi_k^D, t_k\in T_e^D} p_w^{k,D} t_k \le 1, \qquad \forall w,\ \forall l\in\mathcal{L},\ \forall e,\ \forall D \tag{3.4}$$

Constraint defining "link-routes" $x_w^{l,D}$ and allowing multiple flows, belonging to different protection events, to be supported by reusing the same wavelengths, which in turn allows saving of network resources in the case or regular protection:

$$\sum_{k:l\in\pi_k^D, t_k\in T_a^D} p_w^{k,D} t_k \le (N+1) x_w^{l,D},$$

$$\forall w,\ \forall l\in\mathcal{L},\ \forall D \tag{3.5}$$

Constraint limiting the overall traffic received on wavelength w at destination d, in direction D, for each traffic matrix T_e^D:

$$\sum_{k:d=dest(t_k), t_k\in T_e^D} p_w^{k,D} t_k \le 1 \qquad \forall d\in V,\ \forall w,\ \forall e,\ \forall D \tag{3.6}$$

Constraint forcing the use of the receiver on wavelength w, at destination d, in direction D, for the flows sharing this receiver:

$$\sum_{k:d=dest(t_k), t_k\in T_a^D} p_w^{k,D} t_k \le (N+1)\cdot r_w^{d,D} \qquad \forall d\in V,\ \forall w,\ \forall D \tag{3.7}$$

Optional constraint forcing the use of the same wavelength for both working and protection path, for each connection k and on each wavelength w (the so-called "color constraint"):

$$p_w^{k,1} = p_w^{\overline{k},2}, \quad \forall w, (\forall k)(t_k \in T_a^1, t_{\overline{k}} \in T_a^2, t_k = t_{\overline{k}}, \mathcal{P}(k) \geq 1). \tag{3.8}$$

II.4.2 Numerical Results

In this section, the results of the optimal dimensioning are given for a 5 node ring, in diverse network scenarios. The goal is to quantify the impact of protection schemes on the CAPEX cost of the network. The 0-1 ILP formulation is solved by IBM CPLEX optimization software, with an academic license. We have performed an extensive set of simulations for two types of network traffic: random centralized traffic and uniform and symmetric traffic.

In all the examples, the channel rate is set to 10 Gbps, while the wavelength cost per link per km can be calculated with the formula $C_l = (10 * C_r/(100km)) * l$, where C_r is the receiver cost and l is the link distance in km. The formula is derived from [4], where the ratio between the wavelength cost per ring and the receiver cost is estimated on a ring circumference of 100 km. Both wavelength cost and receiver cost are given in arbitrary units (a.u.) and it is always considered that $C_r = 0.1$.

II.5 Random Centralized Traffic

Here, the results are averages collected on 100 designs with randomly generated traffic flow matrices. The generation of the matrices is modeled with two parameters: the load, σ, which indicates the overall amount of traffic supported by the network and the traffic distribution factor, θ, which indicates the amount of traffic passing through a selected node acting as a hub. For $\theta = 0\%$, all the traffic passes through the hub, while for $\theta = 100\%$, the traffic matrix is fully decentralized. The channel rate is set to 10 Gbps, and the value of σ is in $\{100, 150, 200, 250, 300\}$ Gbps. All the links are supposed to be of the same size and the overall ring circumference is taken to be 100 km, resulting in $C_l = 0.2$.

The first diagram, in Fig. 3.4, shows the optimal design cost for premium and regular traffic, when "color constraint" is used. As it can be seen, the premium protection is up to 20 % more expensive than the regular one. This difference is the greatest for the highest traffic load, and increases with traffic increase. On the other hand, it can be seen that the parameter θ has a mild impact on the price of design: with the increase of this parameter the difference in price between premium and regular protection also increases. Since for higher values of θ traffic is more evenly distributed between network nodes, we can conclude that the centralized matrices do not allow traffic to take the advantage of a more efficient use of bandwidth in the case of regular traffic.

To get even better view on the solution of the optimal network dimensioning, in Fig.3.5, the number of link-routes and receivers in the optimal solutions is presented. As expected, premium traffic demands more link-routes in the solution (more than 20% for traffic load of 300 Gbps).
The number of link-routes increases with the increase of σ and θ. On the other hand, the number of receivers in the solution for premium traffic is only

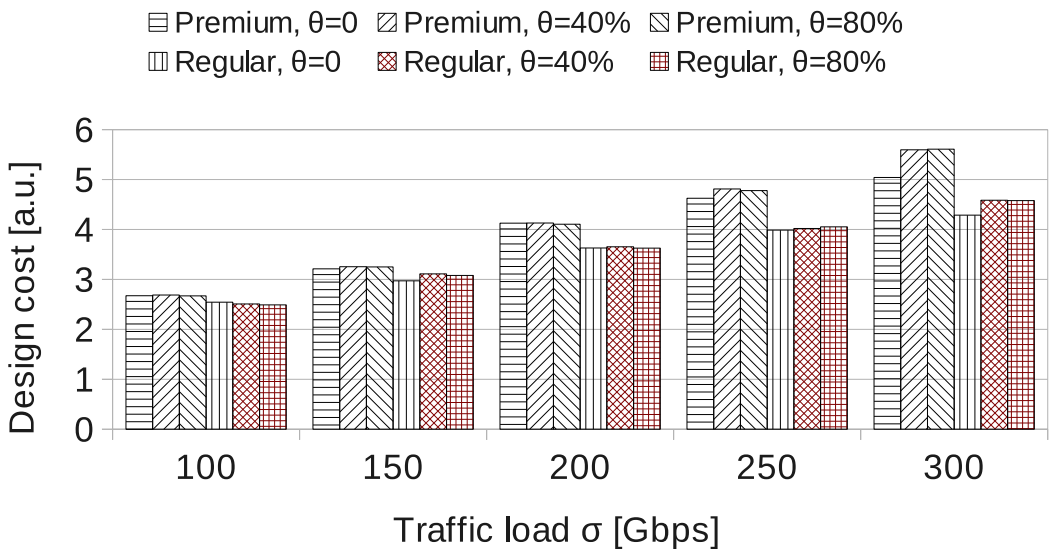

Figure 3.4: Design cost of optimal solution vs traffic load for premium and regular traffic

slightly greater than in the solution for regular traffic, and this difference seems to mildly increase with the increase of σ. Interestingly, the number of receivers is clearly smaller for more distributed traffic matrices (higher values of θ). It is because the traffic is more balanced and destinations receive traffic on more equally charged wavelengths for higher values of θ.

Fig.3.6 shows the results of the comparison of design cost for premium, regular and best-effort traffic, in the same scenario, but only for $\theta = 40\%$. Premium protection is more expensive than regular one (up to 20% here) and much more expensive than in best-effort case (up to 55%). At lower loads, the difference in cost between regular and premium protection is smaller, so it is probably wise to use premium protection for such loads.

The impact of "color constraint", given by Eq. (3.8) in Section IV, is evaluated next for regular protection and distribution factor $\theta = 40\%$. The design cost, the cost of link-routes and receivers are compared in Fig. 3.7 for the dimensioning with and without "color constraint". The conclusion is that forcing color constraint leads to an increase in network cost, but this increase is rather small. Indeed, a pick in design cost increase due to this constraint is below 10% (achieved for $\sigma = 300$Gbps). This increase is higher for higher network loads, because more connections are present and more wavelengths are needed in the optimal solution. Furthermore, the color constraint mainly affects the number of link-routes, by increasing them, while the increase of the number of receivers in the solution, although it exists, is negligible. It is mainly because the receiver cost is much smaller than a cost of using the links on a wavelength for transport, so consequently their number has already reached the maximum for the studied traffic loads.

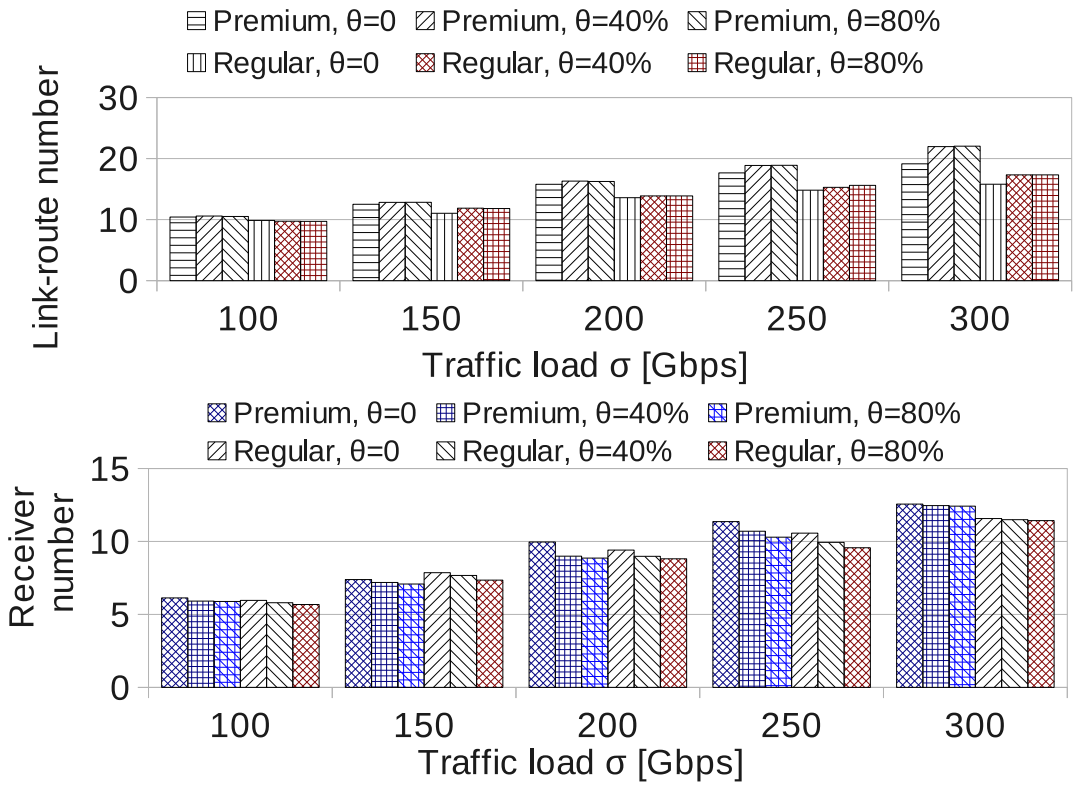

Figure 3.5: Number of link-routes and receivers in optimal solution vs traffic load for premium and regular traffic

II.6 Uniform and Symmetric Traffic

In this section, the focus is on a 5-node ring with uniform and symmetric traffic matrix. It is supposed that between any two stations, in both directions, there is a traffic connection of amplitude a (in diagrams normalized to a single wavelength capacity). The CAPEX costs of components are like in the previous section.

Design cost for different protection schemes for the uniform and symmetric traffic is shown in Fig. 3.8. Premium protection cost is still dominant, starting from small loads. For the highest traffic load, premium is up to 60% more expensive than best-effort and 40% more expensive than regular protection. In this example, premium traffic is not impacted by the color constraint, so the results of dimensioning without this constraint are given for regular traffic only. As in the previous example, the use of color constraint results in a design cost increase. This time, the increase in cost is higher (more than 30%, achieved for $a = 0.24$). In the following the color constraint is always used.

Fig. 3.9 shows the CAPEX cost components in the previous scenario: receiver cost (Rx), the cost of link-routes (second sum in Eq. (3.2)), but also the number of wavelengths in design, for different protection levels. As in previous examples, the link-route cost is dominant, which is expected, as the cost of a wavelength per single link is greater than a single receiver cost. Both link-route

and receiver cost increase with traffic amplitude a, and also the same cost hierarchy of the protection methods exists like in the previous section: premium has the larger link-route and receiver costs than regular protection. Regarding the wavelength cost, it is also the highest for premium, and the cheapest for best-effort traffic, which is expected, as the number of wavelengths in the optimal solution is related to the number of link-routes in it.

Next, a mixture of traffic with premium and regular protection schemes is studied by simulation for uniform and symmetric traffic with $a = 0.42$, and the results are presented in Fig. 3.10. Here, η is the "traffic mix ratio", defining the percentage of traffic connections using the regular protection in the traffic mixture. The design cost and number of wavelengths in the solution are given for the case where different protection schemes use "separate" wavelengths, and when they use the wavelength "sharing". In the former case, the optimization is performed separately for different protection schemes, and the results are obtained by the addition, while in the later case the optimization is performed for the mixture of traffic. The separation can be interesting from the operator point of view, and this is why it is studied here. From Fig.3.10 we can see that the "separation" of wavelengths is more expensive, which is an expected result. The difference in cost goes to more than 20% (for $\eta = 40\%$) and up to 50% more of wavelengths is needed (for $\eta = 60\%$).
Next, for instance, a mixture of 40% of premium and 60% of regular traffic leads to a network that is almost 20% cheaper than a network protected exclusively according to the premium scheme. Finally, when "sharing" the wavelengths, the introduction of the regular traffic into the mixture at a rate lower than $\eta = 20\%$ does not make sense, as the cost reduction is negligible.

In the final example, the impact of topology size on dimensioning results is studied by scaling one of the network links (in notation L_1), with factor α and for $a = 0.42$. Fig. 3.11 presents the dimensioning results obtained for different values of α and in case of regular traffic. Both the design cost and the number of traffic demands routed over link L_1 (i.e. the number of link-routes) in the optimal solution, are given. With the increase of α, the design cost rapidly increases, as expected. The main conclusion is that by increasing the link size, the number of traffic demands routed over the link is decreasing. Such behaviour can be justified by the fact that the route through this link becomes more and more expensive with the increase of α. However, regardless of the link cost increase, the number of the demands routed on it cannot decrease down to 0, as some minimum number of connections will always need to use it for the capacity in the case of network failures.

III Protection Mechanism for Multicast Flows

Earlier in chapter II we proposed Source stripping and segmented bidirectional multicasting.
Here in this section we examine 1+1 and 1:1 protection on both multicast methods and compare the performance [5].

III.1 Source strip Multi-cast Flows Protection

- **1+1 Source Stripping: This scheme is illustrated in Fig3.12 Each source sends an "original" flow in the "working" direction and a "backup" flow in the other direction. The status of the flow (either "backup" or "original") is distinguished by a flag within the control message. This allows to use the same label for both copies which implies that there is no change in label consumption. In nominal operation mode, nodes that receive the flow copy only the packets from the original flow, and ignore "backup" packets.**

  In case of failure, the nodes that receive the flow, and are situated after the failure in the working direction, stop ignoring "backup" packets. Moreover, the source toggles the status of the "backup" flow to "original" during the recovery period. The source resumes to nominal operation mode (i.e. marks packets as "original" in the working direction and "backup" in the other direction) as soon as it receives OAM repair notification messages that is sent by the nodes adjacent to the failure. Receiving nodes learning about the repaired failure resume to ignoring the "backup" packets.

- **1:1 Source Stripping Multicast flows are sent in a single direction (the "working" direction). In nominal operation mode, nodes that receive the flow get this single copy. In case of failure, the source sends the multicast flows in the backup direction with the labels used by the original flows in nominal operation mode. The destinations of the multicast flow then start receiving the packets on the backup direction. As usual, 1:1 protection uses less resources than 1+1 traffic in nominal operation mode, although it is easily seen that the total amount of resources to be provisioned is the same for both 1+1 and 1:1 single failure protection.**

III.2 Segmented Multi-cast drop off Protection

In this method, for a given source node, the network is segmented into two (more or less) equal sections in opposite directions. A multicast flow is duplicated at source and sent in both directions with the same label, and Flow-Version= "original". The two copies reach a point that is called multicast drop off point (MuD) Fig3.13.

