Université de Montréal

Protection partagée pour les réseaux de transport multidomaines

par
Thi Dieu Linh Truong

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Juin, 2007

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Protection partagée pour les réseaux de transport multidomaines

présentée par:

Thi Dieu Linh Truong

a été évaluée par un jury composé des personnes suivantes:

Abdelhafid Hakim
président-rapporteur

Brigitte Jaumard
directeur de recherche

Bernard Gendron
membre du jury

Alain Houle
examinateur externe

Jean-Francois Angers
représentant du doyen de la FES

Thèse acceptée le 20 septembre 2007

RÉSUMÉ

La protection des réseaux multidomaines joue un rôle très important étant donné les conséquences négatives qu'une seule panne peut engendrer, autant en termes de coûts financiers, que de l'envergure géographique des interruptions de services. Il existe un nombre important de recherches portant sur la protection partagée. Cependant, la majorité d'entre elles s'applique uniquement aux réseaux d'un domaine simple compte tenu de leurs hypothèses sur la disponibilité d'informations complètes ou partielles, mais toujours globales. Ces hypothèses ne sont pas réalistes et ne peuvent pas être satisfaites dans le contexte des réseaux multidomaines, en raison des exigences pour l'extensibilité. Très peu de recherches ont été effectuées spécifiquement pour les réseaux multidomaines. Parmi elles, on trouve des solutions de protection dédiée, ou partagée mais incomplète (qui laissent certains liens ou noeuds sans protection), ou valables seulement pour un type de réseau en particulier. Il existe donc un réel besoin quant à une solution complète s'adressant aux réseaux multidomaines génériques.

Cette thèse propose des solutions complètes de protection partagée par chemins et par segments se chevauchant ainsi que des solutions de ré-optimisation des chemins de protection existants permettant de compresser la capacité de protection. Nos solutions se basent sur des agrégations de la topologie du réseau multidomaines qui permettent d'obtenir une image simple et enrichie d'informations agrégées du réseau ainsi que sur des routages à deux niveaux en utilisant, au premier niveau, le réseau agrégé et, au second niveau, les réseaux originaux. L'utilisation des informations agrégées au lieu des informations complètes ou globales durant le routage rend nos solutions extensibles dans le contexte des réseaux multidomaines.

Chacune de nos solutions a été comparée avec les solutions de problèmes similaires dans les réseaux d'un domaine simple. Ces comparaisons pénalisent nos solutions puisqu'elles doivent satisfaire les exigences d'extensibilité ce qui n'est pas le cas des solutions concurrentes. En dépit de ce handicap, les comparaisons montrent que nos solutions ne sont pas loin de la solution optimale d'un domaine

simple, et sont parfois très proches. De plus, le temps de calcul, de l'ordre de quelques millisecondes, satisfont parfaitement les exigences pour une solution de routage dynamique. La quantité de messages à échanger entre les domaines est aussi limitée conformément aux exigences d'extensibilité.

Enfin, les solutions proposées sont génériques pour tous les réseaux multidomaines avec connexions à bande passante garantie tels que MPLS, ATM, SONET/SDH et WDM avec conversion de longueurs d'onde.

Mots clés: Protection, Réseaux multidomaines, Routage.

ABSTRACT

Protection of multi-domain networks is very important because of the highly significant impact of a single failure in terms of cost and geographical scope. Although there have been many studies on shared protection, most of them remain limited to single domain networks due to their requirements of complete or partial but always global information. These requirements cannot be satisfied in multi-domain networks because of the scalability constraint. Few researches have been conducted specifically for multi-domain networks. Amongst them, we find dedicated protection solution, incomplete shared protection solution which leaves some nodes or links unprotected, or a solution scheme that is valid only for a special type of networks. There is thus an actual need for efficient shared protection solutions which deal with general multi-domain networks.

This thesis proposes complete solutions for shared path protection, overlapped segment shared protection as well as solutions for re-optimizing existing backup paths leading to the reduction of the backup capacity. These solutions are based on different topology aggregations which allow obtaining a simple network, enriched with aggregate information; and two-step routings using the aggregate network at the first step and the original ones in the second step. The use of the aggregate instead of complete and global information in the routings makes our solutions scalable for multi-domain networks.