1+1 MuD protection: Each packet from a multicast flow carries 2 control flags: one flag which indicates whether the packet is "original" or "backup", and another flag which indicates whether the packet has to be received by a reception node or not. In nominal mode, the MuD toggles the "original" mark to "backup" on each packet which then continues to the source which then drops both ("backup") packets. Reception nodes only receive packets marked as "original". As soon as a failure is detected by a node (either because it is adjacent to the failure, or because it receives a failure notification message), the protection mechanism within the node operates as follows:

- **The source behaves as in the nominal mode;**

- the MuD of a given multicast flow toggles the reception flag of the packets towards the failure in order to notify nodes (between MuD and failure) to receive backup packets;
- a receiver node continues receiving "original" packets (as in the nominal mode), and also receives "backup" packets when the reception flag is set;
- the two nodes adjacent to the failure drop the flows coming from both directions (this requires a multicast indicator within the flow specific control information. This function is set to avoid duplications on both sides of failure.). When the end of the failure is detected, the MuD stops setting the reception flag and the operation resumes to normal.

1:1 MuD protection: In nominal mode, the MuD drops the multicast packets (instead of marking them as "backup" as in 1+1).
Reception nodes receive a single copy of each multicast packet. This is illustrated in Fig.3.13. In this mode, the source node does not have to drop multicast packets since they are stripped at the MuD (which obviously is more efficient in terms of resource usage). As soon as a failure is detected by a node (either because it is adjacent to the failure, or because it receives a failure notification message), the protection mechanism within the node operates as follows:

- the source sends packets in both directions as in the nominal mode, but starts dropping any packet marked as "backup";
- the MuD stops dropping the multicast packets, but marks them as "backup";
- a receiver node continues receiving a single packet (either marked as "original" or "backup");
- the two nodes adjacent to the failure drop the flows coming from both directions (performs the same operation as in 1+1 mud as expected to avoid duplications).

When the end of the failure is detected by the MuD, it stops marking packets as "backup" and drops them instead. The source continues dropping any packet marked as "backup", which may arrive during the restoration period.

III.3 Performance Results

Multicast traffic is considered to be part of the traffic matrix that has been used to dimension the network. Therefore this service does not suffer from resource depletion or congestion both in nominal and protection modes. However they may experience temporary service degradations, during the transition from nominal to protection mode and then from protection back to nominal mode. The amount of transient degradation depends on the actual protection technique. In this section we study the transient degrading factors that are caused during fault recovery and restoration by the methods discussed earlier. The possible degradation factors are: packet loss, duplication and disordering. To

evaluate them, we assume that time to detect the failure and initiate the OAM message is negligible, in comparison to the propagation time within the network. This assumption is consistent with the hypotheses of a 10 microsecond time slot duration and OAM messages sent per time slot. To perform the analyses of each method, we consider a bidirectional time slotted ring with a total cycle time equal to T_R.

Derivation of the degradation times

During the failure recovery phase, loss occurs when the emission time of the first received backup packet by node N, $T_e^{FB}(N)$ is greater than the emission time of the last received working packet, $T_e^{LW}(N)$, whereas duplication happens in the opposite case. Thus, loss and duplication durations, $T_{Loss}(N)$ and $T_{Dup}(N)$ are calculated as follows:

$$T_{Loss}(N) = T_e^{FB}(N) - T_e^{LW}(N) \tag{3.9}$$

$$T_{Dup}(N) = T_e^{LW}(N) - T_e^{FB}(N) \tag{3.10}$$

In the rest of this study, we simply denote $T_{Loss}(N)$ as T_{Loss} and $T_{Dup}(N)$ as T_{Dup}.

There is no loss when restoring the network to nominal mode after repairing the failure, but duplication may still happen. The condition for duplication is expressed as a function of the emission times of the first working, $T_e^{FW}(N)$ and last backup packets, $T_e^{LB}(N)$received by node N. Therefore, in restoration operation:

$$T_{Dup} = T_e^{LB}(N) - T_e^{FW}(N) \tag{3.11}$$

Packet disorder happens when the emission times of two consecutively received packets are decreasing. Thus, in case of loss, this condition reads as $T_r^{FB}(N) < T_r^{LW}(N)$ where r stands for reception time. Disorder may also happen concurrently with duplication but it is not treated as a separate degradation factor. The relationship between emission and reception times obviously depends from the node position. We first introduce the notation $T_{N,P}$, which is the propagation time from node N to node P, measured along the working direction. (Thus $T_{N,P} = T_R - T_{P,N}$.) Using this notation, and naming Src the source node of the multicast flow, we may write:

$$T_r^B(N) = T_e^B(N) + T_{N,Src} \tag{3.12}$$

$$T_r^W(N) = T_e^W(N) + T_{Src,N} \tag{3.13}$$

Thus from equations (3.12) and (3.13), the disorder during failure recovery is derived as:

Protection	1:1 Src-S	1+1 Src-S	MuD 1:1	MuD 1+1
Loss	$N \in [F, Src]$, $T_{Loss} = T_{Src,F} + Min(T_{Src,F}, T_{F,Src})$	$T_{Loss} = Min(2T_{Src,F}, T_R - 2T_{N,Src})$	$N \in [F, D]$, $T_{Loss} = T_{Src,D} - Min(T_{Src,D}, T_{D,Src})$	NO
Duplication	NO	$N \in [F, M]$, $T_{Dup} = 2T_{N,Src} - T_R$	$T_{Dup} = T_{D,Src} - Min(T_{Src,D}, T_{D,Src})$	$T_{Dup} = T_R - 2T_{Src,N}$
Disorder	$T_{Dis} = T_R - 2(T_{Src,F} + T_{N,Src})$	$2T_{N,Src} \leq (T_{F,Src} - TSrc, F)$, $T_{Dis} = (T_{F,Src} - TSrc, F)$	NO	
Loss	NO			
Duplication	$N \in [Src, F]$, $T_{Dup} = T_{F,Src} + Min(T_{Src,F}, T_{F,Src})$, $N \in [F, Src]$, $T_{Dup} = T_{Src,F} + Min(T_{Src,F}, T_{F,Src}))$		NO	
Disorder	NO		$T_{Dis} = T_R - 2T_{Src,N}$	

Table 3.1: Durations of the transient degradation factors during the failure recovery (upper part) and restoration phases (lower part) for the four protection methods under study.

$$T_{Dis} = T_r - 2T_{N,Src} - T_{Loss} \tag{3.14}$$

during restoration, when disorder occurs, while there is no duplication, its duration is given by

$$T_{Dis} = T_r - 2T_{Src,N} \tag{3.15}$$

Using the previous equations, the duration of the transient degradation for each of the four protection methods described in the previous section may be straightforwardly derived.The network time diagrams are very useful to determine the pertinent emission times of either working or backup packets.
For example, Fig.3.14 shows such a time diagram in the case of multicast source stripping with 1:1 protection method, with a failure happening at t_f. It is clear from Fig.3.14 that the nodes located after failure position (F) suffer from packet loss during the failure recovery. In this example, $T_{Src,F} < T_{F,Src}$. Thus, $T_e^{LW}(N) = -T_{Src,F}$ and $T_e^{FB}(N) = T_{Src,F}$ as, in 1:1 source stripping method, after the source is notified, the backup flow is transmitted on the alternative direction. Using equation (3.9), we obtain the loss duration for the nodes after the failure in 1:1 source stripping method: $T_{Loss} = 2T_{F,Src}$. Accordingly, the duration of all transient degradation factors for the protection methods, are calculated and summarised in Table4.1.

The results of Table 4.1 show that all degradation times values have worst case value equal to T_R. However, the analytical form of these results does not

easily show the variation of the degradation factors experienced in a given node as a function of the positions of this node and of the point of failure. The analysis of the results is greatly facilitated by using the graphical representation provided in Figure 3.15 and Figure 3.16.

Figure 3.15 corresponds to the source-stripping method and shows, for a given couple of (node, failure) positions, the type of degradation experienced by the node during failure recovery and restoration phases. For instance, it makes clear that 1+1 protection results in less occurrences for packet loss than 1:1 protection, but may lead to duplication in some configurations.

Figure 3.16 illustrates the results obtained using the MuD approach. It is important to note that the maps shown in this figure are plotted in the (node position, drop-point position) space and not in the (node position, failure position) space. This is consistent with the results of Table 4.1, which show that degradation factors in the MuD scenario depends on the drop point position D but not on the failure position. More precisely, the failure still has an obvious impact because the type of degradation will depend whether the failure affects the working or the backup flow received by the source, but the value of the degradation time will not depend on the precise location of the failure. Another noteworthy element is the MuD point position. This position has to be chosen as close as possible to the half way of the multicast flow. Thus, the areas of interest in both maps shown in Figure 3.16 is the central horizontal strip limited by dashed lines.

All techniques for protecting multi-cast traffic that we considered in this work are characterized by short degradation times, bounded by the ring round-trip time T_R, regardless to the type of degradation (packet loss, disorder or duplication). These times can thus be less than 3 ms for 500 km ring. This common feature results from the main hypothesis that we made, namely the ability to generate and process OAM messages during each time slot and a time slot duration much shorter than T_R.

A further simplification has been made to derive the formulas given in Table 4.1, by neglecting the time for generating the notification messages inside the nodes situated at both extremities of a failed link. In practical implementations, such messages would be likely generated after missing a few time slots, instead of only one. But these adjustments would not significantly modify the fact that the maximum degradation times are close to the propagation time limit.

The different protection methods slightly vary in terms of performance. Most applications will not be sensitive to these variations. But it still worth noting that in several cases the optical packet layer behaves very close to an ideal lossless physical layer (in presence of a single failure).

In the source stripping 1:1 protection method, destination experiences packet loss if the failure occurs upstream on the Working direction. This occurs in 50% of single failure situations (assuming an uniform distribution of node and

failure positions in the ring), as shown in Figure 3.15. The main advantage of this method is its simplicity in triggering failure recovery and restoration procedure. Moreover, the protection resources can be used to transmit best effort traffic in normal operation mode.
When using the source stripping 1+1 protection method, contrary to the unicast traffic case, loss may occur for receiving nodes that do not receive "backup" packets because they have not yet received the OAM failure notification message.

However the probability of experiencing loss and the duration of this loss are both smaller than the corresponding indicators in the 1:1 protection method. For example, Figure 3.15 shows that the loss may occur in 37.5% of the single failure situations.

The MuD protection methods that we have presented in this work, are based on a (per multicast source) segmentation of the network. Both 1:1 and 1+1 MuD protection methods result in a reduction of the probability of experiencing packet loss, when compared to the source stripping case. For 1:1 MuD protection, loss will occur in a residual number of situations, which vanishes when the drop point is exactly located opposite to the source. In the MuD 1+1 protection case, loss never takes place. In both cases, duplication or disorder will be observed in nearly 50% of the situations. Another noticeable difference between the MuD and source stripping cases, is that, in the former, the degradation times will depend, for a given MuD position, only on the receiving node position and not on the position of the failure. The difference between the performances of source-stripping and MuD protection methods is illustrated in Figures 3.17 and 3.18, for a ring with 9 nodes. The first node is the multicast source, while the other nodes are registered as destination of the multicast flow.
In these figures, the nodes are characterized by their normalized position along one reference direction of the multicast flow. The failure is assumed to take place between the third and the fourth destination node. The source-stripping scenario is compared with the multicast drop point one; in this latter case, drop in performed by the fifth destination node. The duration of the different degradation factors in the failure recovery phase (normalized to T_R) is given for all nodes impacted by the failure.

Figure 3.17 shows the temporary loss duration experienced by the different nodes. With 1+1 protection, the MuD method results in no loss and, with 1:1 protection, an extremely low loss for a couple of nodes. Transient duplication and disorder durations are shown in Figure 3.18. MuD and source stripping methods behave similarly regarding duplication duration, while there is no disorder in the MuD case, in agreement with the results of Table 4.1. In case of MuD protection, the network operator has a full knowledge of the various degradations that may affect multicast traffic, including the cases where no loss is experienced (in the 1+1 MuD protection case) as shown in Figure 3.16 and Figure 3.17.
The above considerations are reasonable when the MuD, and thus the source of multicast flow, are well identified. This is actually the case for IPTV sup-

port where the corresponding multicast flows will likely be sent by a single source node in the ring. Thus, the MuD approach seems well suited to IPTV applications.

On the other hand, source stripping is simpler to configure for regular multicast traffic, where each node can be as source, as it limits the configuration burden for the operator who does not have to configure a MuD per potential source node. Clearly, as noted previously, the needs of current applications can easily tolerate degradation times up to T_R. However, for future highly demanding applications, other types of protection based on the mentioned architecture are interesting to be considered, since they offer the possibility of quasi-lossless protection.

Conclusion

We have a presented a dual ring optical packet ring architecture based on the POADM ring concept. The dual ring structure enables implementing protection mechanisms. We first considered unicast traffic. For this traffic, we proposed a class-based approach; premium and regular classes mechanisms are mimicking classical 1+1 and 1:1 techniques. A major difference with respect to protection in either circuit (e.g. SDH)- or packet (e.g. RPR)-based ring is the ability to send OAM messages during each time slot, which enables faster reaction times in presence of a failure. Moreover, when a node switching mechanism is necessary (for regular or possibly best effort traffic), this mechanism is operated at the packet rhythm (in a directionless packet-OADM) and thus extremely fast. The performance of these mechanisms has been evaluated in the framework or shortest path routing (i.e. minimum distance, but not necessarily minimum hop number). In these conditions, it is easy to derive upper bounds for the amount of lost traffic for both premium and regular mechanisms. The loss is very limited, and independent of the ring circumference for premium traffic. We have then presented an efficient full protected multicast service for a bidirectional time slotted WDM ring. Two variants have been discussed, relying either on conventional source stripping or on an original Middle Drop Point stripping. The performance study shows very low packet loss in all cases (and practically lossless in the MuD approach) and the maximum delay is less than a whole cycle. The Middle Drop Point stripping approach is practical for services relying on a specific multicast source, which is the case for demanding applications like IPTV.

Bibliography

[1] Sivarajan K.N Ramaswami R.

[2] L. Sadeghioon, A. Gravey, and P. 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, 2012.

[3] Bogdan USCUMLIC, Lida Sadeghioon, Annie GRAVEY, and Philippe GRAVEY. The Cost of Traffic Protection in Bidirectional Optical Packet Switching Rings. In *ISCC 2013 : 18th IEEE Symposium on Computers and Communications*, Split, Croatie, 2013. 13505 13505.

[4] Uščumlić B. Cerutti I. Gravey A. Gravey P. Barth D. Morvan M. Castoldi P. Optimal dimensioning of the wdm unidirectional ecoframe optical packet ring. *Photonic Network Communications*, 22:254–265, 2011.

[5] L. Sadeghioon, P. Gravey, and A. Gravey. Reliable multicast on a wdm optical packet ring. In *Next Generation Internet (NGI), 2012 8th EURO-NGI Conference on*, pages 103–110, 2012.

Figures and tables

Figures

Tables

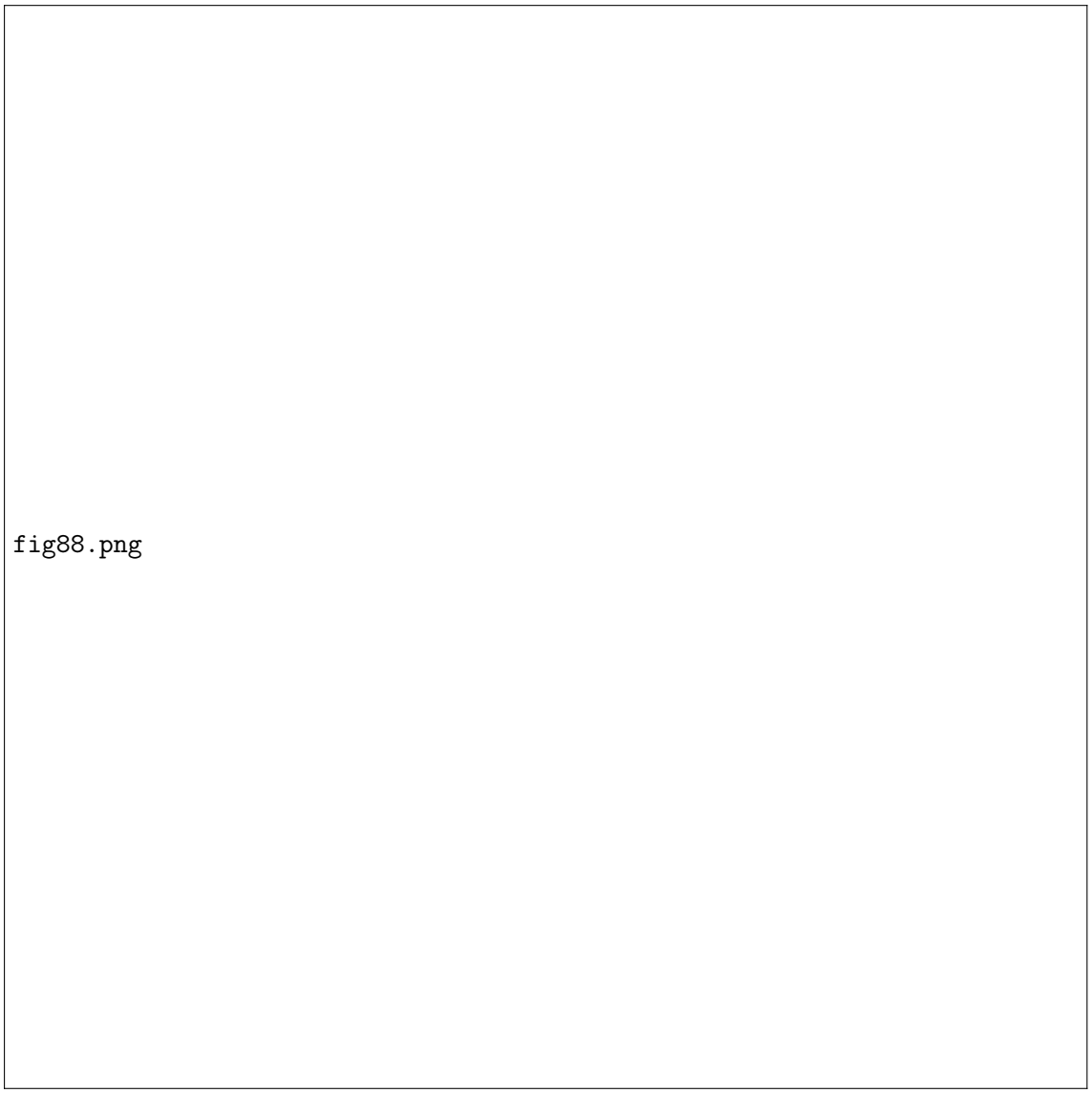

Figure 3.6: Design cost for different protection levels and $\theta = 40\%$

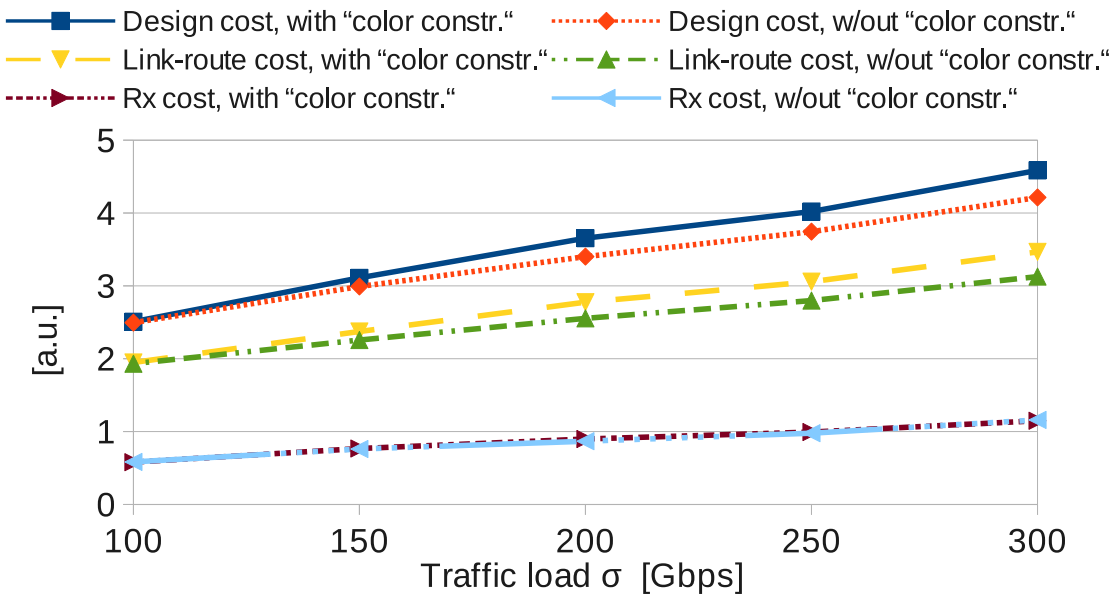

Figure 3.7: Evaluating the impact of "color constraint" for regular traffic and $\theta = 40\%$

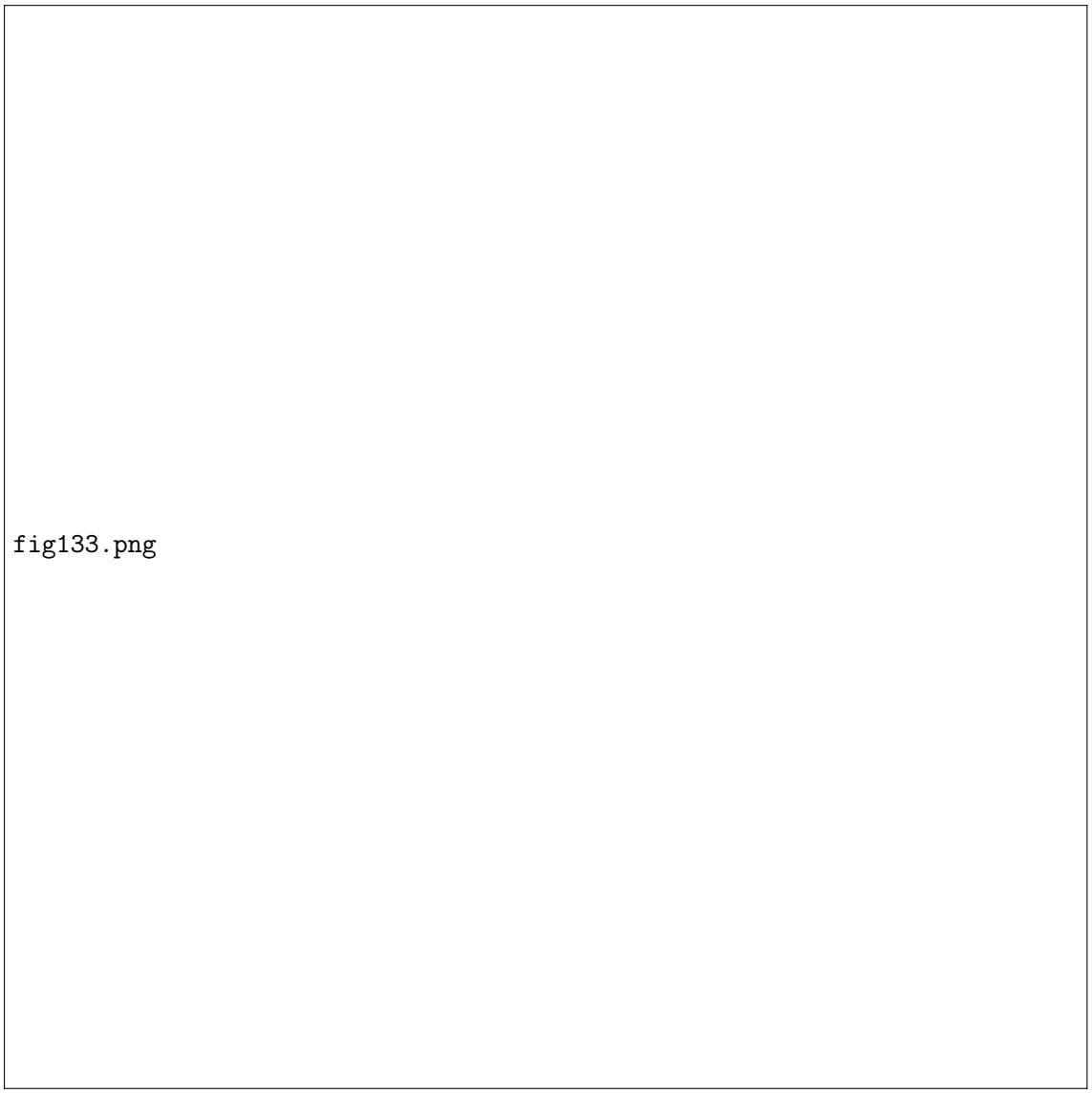

Figure 3.8: Design cost for different protection levels in uniform and symmetric scenario

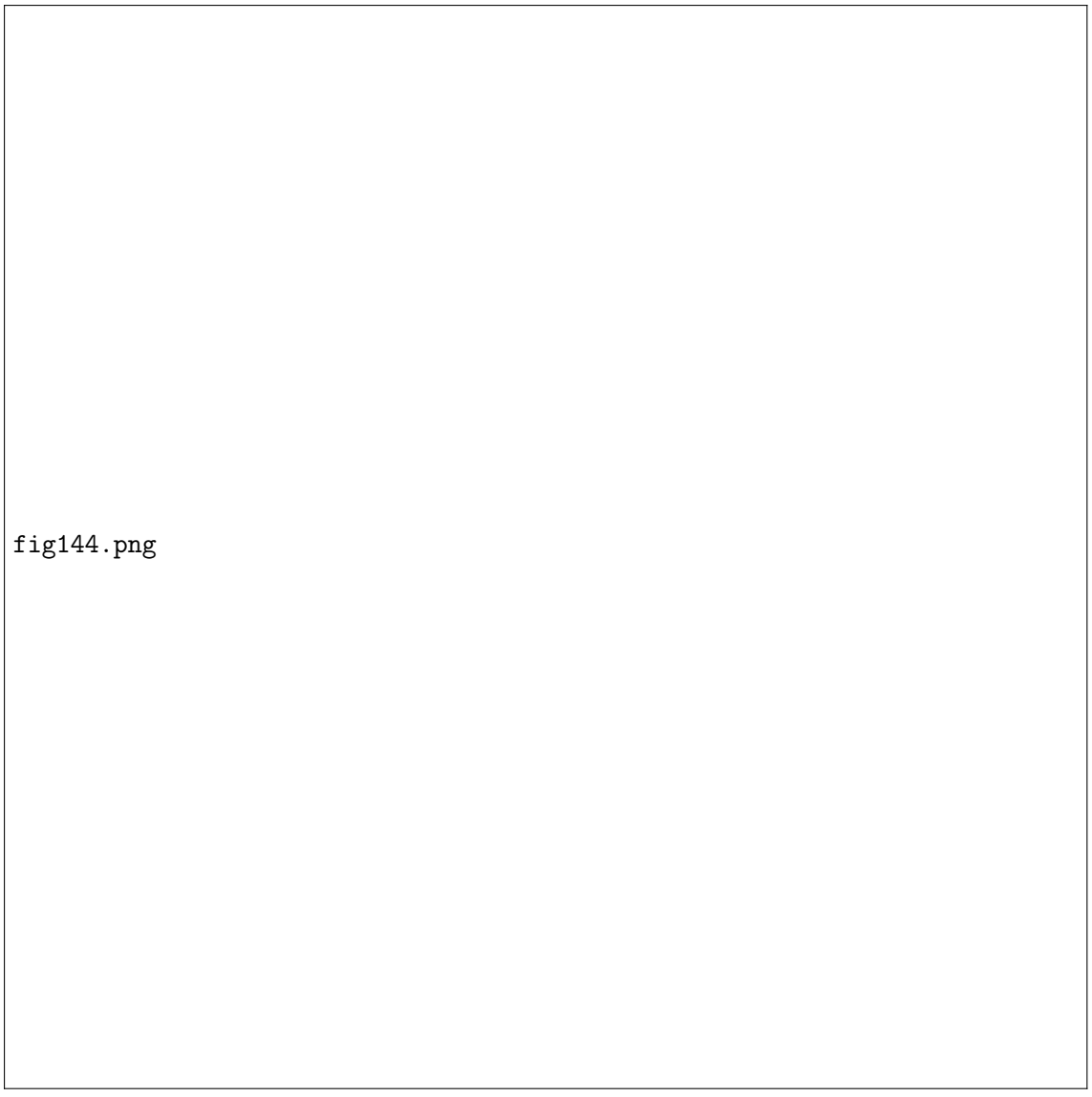

Figure 3.9: CAPEX cost components and wavelength number in solution for different protection levels

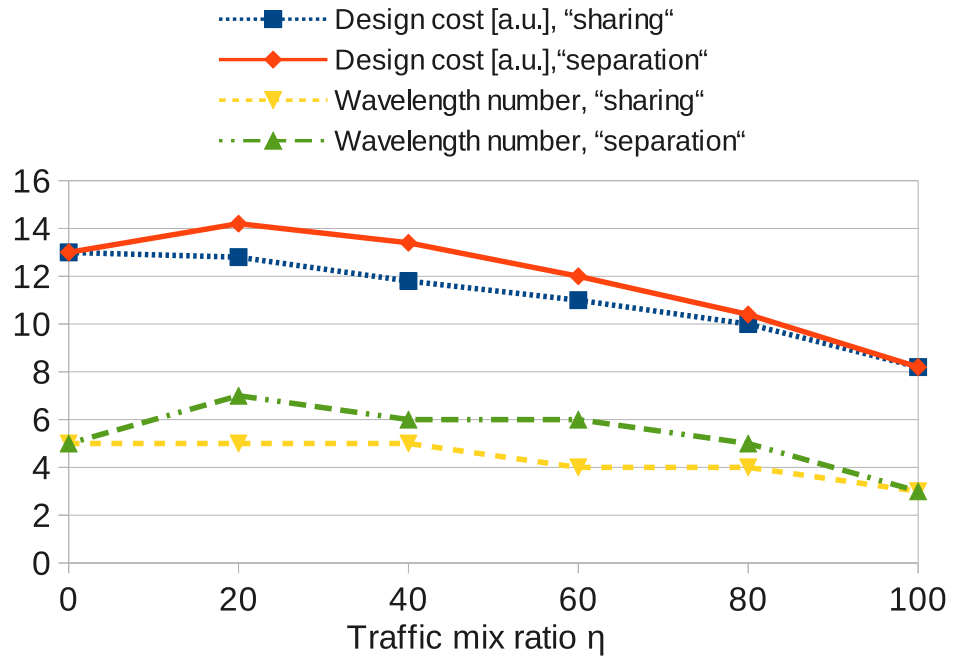

Figure 3.10: Mixed traffic scenario (Premium + Regular) for $a = 0.42$

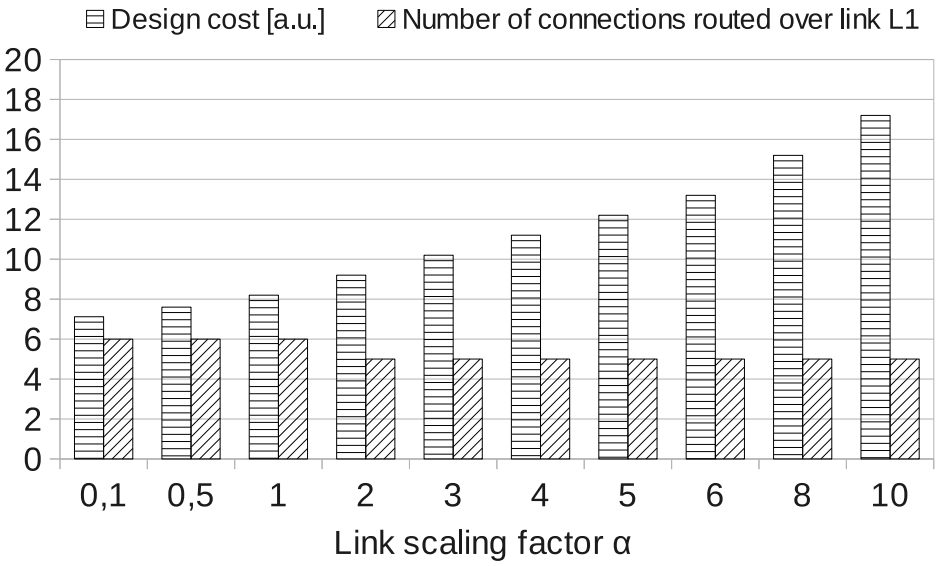

Figure 3.11: Impact of link size change on the network configuration

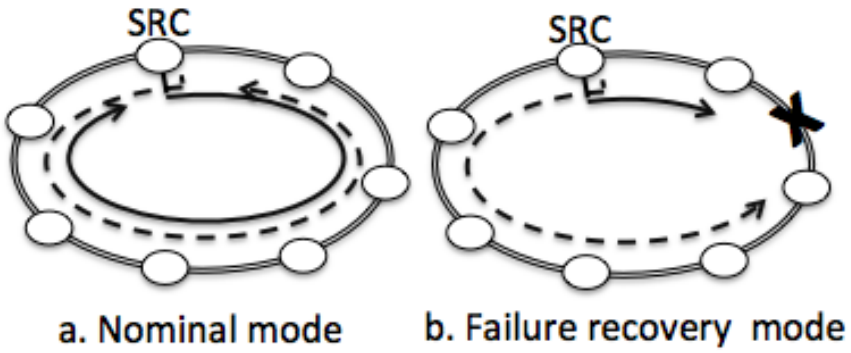

Figure 3.12: Bidirectional ring with 1+1 multicast source stripping protection mechanism.

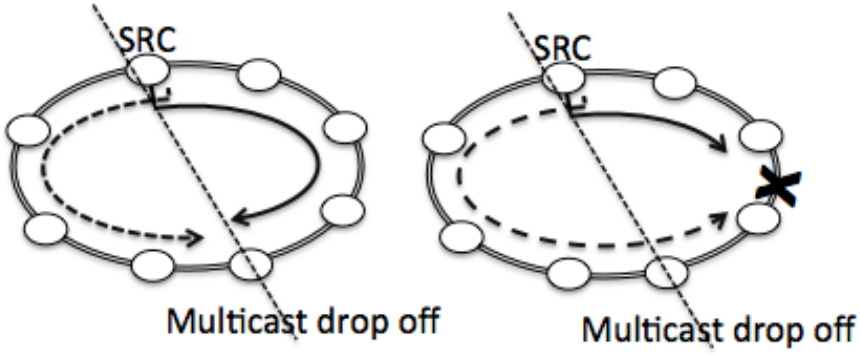

Figure 3.13: Dual ring Multicast segmented scenario in nominal operation and protection mode.

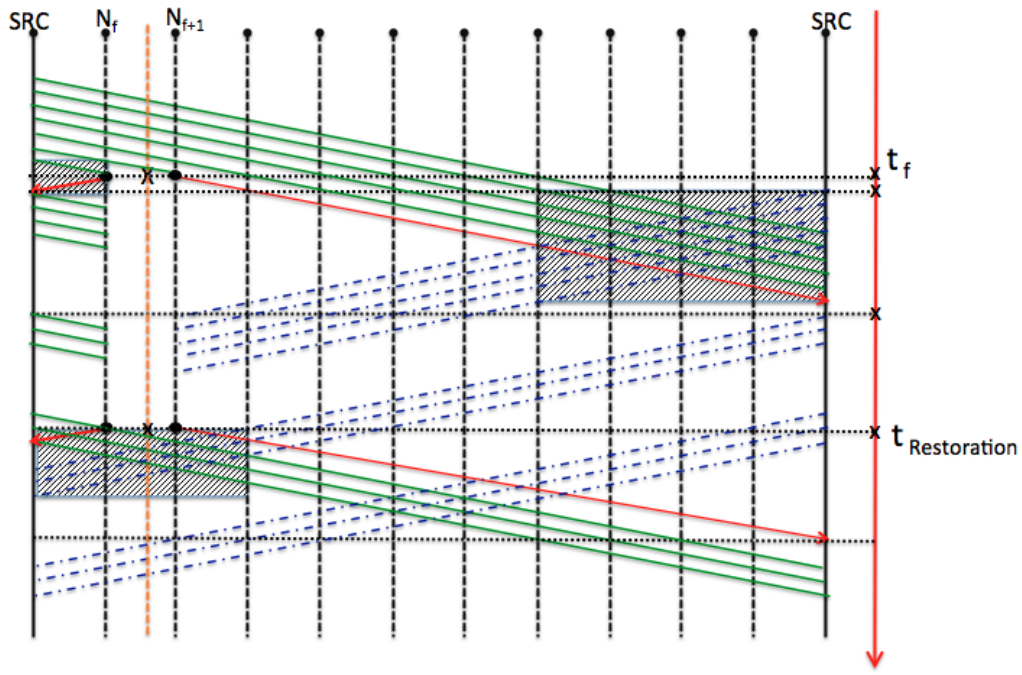

Figure 3.14: Time diagram of 1:1 Source Stripping Protection Method Multicast Service.

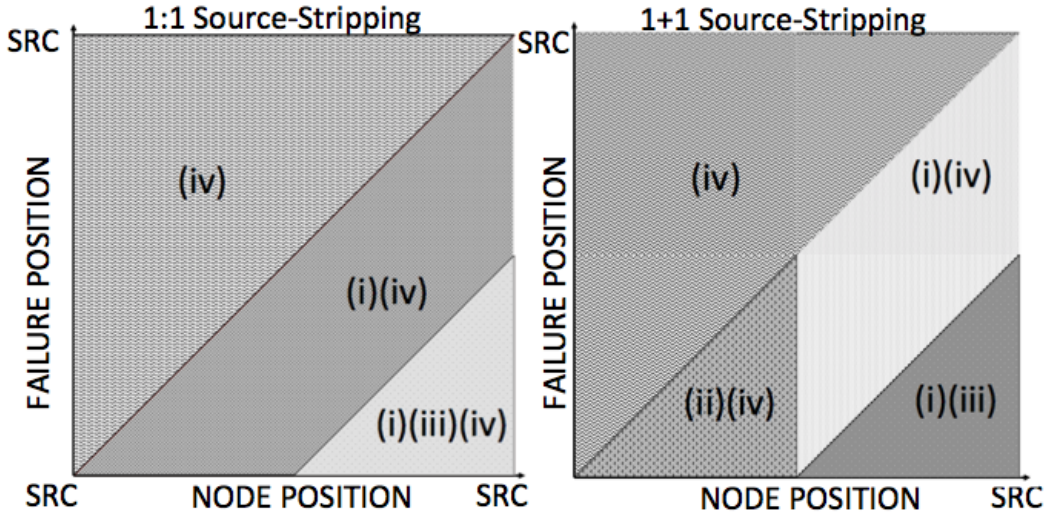

Figure 3.15: Degradation factors map for 1:1 and 1+1 source stripping, as a function of receiving node (horizontal axis) and failure (vertical axis) positions. Both axis are oriented along the working direction (i), (ii), (iii) stand respectively for loss, duplication and disorder during failure recovery phase, while (iv) and (v) stand respectively for duplication and disordering in the restoration phase.

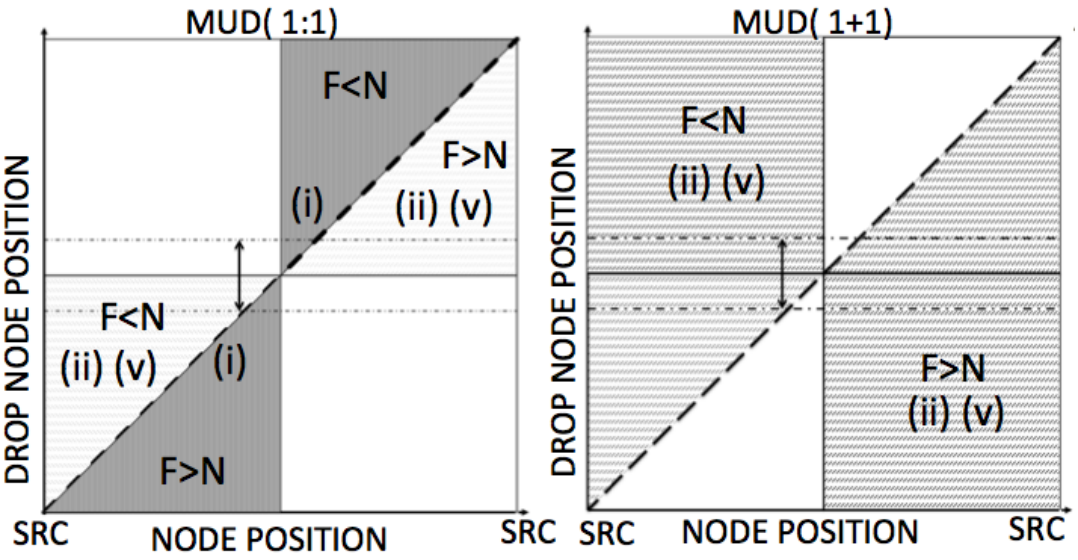

Figure 3.16: Degradation factors map for 1:1 and 1+1 MuD protection, as a function of receiving node position (horizontal axis) and MuD position (vertical axis). F and N indicate the failure and Node distance to the source node, (considering one ringlet direction as the reference direction). Both axis are oriented along the working direction (i), (ii), (iii) stand respectively for loss, duplication and disorder during failure recovery phase, while (iv) and (v) stand respectively for duplication and disordering in the restoration phase. When the failure is downstream (respectively upstream) the receiving node in the working direction, the relevant portion is the area below (respectively above) the first diagonal.

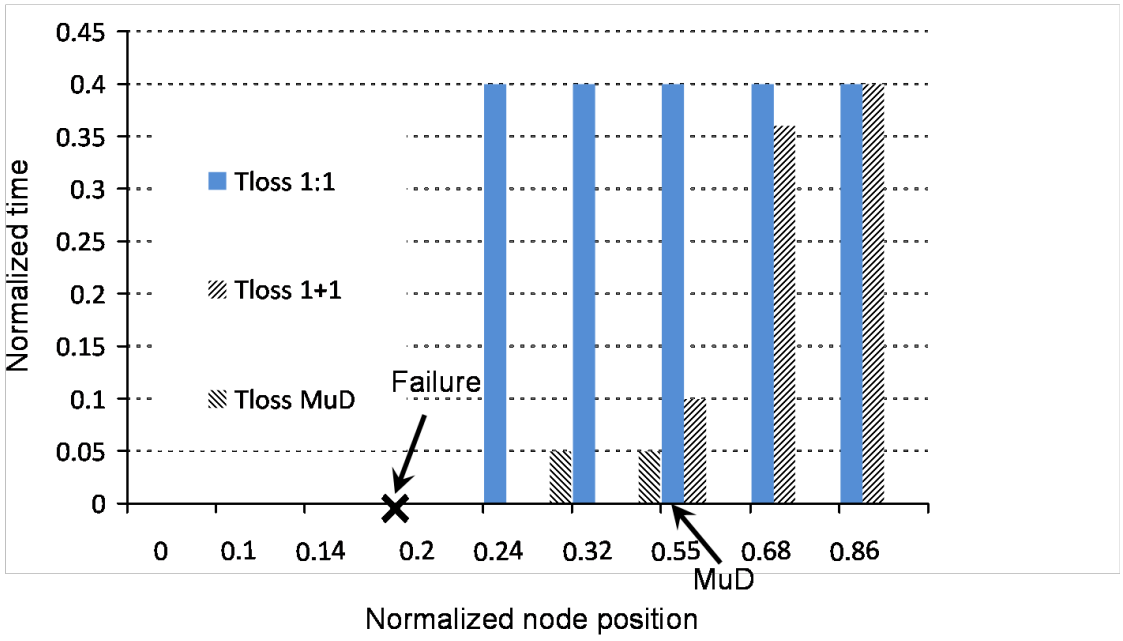

Figure 3.17: Duration of temporary Loss for 1:1 and 1+1 source stripping and 1:1 and 1+1 MuD protection methods for multicast flows. The horizontal axis scale is normalized to the ring length, and the vertical axis is normalized to the round trip time.

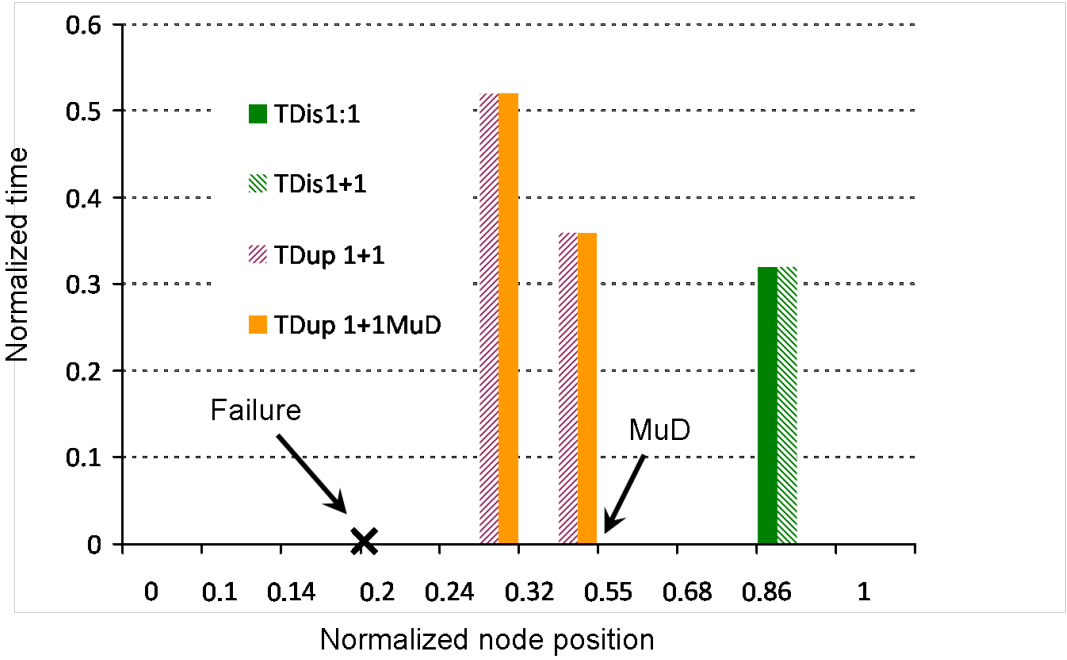

Figure 3.18: Duration of temporary degrading factors, Disordering and Duplication, for 1:1 and 1+1 source stripping and 1:1 and 1+1 MuD protection methods for multicast flows. The horizontal axis scale is normalized to the ring length, and the vertical axis is normalized to the round trip time.

Chapter 4

Optical Transparency in advanced topologies for metropolitan area networks

Contents

Introduction

The focus of this chapter is to offer an optically transparent transport network that supports sub-wavelength granularity. In the previous chapters we have studied ex- tensions of POADM rings. This network relies on fast tunable transmitters, fast optical switches and no optical buffering within a single ring. However in ring inter- connection there is still a need for temporary optical/electrical buffers to transmit the inter-rings traffic.

In this chapter, the first part proposes a multi-granular all optical hybrid (POADM plus OADM) network. An algorithm is then defined to recognize the optimum topology (of an arbitrary input topology) configuration complying with the switch- ing technology according to the traffic matrix. The result of the algorithm leads to a fully transparent multi granular network including POADM rings and a semi meshed ROADM based optical circuit switched network.

In the the second part of this chapter we consider another sub-wavelength granular technology, Time Wavelength Interleaved Network (TWIN)that offers end to end optical transparency and sub-wavelength granularity with no scalability prob- lem.

I Hybrid Optical Packet/Circuit Switched Network Design

There are plenty of research on optical hybrid packet/circuit switching techniques [1]. The introduction of multi-degree ROADMs is major element in favor of such approach, compared to a pure OPS or OBS network. Indeed, in many situations, in particular in core networks, it would make little sense to replace all ROADMs by a new kind of optical packet-switched nodes. It is likely that optical circuit switched technologies stay expensive and power consuming than optical packet switched technologies. Although the so-called "packet optical transport nodes", where ROADM and electronic switching functions (with, e.g. OTN circuit switching and Ethernet packet switching) are now widely deployed.
In the following section we propose [2] an all optical hybrid packet/circuit switched network combining POADM and RODAM functionalities. Obviously, any attempts to implement a transparent optical network will have to face the limitations resulting from optical impairments along optical paths and reach limitation. However, these issues are less critical in a metro-regional network, that is the scope of this study.

I.1 Multi-granular Node Structure

We propose a hybrid node structure based on POADM and CD-ROADM integration capable of both optical packet and circuit switching within optical transport layer. The goal here is to combine both benefits of the sub-wavelength granularity and optically transparent grooming capability, in addition to robustness of optical circuit switched technologies and flexibility to place the node within any topology configuration, see Fig 4.1.

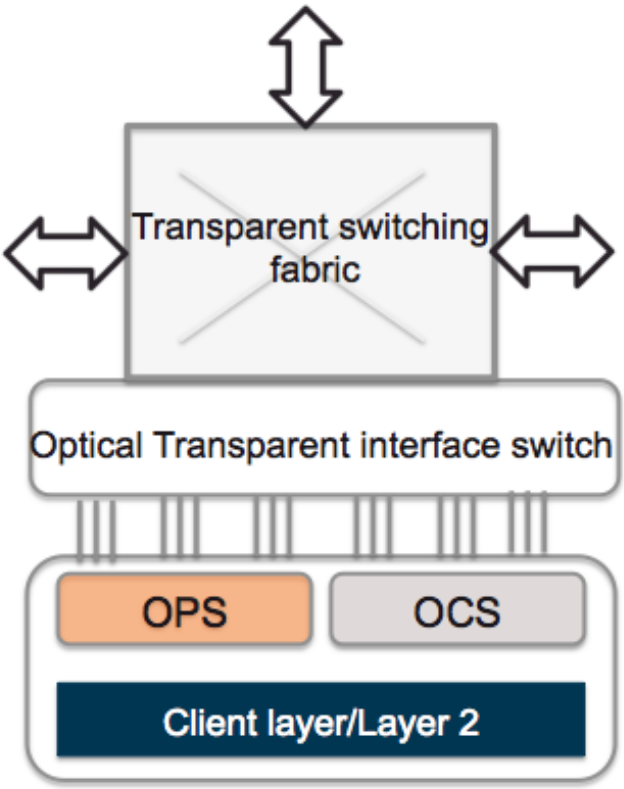

Figure 4.1: Optical Hybrid Node Structure Function Blocks

The OPS block represents the POADM structure that offers optical packet add/drop architecture by deploying SOA based fast switches within a pair of wavelength MUX/DEMUX. The transparent switching block represents a direction-less add/ drop structure. The node is equipped with M tuneable transmitters each connected to MxM WSS to be able to transmit at any wavelength and to all directions. At receiver side the optical transparent interface switch demonstrates a MxN optical switch that is connected to fixed WDM array of N receivers that makes the node capable of receiving from all directions.

The colourless functionality is provided only per wavelengths groups, that means for OCS and OPS separately. Moreover if OPS technology is based on tunable transmitter like in Packet-OADM rings, then colorless operation is automatically ensured.

I.1.1 Topology Design for Optical Hybrid Network

Overlay rings problem on physical mesh topologies for POADM based network has been studied in [?].
Here we defined a procedure (COHYB), to design a hybrid all transparent network, first by dividing the input traffic matrix between OCS and POADM technologies, and then by finding the solution with minimum POADM CAPEX

cost,(see Algorithm 1).

Algorithm 1 COHYB: Choosing Optical packet switching rings for Hybrid network

Require: $G(V,E), T, B, W, C_r, C_w, f_g, D$
  Construct set $\mathcal{R}$ of all possible rings on G;
  for $t_{ij} \in T$ **do**
    $t_{ij}^{OPS} = 0$;
    if $\lfloor t_{ij}/B \rfloor < f_g$ **then**
      $t_{ij}^{OPS} = t_{ij} - \lfloor t_{ij}/B \rfloor$
    end if
  end for
  for $r \in \mathcal{R}$ **do**
    $\tau(r) = \emptyset$;
    for $t_{ij}^{OPS} \in T_{OPS}$ **do**
      if $P(t_{ij}^{OPS}) \in r$ **then**
        $\tau(r) = \tau(r) \cup t_{ij}^{OPS}$
      end if
    end for
    $c(r)$ =POADM_OPT$(\tau(r), r, C_w, C_r)$;
  end for
  Construct set $\mathcal{S}$ of D randomly chosen ring coverings;
  for $S \in \mathcal{S}$ **do**
    $C_S = \sum_{r \in S} c(r)$
  end for
  $C_{S_{min}} = C_\alpha$, $\alpha \in \mathcal{S}$;
  for $S \in \mathcal{S}$ **do**
    if $C_S < C_{S_{min}}$ **then**
      $S_{min} = S$
      $C_{S_{min}} = C_S$
    end if
  end for
  return(S_{min});

COHYB operates on a physical mesh topology, where all links are considered to be bidirectional and to have a separate fiber for each direction. In each link direction, a maximum capacity of $W \cdot B$ can be supported, where $W = 80$ is the maximum number of wavelengths and B is a single wavelength capacity.

Apart from W and B, the input parameters for the algorithm COHYB are the physical mesh topology $G(V,E)$ defined by a set of bidirectional edges E and a set of vertices V, the input traffic matrix T, the threshold f_g and the algorithm depth D (explained below), and the CAPEX cost of components in POADM rings.

These costs include the cost of a POADM single wavelength receiver (C_r) and the wavelength leasing cost per km of a link (C_w).

The cost of tuneable transmitters in POADM rings is not taken into account as we have considered that each POADM node has the needed number of tune-

able transmitters in both ring directions. The costs of optical circuit switching technology is also not considered, as COHYB focuses on ring covering problem.

In the first step, the procedure finds a set $\mathcal{R}$ containing all the rings that can be physically constructed on the topology G.

In the next step, the traffic matrix T_{OPS} that contains only the flows that will be transported via optical packet rings, is built. Only a part of flow $t_{ij} \in T$ that is greater than integer number of wavelength capacity and that is smaller than threshold f_g is kept in the new matrix T_{OPS}.

After finding the matrix T_{OPS}, the algorithm COHYB calculates for each candidate ring $r \in \mathcal{R}$ a bidirectional traffic matrix $\tau(r)$, according to shortest path routing, that contains only the flows that can be transported by links of the ring r.
Here, $P(t_{ij}^{OPS})$ is the path of connection t_{ij}^{OPS}, i.e. a set of links on which it is routed within the ring r.

In the same loop, COHYB calculates the optimal CAPEX cost for the ring r based on $\tau(r)$, by resorting to the optimal dimensioning solution for the unidirectional POADM ring without flow splitting, from [3]. This solution is called with POADM_OPT in COHYB and as the input it takes $\tau(r)$, the receiver and the wavelength cost. For actual dimensioning with POADM_OPT, a traffic matrix with 1+1 protection, calculated on the base of $\tau(r)$ is used.

Next, a set $\mathcal{S}$ of D ring coverings of the hybrid network is created. A ring covering $S \in \mathcal{S}$ is a minimum set of rings r, elements of $\mathcal{R}$, such that all the traffic connections from T can be transported by the rings from S. The ring coverings S are chosen randomly among the possible ones. The "depth" parameter D is chosen to be $D \leq |\mathcal{R}|$, i.e. to limit the complexity of the algorithm, as the number of possible ring coverings is $|\mathcal{R}|!$. Introducing parameter D also allows the user of the algorithm to trade the calculation time with the price of the solution.

Based on dimensioning results for each ring in $\mathcal{R}$, each ring covering is attributed a cost C_S, and a covering with minimum value of this cost, S_{min} is chosen for the final solution of the algorithm.

I.1.2 Numerical Results

Fig 4.2 shows an example of a mesh topology which is derived from Deutsch-Telecom network with 7 nodes. We consider three different traffic profiles for three scenarios to be examined on this network.

The first two scenarios contain two hotspots with the same total network load and traffic density, while the third one has no hotspot or concentration point. The number of possible rings is 15 on the topology given in Fig. 4.2 (up-left). The rings are listed in Fig 4.2 (down) and are noted with R_i. Some possible ring coverings, as seen by COHYB algorithm are illustrated in Fig 4.2 (up-right).
All the scenarios have a total load of 1000 Gbps for the input traffic matrix, the capacity per wavelength is considered to be 10 Gbps, and the threshold f_g is set to 7 Gbps.
In all the results, the cost of POADM single wavelength receiver C_r is set to 1 (in arbitrary units), while the cost of wavelength C_w is considered to be 10

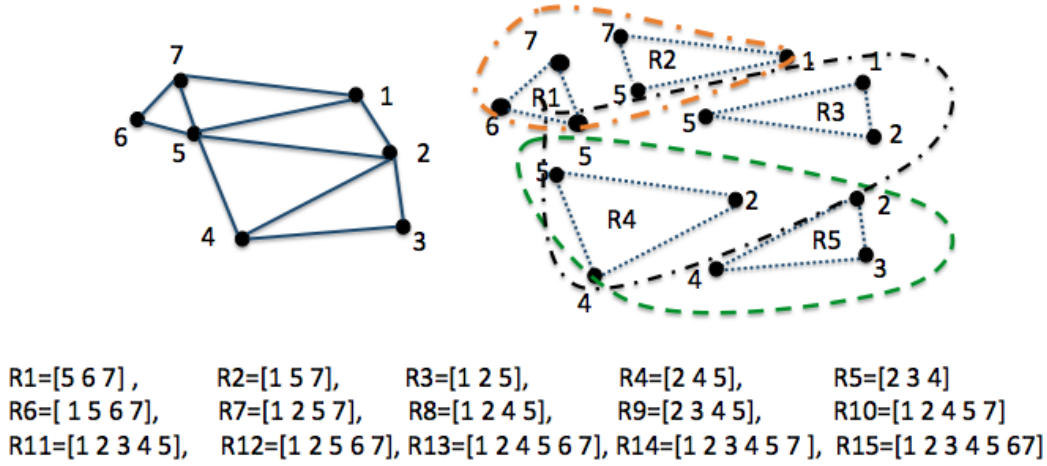

Figure 4.2: 7 Node semi-Mesh network topology derived form Deutsch-Telecom,15 Ring combinations are possible out of this topology.

times greater than cost of receiver for the link distance of 100 km.

Scenario 1	Scenario 2	Scenario 3
S1=R15	S1=R15	S1=R1
S2=R15+R1	S2=R15+R12	S2= R1+R15
S3=R15+R12	S3=R15+R12+R1	S3= R1+R15
S4=R15+R12+R5	S4=R15+R12+R5	S4= R3+R15
S5=R15+R12+R5+R1	S5=R15+R1+R5	S5=R4+R15
S6=R15+R5	S6=R15+R1	S6=R5+R15
	S7=R15+R5	S7=R6+R15
		S8=R7+R15
		S9=R8+R15
		S10=R9+R15
		S11=R10+R15
		S12=R11+R15
		S13=R12+R15
		S14=R14+R15

Figure 4.3: Different solutions (ring coverings) found by algorithm COHYB for Scenarios 1,2 and 3.

I.1.3 Scenario 1

It is supposed that there are two main hotspots in the network, located at nodes 1 and 5. In COHYB algorithm, the depth is chosen to be $D = 6$. Table in Fig. 4.3 for Scenario 1 shows the outcome of COHYB algorithm: there are 6 solutions, i.e. "ring coverings", all of which include the largest ring, R15. For instance, in solution S1, R15 provides a full connectivity (via optical circuits) and an optical grooming capability between all the nodes (by means of OPS).

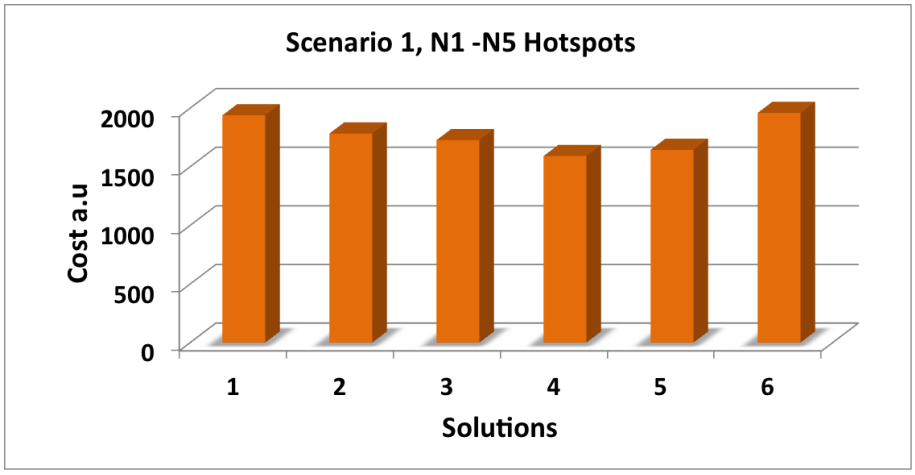

Figure 4.4: Cost of POADM ring design in Scenario 1

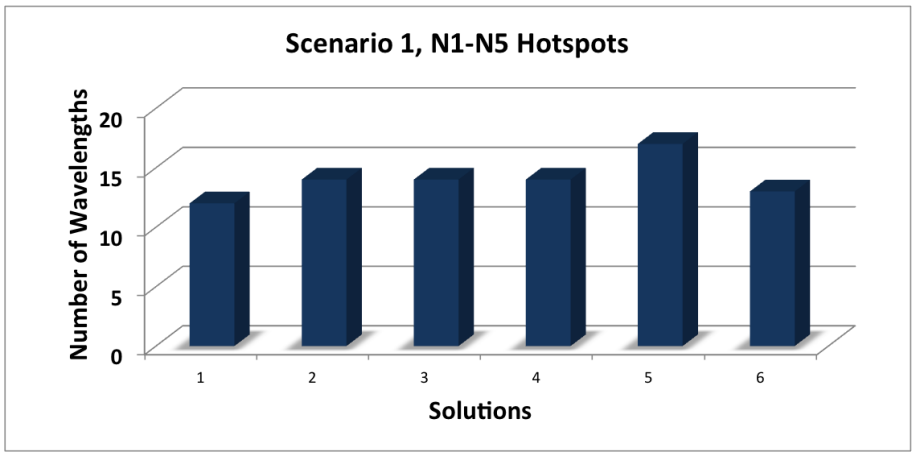

Figure 4.5: Number of wavelengths in POADM ring design in Scenario 1

As it is shown in Fig.4.3, R15 remains in all candidate solutions, as the input traffic matrix imposes a full optical connectivity across the network for some nodes.
The cost of different solutions in Scenario 1 is shown in Fig. 4.4. According to the figure, the cheapest solution is the ring covering S4 that contains rings R12 and R5.

R12 and R5 are the local rings adjacent to the hotspots, covering an area of dense traffic below the threshold. Having these rings in the final solution is obviously the cheapest option for this network configuration.

The overall number of wavelengths in different solutions for Scenario 1, considering threshold f_g equal 7 Gbps is shown in Fig. 4.5. The minimum number of wavelengths in a solution is 11, found for ring covering S1. However, in this solution the wavelength cost is very expensive, as only the largest ring R15 is used. Solution S4 does not achieve the minimum in number of wavelengths (it

has 13 of them), but its overall CAPEX cost design (including receivers and wavelengths in different rings) is the cheapest. These are the total number of wavelengths for POADM solution part of the network that are added to the total number of wavelengths.

I.1.4 Scenario 2

This scenario has a high traffic density area around nodes N1 and N4, in which hotspots are located. In COHYB algorithm, depth is chosen to be $D = 7$. The solutions found by the algorithm are listed in Fig. 4.3 for Scenario 2. The cheapest solution is achieved for ring covering S2 (Fig. 4.6). The traffic exchanged with node N4 is in a high portion switched by circuits, while the majority of traffic with a source/a sink at N1 is below the grooming factor f_g, which explains the result.

Note that the solutions S5, S6 and S7, containing rings R15, R1 and R5 have a much greater cost, which is partly a consequence of the ring topology, an issue that will be discussed in more details a bit later.

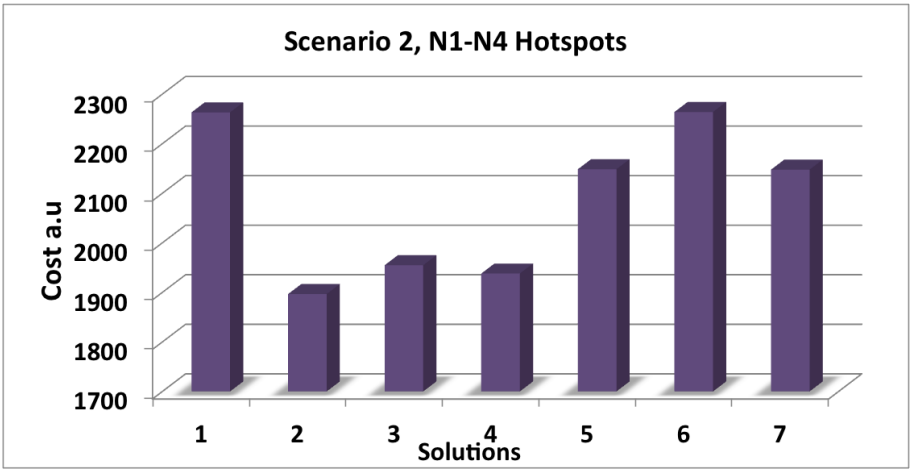

Figure 4.6: Cost of POADM ring design in Scenario 2

In Fig. 4.7 we present the overall number of wavelengths in different solutions for this scenario.
Solution S1 has the minimum number of wavelengths (11), but a high overall cost, due to a high wavelength cost.

I.1.5 Scenario 3

In Scenario 3, a case of "any to any traffic" is studied. Here, it is supposed that traffic is uniformly and symmetrically distributed between different nodes of the starting 7 node topology. This scenario is interesting for the comparison

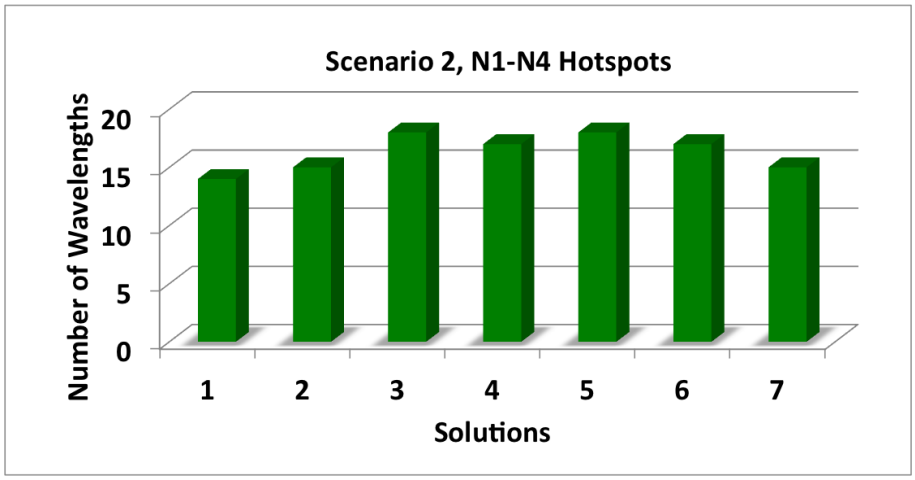

Figure 4.7: Number of wavelengths in POADM ring design in Scenario 2

purposes with the previous examples, as it represents a theoretical case of evenly distributed traffic.

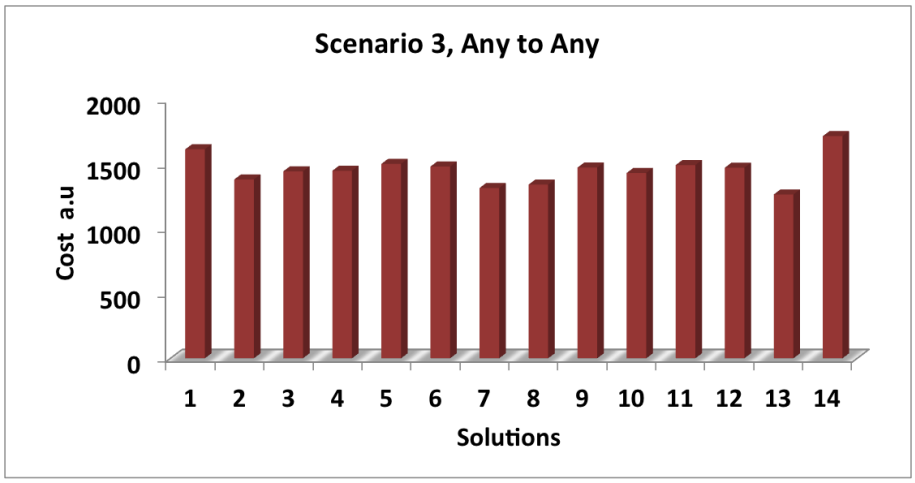

Figure 4.8: Cost of Packet-OADM ring Design in Scenario 3

The solutions found by the algorithm are listed in Fig. 4.3 for Scenario 2. The algorithm depth is set to $D = 14$ in this example.

Fig. 4.8 shows the POADM ring dimensioning cost for the solutions in this scenario.

It could be noticed that the cost differences between different solutions are not as expressed as in the previous scenarios, which is a consequence of the chosen traffic profile.
The minimum cost is achieved for solution S13, which contains the largest ring, R15, and a single nested ring, R12. The number of wavelengths is the same in all solutions and equal to 10.

I.1.6 Impact of the Traffic and the Link Size on the Cost

In Fig. 4.9 we examine the part of traffic carried by optical circuits and compare it with traffic carried by optical packet rings.

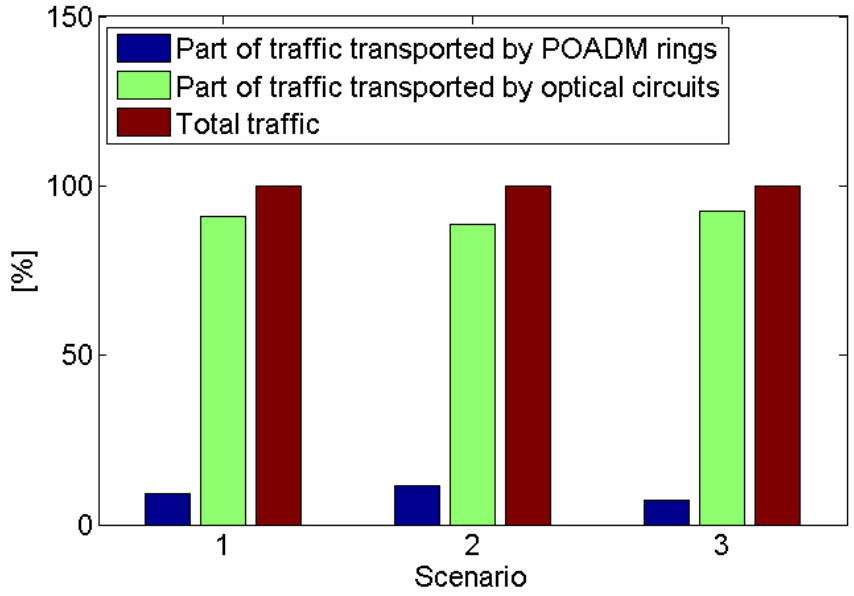

Figure 4.9: Participation of OPS and OCS traffic in the final solution

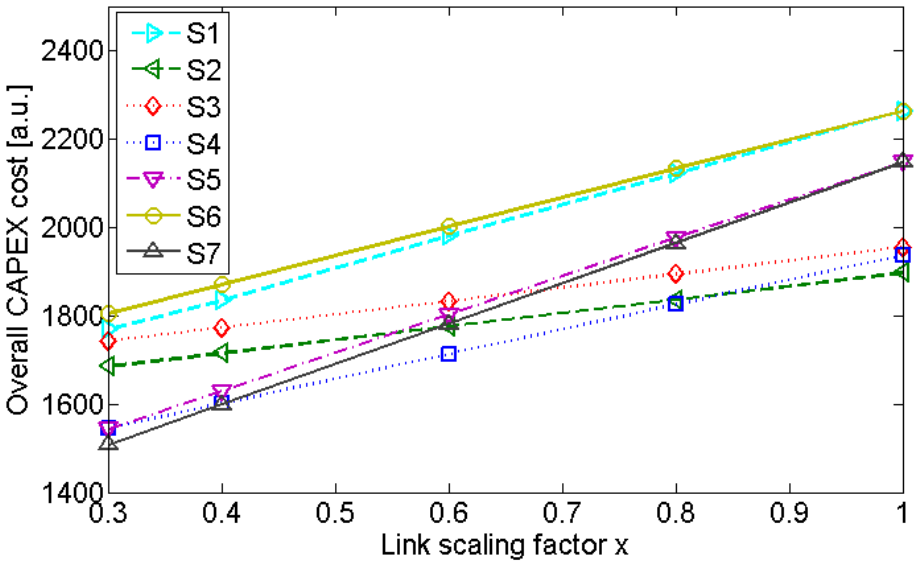

Figure 4.10: Impact of distance on ring covering choice in Scenario 2

According to this figure, the OPS traffic is always around 10%. On one hand, this value depends on the threshold f_g which was set to 70% of the wavelength

capacity in this study. On the other hand, as optical circuits transport the entire part of each traffic connection above the threshold, its high participation in total traffic is also a consequence of the input traffic matrix, in which many flows were above single wavelength capacity.

The physical distance between different nodes in hybrid network has an impact on the cost of ring coverings, as the cost of wavelength leasing in a ring linearly depends on its size.

To illustrate the effect that a change of link size has on the final solution, we have simulated again Scenario 2, but this time for a smaller size of links belonging to ring R5. The size of R5 is scaled-down by the link scaling factor x ($x \leq 1$), and the cost of different ring coverings is compared in Fig. 4.10. As some links are shared between rings R5 and R15, a decrease of ring circumference in R5 leads simultaneously to a reduction in size of ring R15. For small values of x the cheapest solution is S7, for middle values it is S4, while for higher values of x the cheapest solution is S2, confirming that the final solution strongly depends on the physical configuration of the network.

I.2 Conclusion on Hybrid Optical Packet/Circuit Network Design

In this section , we presented a dimensioning method for a hybrid all-optical network design. It provides the minimum cost OPS rings, that provide optical traffic grooming where it is needed. A novel heuristics is proposed for regional transport network planning, that divides the input traffic on the portions to be transported by optical circuits and POADM rings, in addition to optimizing the number of wavelengths and fixed single wavelength receivers needed in such rings. The optimization CAPEX cost includes the receiver cost and the wavelength leasing cost per km of link distance. The heuristic has a given "depth" parameter, which allows a tradeoff between the calculation time of the algorithm and the precision of the solution.
The results for three different traffic matrices, prove the interest for optimizing the topology by ring coverings in a hybrid transport all-optical network.

They also show that the OPS traffic has up to 10% of total traffic share in the network, and that the physical topology has an strong impact on the final solution choice.

II Time-Domain Wavelength Interleaved Network Control and Protection

TWIN technology as mentioned earlier in chapter I, is one of the sub-wavelength networking technologies that provides granular transaction between any two nodes in the network all optically with minimum switching capability referring to WDM passive switching devices. This technology is relying on sending and receiving grant and reports based on the demand at the sources and scheduling performed at source or a central entity to be able to access the medium and provide collision free access. [4] shows a comparison between centralised

and distributed access, resulting in a higher performance of centralised control method. In the following we propose a separate control access mechanism, considering the benefit of centralized control in performance of TWIN network.

II.1 Virtual TWIN Rings for Control

In this section we present a dedicated control channel on a virtual control topology for transmitting scheduling, control and OAM messages based on centralized control.

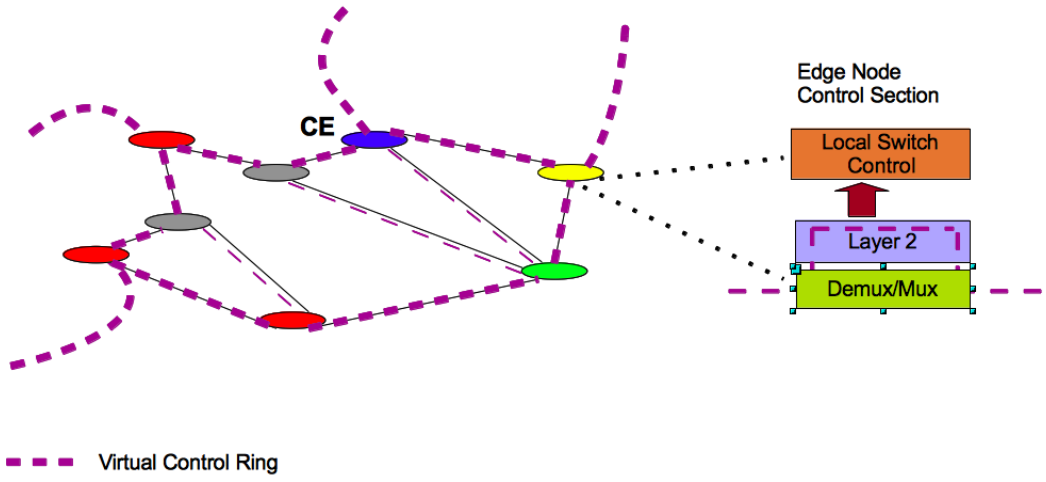

Figure 4.11: Virtual Control Ring on a Mech network

Fig.4.11 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 informations.
Thus all the node within the network can exchange the report/grant between the Control Entity (CE) information and all the other nodes, in addition to other control and management messages via this dedicated path that we call it Virtual TWIN Ring (VTR).

To offer VTR for control it is necessary to consider the scalability in terms of geographical distance and number of nodes in the ring. In section II.4 we propose an algorithm to find the VTR appropriate for any arbitrary topology [5].

II.2 Propagation Delay Calculation

In order to be able to provide a non-blocking scheduling in TWIN, the CE needs to know about distance that is the propagation delay between each source-destination pair. The propagation delay essentially depends on the path that

is chosen on the network topology, should the path change under any circumstances such as link failure, the distance needs to be recalculated.

The CE calculates the propagation delay between itself and all other nodes virtual control ring in addition to the propagation delay between any node pair in the network on the actual working paths [6].

To do so, the control entity exchanges the control messages with timestamps with network nodes, by performing the following a 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.

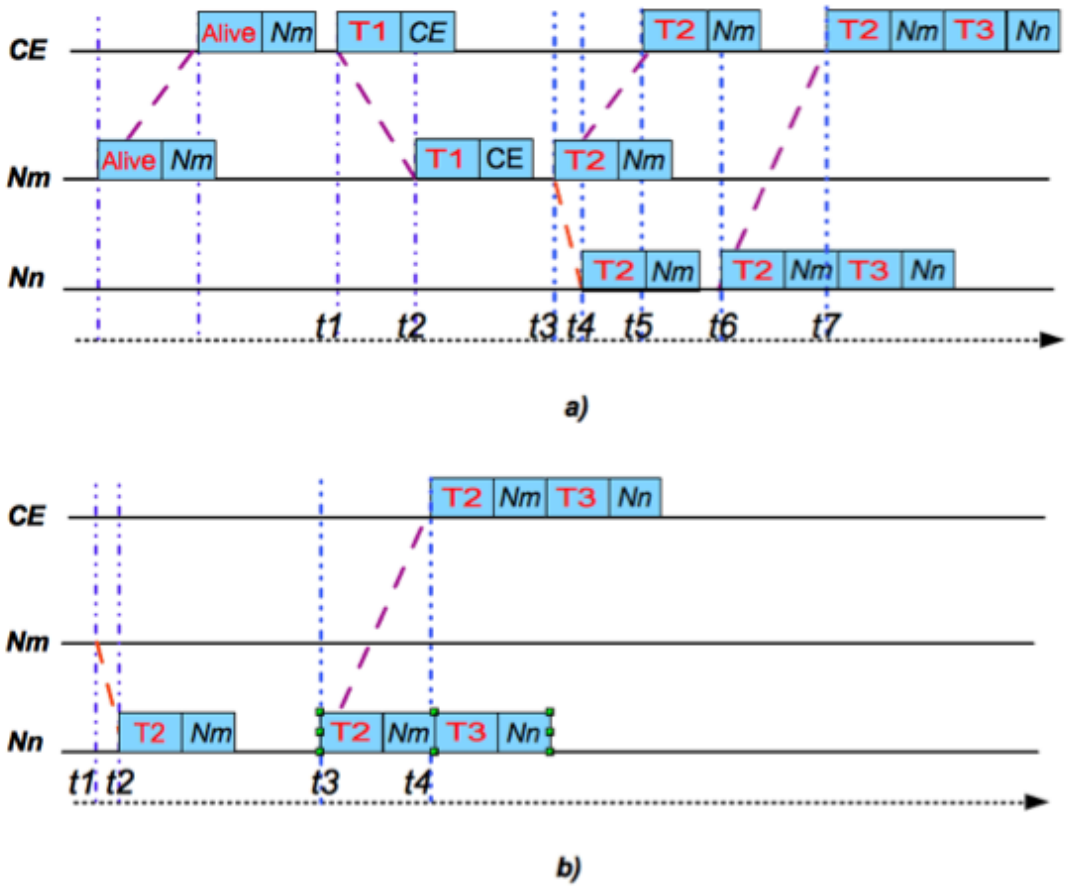

Figure 4.12: Propagation delay calculation

Fig4.12 shows the time diagram of the control messages exchange when:

1. **Calculating the propagation delay between the control entity and the node N_m, marked with $t_{PD}(CE, N_m)$, and**

2. **Calculating the propagation delay between the nodes N_m and N_n, in direction from N_m to N_n (marked with $t_{PD}(N_m, N_n)$).**

Here, $N_m, N_n \in \mathcal{N}$. 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), $t_{PD}(CE, N_m) = t_{PD}(N_m, CE)$, and this is valid for all nodes N_m.

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

- **Node N_m sends the "alive" message on the control ring, and by using the control wavelength, to the CE to initialize its network clock.**
- **CE sends the clock reference $T1 = t_1$ to node N_m on the control ring.**
- **N_m receives the timestamp at the absolute time t_2 and sets its own clock to the CE local time $T1$.**

In the second part of the procedure (Fig.4.12), consider nodes N_m and N_n. If both nodes have the correct value of the clock, the following steps are used to calculate the propagation times $t_{PD}(CE, N_m)$ and $t_{PD}(N_m, N_n)$:

- **Node N_m broadcasts the timestamp $T2$ to the node N_n via TWIN's data-paths (and not the control ring). $T2$ is calculated by node N_m as:**

$$T2 = T1 + (t_3 - t_2) \tag{4.1}$$

- **In addition, node N_m sends towards CE the clock reference $T2$ on the control ring. The propagation delay time $t_{PD}(CE, N_m)$, from CE to N_m, can be calculated as a half of the round-trip time:**

$$t_{PD}(CE, N_m) = [(t_2 - t_1) + (t_5 - t_3)]/2 \tag{4.2}$$

  From Eq. (4.2) and (4.1), we obtain the following result:

$$t_{PD}(CE, N_m) = (t_5 - T2)/2 \tag{4.3}$$

  The last equation is used at CE to calculate the propagation distance to N_m, since both values t_5 and $T2$ are known to CE.

- **Next, CE computes the time difference $\Delta t_d(N_n, N_m)$, i.e. the difference in the nodes' local clock, that is equal to the difference between propagation time from CE towards the nodes:**

$$\Delta t_d(N_n, N_m) = t_{PD}(CE, N_n) - t_{PD}(CE, N_m). \tag{4.4}$$

- **Finally, N_n sends its local time $T3$ to the CE, as soon as it receives the message from N_m. In its message, it will include the timestamp $T2$ received by N_n.**

- **CE calculates the propagation delay time between node N_m and N_n on the working topology, by using the following formula:**

$$t_{PD}(N_n, N_m) = T3 - T2 + \Delta t_d(N_n, N_m) + (t_m^{\lambda_D} - t^{\lambda_C}) \tag{4.5}$$

Factor $(t_m^{\lambda_D} - t^{\lambda_C})$ is needed to account for the chromatic dispersion effects. It is supposed that CE has knowledge of this value, since it knows for the receiving wavelength of each TWIN node.

The procedure shown in Fig.4.12 is repeated for all the nodes and node pairs in the network, in order to determine the propagation distances in the entire network.

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

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

$$t_{PD}(N_n, N_m) = T3 - T2 + (t_m^{\lambda_D} - t^{\lambda_C}) \tag{4.6}$$

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

II.3 Protection on Virtual Vontrol Rings

Fig.4.13 shows the TWIN architecture working in the nominal operation mode with all nodes connected via multipoint-to-point TWIN trees. In Fig. 4.13 b) sources S1, S3, are sending traffic to D1 via G1. S2 path is passing through S1,G1. Here G1 is a passive optical switch. It is possible to assume that all the nodes have this passive wavelength switch functionality throughout the network. In Fig.4.13 c) and d) the path between S3 and D1 is broken each time at different link. Depending on which type of protection is chosen there are multiple ways to send the traffic between S3 and D1 such as: running a routing algorithm to find the second shortest path if the short distance is the routing constraint or any other possible criteria. Alternatively it is possible to store a disjointed path pair each source destination in a way that it does not change any passive node configurations. However we have to remember that it is necessary to minimise the need for dynamic switching in the core of TWIN based network for the alternative protection path.

We propose to use the virtual ring we introduce for the control purpose earlier, for data transport on the data wavelength in both nominal and failure recovery mode to avoid this complexity and need for reconfiguration. Fig 4.13 e) shows the network has become a self healing ring. with no need for reconfiguration.

In the case of failure, the following events take place between CE and nodes $S1$, $S2$ in order to restore the network:

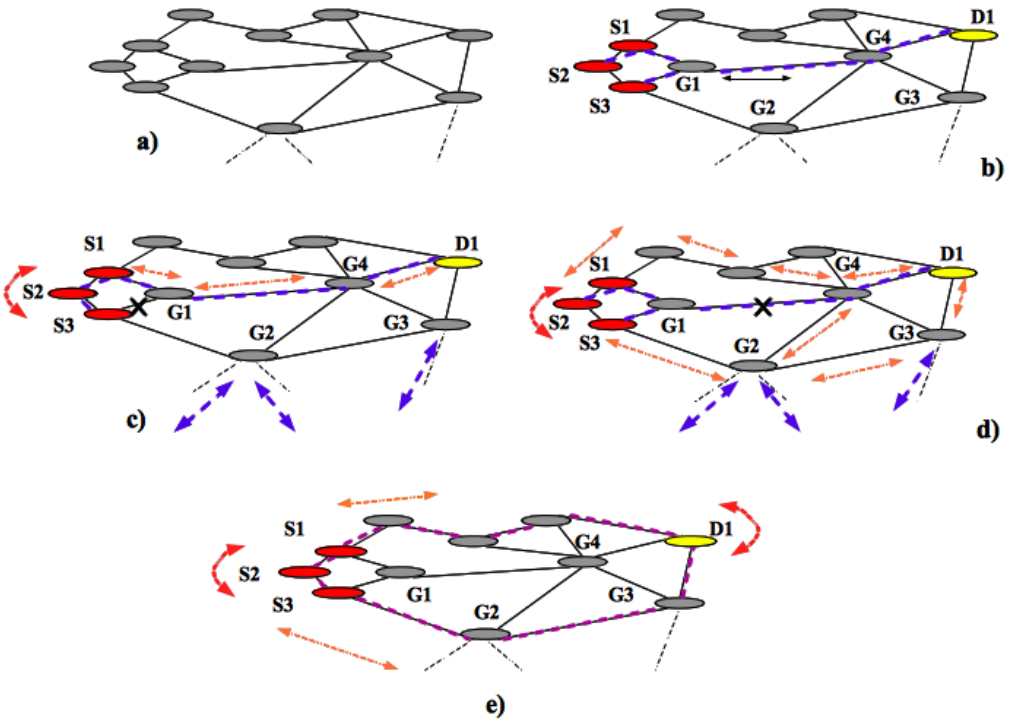

Figure 4.13: protection path options for different failure locations

1. **Failure Detection - optical signal loss is detected by fault manager entity of the two adjacent nodes near to both ends of the link failure in the WDM layer. The failure alarms generated in WDM layer is processed at control/ management layer to localize the link failure [7].**
2. **At CE, the propagation delays for the affected nodes are updated according to the earlier section described procedure.**
3. **Via failure notification message all the nodes are informed of the routes that are no longer available therefore previous grants are no longer valid.**
4. **At CE the new schedules and grants are generated in such a way that no reconfiguration at the intermediate nodes of backup paths is necessary. Considering that the backup paths are located on the virtual TWIN ring**
5. **Source switches on the alternative path on the ring, and uses the new grants.**
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 regarding the backup paths are no longer valid.**
8. **The CE generates the new grants for the new reports and the regular working paths.**

The proposed method is more efficient than a traditional protection scheme offering 1:1 link protection. The reduction in complexity is obtained at the cost of potentially increased propagation distance between some source destination pairs in the network. This effect is quantified in section II.5.

II.4 Algorithm

Here we present the VTR algorithm that has no limitation in terms of number of nodes, and allows the extension of the proposed concept to a general transport 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 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. The optical signal reach constraint is accounted by limiting the maximum size C of rings used for network covering. C depends on many parameters, in particular the modulation format and baud rate used at transmitter side [8]. Assuming a standard single mode fiber and a 50 GHz channel width, conservative values could be e.g. $C = 200$ km at 200 Gbsps, and $C = 1500$ km at 100 Gbps.

The VTR algorithm is based on a 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 be populated by resorting to the cycle finding algorithms, e.g. the one given in [9].

In the first step, the algorithm finds a referent "network covering" solution θ_{ref}. The network covering solution consist 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, VTR builds a set of all candidate solutions for network covering, ie. the set Θ of all subsets of $\mathcal{R}$, consisting 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 benefits of such approach are in reducing the algorithm complexity, but also in finding the network covering that is easier for administrations, 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 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 if sent over the other rings in θ_k, and then, the corresponding propagation distances between the node S_i and the outside nodes are calculated. When doing so, the maximum over both protection ring directions is taken. This is indeed necessary, since a link failure might force the protection flow to be sent to any of two directions. At the end of this step of VTR, for each ring R_i, the overall sum $\delta(R_i)$ of propagation distance is calculated. For all the interior nodes, the propagation

Algorithm 2 VTR: Virtual TWIN Ring

Require: $G(V, E)$, 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}. θ_{ref} is found for $|\theta_{ref}|$ ==MIN, break the loops; **then**
    end if
  end for
end for
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. $\Theta = \Theta \cup \theta_k$; **then**
  end if
end for
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 in the case of protection, $\Delta(S_i, D_j)$, 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 protection 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 for
    calculate the sum of propagation distance in the protection mode for 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 for
  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
for the final ring covering solution choose the set θ_k,

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

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, VTR 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 the 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.

II.5 Result and Analyses

To illustrate the operation of the VTR algorithm, in this section we consider the Deutsche Telekom network in Fig. 4.14, composed of 17 nodes and 26 links. In the figure all the link lengths are expressed in km.

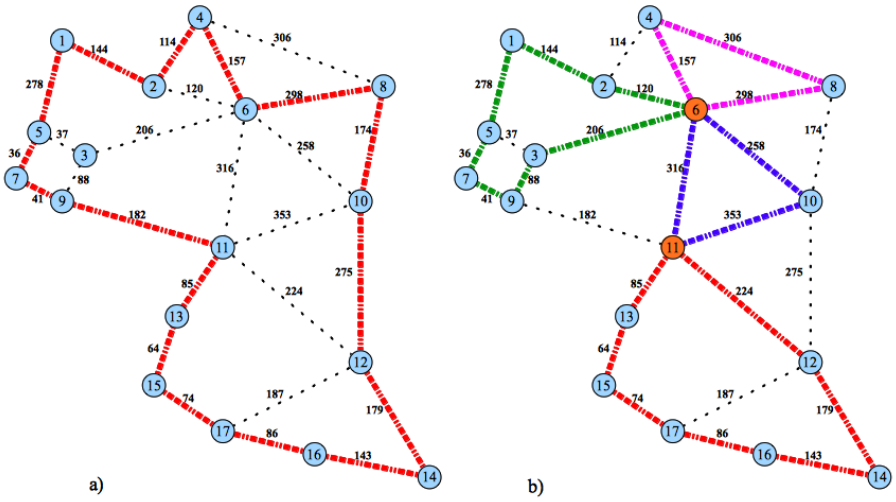

Figure 4.14: Deutsche Telekom network: a) covering with a single ring, b) optimal ring covering according to VTR algorithm

In accordance to the protection scheme we proposed, all the nodes should be covered by at least one bidirectional ring.The minimum single ring covering the entire network is shown in Fig. 4.14(a).
The size of this ring is 2330 km, which greatly surpasses the optical reach of signal at 100 Gbps, and is not, thus, an acceptable solution.

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

The set of candidate rings, found by VTR is given in table 4.1. There are 13 potential candidate rings, each of them with circumference limited to C. Table 4.1 shows the nodes contained by each ring, and their circumference in

Table 4.1: 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

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 rings needed to cover the studied topology. Table 4.2 summarizes the set of candidate solutions (with 4 rings) that the algorithm has found. The number of the candidate solutions is 13.

Protection with TWIN virtual rings simplifies the rerouting of traffic flows in the case of link failure in the network. On the other hand, some flows might be penalized with an increased distance of backup path, in comparison to a mesh network protection scheme, where a protection path would be chosen according to the second shortest path rule. The number of paths exceeding 1000 km, for each of the candidate ring coverings is given in table 4.2.

These results show that choosing θ_3 or θ_4 would have bad impact on the protection performance. We can see that the minimum number of protection 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. 4.14(b)) has the minimum number of source-destination paths exceeding $C = 1000$ km (as shown in table 4.2), so it is selected as the final result of the algorithm.

To get the insights on the increase of propagation distance brought by virtual ring protection in case of a single link failure in the network, we have compared two cases on the same topology:

Table 4.2: 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

1. **TWIN network protection is realized by traffic rerouting after the failure according to the shortest path rule, if the working traffic path has been affected;**

2. **TWIN network protection is provided by virtual ring protection (solution θ_9); backup traffic is rerouted by using the protection rings.**

The metric that is used for the comparison is the "maximum backup path length", which can be defined as the maximum distance that a backup route between any node pair in the network can reach, by considering all single link failures. Note that a single link failure is considered here in the context of the "link protection".

The results are shown in Fig. 4.15, for all network links. As expected, the decrease of the complexity coming from the use of virtual rings, comes at price of worse propagation distance.

The maximum relative difference between standard mesh and virtual TWIN ring protection is achieved in the case of failure between nodes 11 and 13 (Fig. 4.14). The relative difference between two protection methods goes up to 27% in this case.
Note the maximum distance is less than 1400 km that is less than typical optical reach of 100 Gbps [8].

Fig. 4.16 shows maximum source-destination distance per node on TWIN tree topology and TWIN virtual rings according to solution 9. We can observe that only 2 nodes among 17 experience 45% of increase in maximum distance per source-destination pair for each node. For other nodes the maximum distance stays almost the same for both control and working path. Finally, we emphasize that the difference between control channel propagation path and data channel in nominal operation mode does not affect the performance of the system operation.

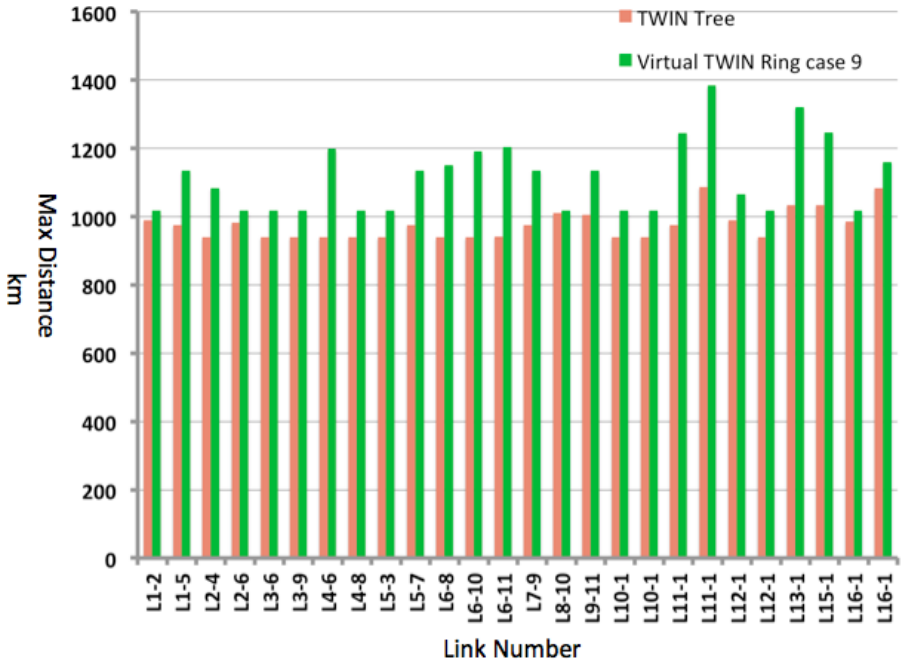

Figure 4.15: Maximum backup path length in case of "link protection": TWIN tree protection (orange) vs TWIN virtual ring protection (green) in solution θ_9.

II.6 Conclusion on TWIN Control and Protection

In section II we have proposed a new method for the control and the protection in a TWIN based optical burst switching network. The method facilitates TWIN's control and operation with an out-of-band control channel, carried on one or multiple virtual rings 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, all of them exchanged between the nodes and the central control entity. The proposed protection scheme for TWIN relies on the virtual rings as the alternative path in case of failure, while the working 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 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. By using the algorithm, the mentioned features are ensured with the shortest propagation distance on the protection rings, cal-

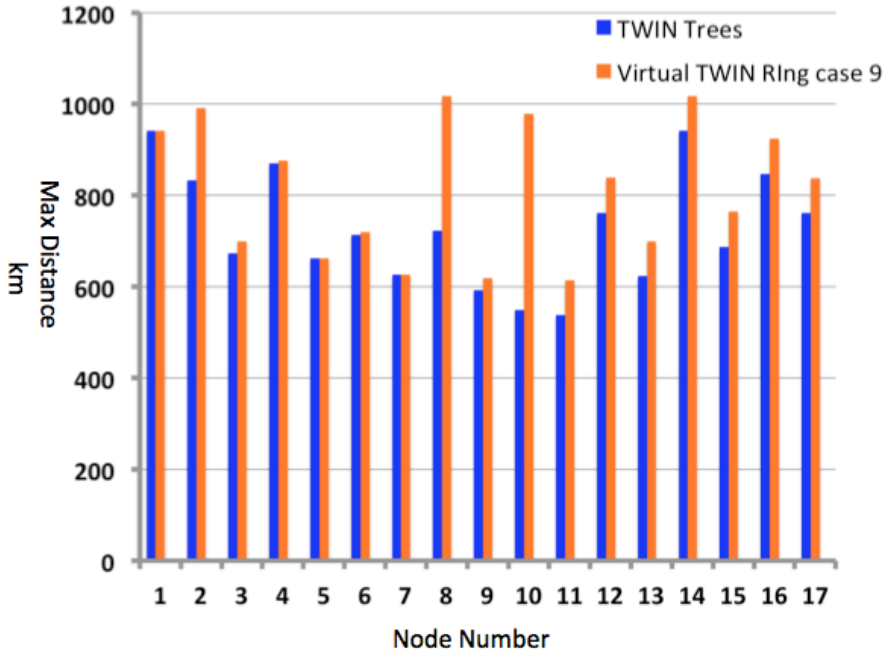

Figure 4.16: Maximum length on TWIN tree topology in nominal operation mode (blue) comparing to maximum length on virtual TWIN rings (orange) for control and protection path, solution θ_9.

culated over all source-destination pairs and for all possible single link failure events. 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 can cover all the nodes and ensure the protection with a minimum network reconfiguration burden, in comparison to the standard "link protection" solution. Of course, the simplicity of protection is achieved at the cost of increase in average propagation distance between the nodes (limited to 27 this example).

Conclusion

In this chapter we studied two type of all optical transparent network. The first network relied on a hybrid OPS-OCS. Since the OPS (POADM based rings) part performs all optical grooming and the OCS provided point to point optical light path we defined an algorithm to integrate the advantage of both technologies, since we intelligent planning we can deploy the technology where it is most appropriate and result into lower cost in total design. The second part we focused on TWIN technology, by defining a dedicated control channel on a virtual overlay topology. Then we observed that on the virtual ring topology we can also protect the network. On a large scale topology this solution gives a multiple interconnected rings as an overlay topology yet still all the path can be setup all optically due to the TWIN technology characteristics.

Bibliography

[1] H. Wang, A.S. Garg, K. Bergman, and M. Glick. Design and demonstration of an all-optical hybrid packet and circuit switched network platform for next generation data centers. In *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pages 1–3, 2010.

[2] L. Sadeghioon, B. Uscumlic, P. Gravey, and A. 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, 2013.

[3] Bogdan Uscumlic. Optical architecture and traffic engineering in optical metropolitan networks. In *PhD Thesis*, pages 1–6, 2010.

[4] Ahmed Triki, Paulette Gavignet, Bernard Arzur, Esther Le Rouzic, and Annie Gravey. Efficient control plane for passive optical burst switching network. In *International Conference on Information Networking (ICOIN)*, jan 2013.

[5] 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 European Conference on and Optical Cabling and Infrastructure (OC i), 2013 8th Conference on*, pages 121–128, 2013.

[6] L. Gravey P. Gravey A. Morvan M. Popescu, I. Sadeghioon. Synchronization of the time-domain wavelength interleaved networks.

[7] Y. Li, S. Ranka, and S. Sahni. Wavelength scheduling in time-domain wavelength interleaved networks. In *Computers and Communications (ISCC), 2011 IEEE Symposium on*, pages 519–524. IEEE, 2011.

[8] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. 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, Jan 2011.

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

Figures and tables

Figures

Tables

CONTRIBUTION TO THE DESIGN OF OPTICAL-PACKET BASED METROPOLITAN AREA NETWORKS

This study was an attempt to address the issues regarding combining the transparency and sub-wavelength granularity appropriately in optical metro network architectures by offering two main sets of contributions.

The first set of contributions **consists in extending the applicability of optical packet switching in MAN as a solution for reliable multi-service metro network providing sub-wavelength granularity on top of an infrastructure based on optically transparent rings. The Packet-OADM time slotted WDM unidirectional ring was chosen as the basis of the first study to acquire the adapting tools for multi-service provision and reliability. The first step was to extend the POADM unidirectional ring into a bidirectional one by extending the node structure and keeping the main principle of all optical transit for data channel at intermediate nodes. Thus POADM bidirectional WDM time slotted ring as the layout of this part of the work was founded on the following extensions:**

- **Two tuneable transmitters per node (one per direction)**
- **Separate sets of fixed WDM receivers per ringlet per direction.**
- **Out of band control channel per ring (per direction) that is synchronised by the data channels on that ring and direction.**

Then, we extracted the key elements to provide different features such as QoS, Multicast, and protection by defining identifier fields in the control packet content per each time slot. Moreover, we 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, we defined an original label based MAC offering multiple services relying on separate control packet messaging system capable of the following:

- **Identifying each flow by two level labels (SDU & PDU). This enabled us to extend the MAC protocol functionality for inter-ring packet intercon-**

nection using O/E/O conversion, and to propose two switching methods either at PDU or SDU level at interconnecting nodes. Moreover, labels are particularly suited to offer multiple services such as multicast and protection.

- **Defining two methods for insertion and extraction of multicast traffic.**
- **Using a reservation-based access method for best effort traffic. The performance of the method then was evaluated and it was shown that the impact of best-effort traffic on the guaranteed one can be neglected.**
- **Integrating OAM messaging system intro the control packet content and obtain a global multi-service MAC protocol.**

According to the developed MAC for POADM bidirectional multi-ring, we studied how to adapt the conventional 1+1 and 1:1 protection methods both for multicast and unicast traffic. At first, we proposed a novel flow based procedure to identify and localise the failure by defining a section for global notification messages in the control message, in order to broadcast the failure notification in distributed manner. In addition, we proposed a table structure in each switch to correlate the flow labels and the location of the failure. Then 1+1 (premium) and 1:1 (regular) class-based protection with packet granularity methods were studied and three main degrading parameters were identified as: packet loss, packet disorder and packet redundancy. Where the loss defect is completely suppressed with premium method, as expected, the cost of it is about twice of regular protection method. However combining the two methods can offer a practical solution to service providers. Finally we studied the premium and regular protection method for the multicast traffic. Two variants have been discussed, relying either on conventional source stripping or on an original Middle Drop Point stripping, that offers practically no loss. The performance study showed that the packet loss is very limited in all cases and the maximum delay is less than a whole ring cycle, which amounts to a few milliseconds in regional networks.

A second set of contributions, **concentrates on studying end-to-end optical transparency on complex topologies in addition to offering sub-wavelength granular capacity. In this study two different approaches were investigated:**

In the first one, we proposed a hybrid POADM/OCS architecture. POADM-rings are used to carry sub-wavelength granular traffic and offer efficient optical grooming where it is needed, while OCS network (with Colorless and Direction less ROADM) provides reconfigurable optical circuits at wavelength level where it is necessary. We proposed a node structure to combine the two technologies in order to benefit from the advantages of both of the technologies. In addition, we developed an algorithm to map an overlay topology based on the switching technology and a given traffic matrix. Then we derived a cost function and found the optimum solution considering the amount of wavelengths and the number of receivers per node for POADM technology. This method was applied to a 7 node meshed network. By splitting the traffic matrix according to a defined grooming factor, we managed to reduce the number of

wavelengths to four times less than if no optical grooming would be used.

In the second approach, we chose TWIN as another sub-wavelength granular technology that offers end-to-end transparency. We suggested a 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 has provided new insights on the applicability of transparent optical-packet based networks for metro applications. However, several issues deserve further investigations.

First, further work would be necessary in order to assess the respective merits of the optical packet switching based on POADM or TWIN and optical circuit switching. For instance, the structure of different nodes (e.g. the nodes interconnecting two POADM rings or the "passive" TWIN nodes, which could be OCS nodes) should be described. This would enable to compare the different approaches in terms of cost and power consumption by using realistic figures. Also, dimensioning tools for both TWIN and POADM have to be improved to include and QoS requirements (this issue is presently addressed at Telecom Bretagne) and by developing heuristic suited to large scale networks.

Secondly, the fast evolution that optical communications experienced in the last five years, with the progress of coherent detection and digital processing techniques may provide new solutions to offer flexible WDM channels (with multi-line rate transponders) and a sub-wavelength granularity by optical means, e.g. by using multi-band OFDM. These approaches could challenge optical-packet techniques and further studies would be required to compare flexible circuits and optical packets.

Finally, new paradigms, that were initially considered in upper layer networks or in wireless access networks, such as network virtualization, software defined networks and cognitive MAC will probably impact the architecture of future optical metro networks. For instance, the work carried on the MAC could be extended by developing an intelligent cognitive MAC allowing to adapt the resources to the application.