Each of our solutions has been compared with the solutions of similar problems in single domain networks. Our solutions are penalized in the comparisons as they have to satisfy the scalability constraint, by using aggregate information, while the concurrent solutions do not. In spite of this disadvantage, the comparison results show that our solutions are not far from, sometimes close to, the single domain optimal ones. In addition, the computational efforts, in the order of few milliseconds, are definitively appropriate to an online routing. The number of messages to be exchanged among domains complies with the scalability constraint.

Finally, the proposed solutions are general for all multi-domain networks with

bandwidth guaranteed connections such as MPLS, ATM, SONET/SDH and WDM with wavelength conversions.

Keywords: Multi-domain network; Protection; Routing.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES

ATM	Asynchronous Transmission Mode
BGP	Border Gateway Protocol
DWDM	Dense Wavelength Division Multiplexing
DYPOS	DYnamic Programming Overlapped Short segment shared protection
DYPOS-BGB	DYPOS with Blocking-go-back option
GMPLS	Generalized Multi-Protocol Label Switching
GROS	GReedy Overlapped Short segment shared protection
GROS-BGB	GROS with Blocking-go-back option
ILP	Integer Linear Programming
JDP	Joint Computing of Directive Paths
MaR	Map and Route
MaR-G	Map and Route-Greedy
MaR-O	Map and Route-Optimal
MPLS	Multi-Protocol Label Switching
MPLS-TE	Multi-Protocol Label Switching-Traffic Engineering
MSPP	Multiservice Provisioning Platform
O/E/O	Optique-Electrique-Optique
OSSP	Overlapped Segment Shared Protection
PNE	Programmation linéaire en Nombres Entiers
RaM	Route and Map
RRLocal	Reroute Local

RSVP	Resource Reservation Protocol
SCI	Sharing with Complete Routing Information
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical Network
SP	Shortest Path
SPP	Shared Path Protection
SRG	Shared Risk Group
SSP	Shared Segment Protection
WPF	Working Path First

À mes parents et toute la famille.

REMERCIEMENTS

Je voudrais tout d'abord exprimer toute ma gratitude envers mon directeur de recherche professeur Brigitte Jaumard, titulaire d'une chaire de recherche de l'université Concordia en Optimisation des Réseaux de communication, pour avoir accepté de superviser cette recherche, pour sa disponibilité, sa patience, sa confiance et ses conseils précieux sur le plan scientifique ainsi que pour sa grande empathie sur le plan humain au long de ces dernières années.

Je remercie les membres de jury, en particulier, l'examinateur externe le professeur Alain Houle, pour ses commentaires constructifs.

Je remercie la direction du Département de Technologie d'Information de l'Institut Polytechnique de Hanoï, Vietnam, où j'enseigne depuis 2000, pour m'avoir permise de m'absenter pendant toutes ces années afin de venir à Montréal pour perfectionner mes connaissances professionnelles.

Durant la quatrième année d'études, j'ai eu la possibilité de bénéficier d'une bourse de formation initiale de l'AUF de la région Asie-Pacifique. Je remercie la direction de l'AUF pour cette bourse.

Une grande partie des recherches de la thèse a été effectuée au laboratoire d'Optimisation des Réseaux de Communication (ORC) du Centre de Recherche sur les Transports (CRT). Je souligne le support technique des administrateurs du système informatique, techniciens et employés du centre.

J'ai beaucoup apprécié l'ambiance joviale qui régnait toujours dans notre bureau grâce à la présence de Benoît, Caroline et les autres collègues du laboratoire ORC. Je remercie particulièrement Antoine pour ses corrections instantanées du français de ma thèse.

Je n'ai pas oublié mes amis du laboratoire de Téléinformatique, UQAM : Sabri, Nhat, Boubker, Elmi, Larbi, Maher, Jason, Rudy..., je me souviendrai pour toujours de leurs encouragements et des bons comme difficiles moments que nous avons passés ensemble.

Je remercie infiniment mes parents pour leurs sacrifices, leurs pensées envers moi, leur confiance en moi et leur soutien émotionnel sans fin. Je remercie mon frère qui a pris soin de mes parents durant mon absence, sa famille et particulièrement mes adorables neveux pour les conversations téléphoniques amusantes et rigolotes des fins de semaine qui me ressourçaient.

Mes recherches n'auraient pas pu bien avancer sans les sorties avec ma grande gang d'amis vietnamiens à Montréal. Je garde dans mon coeur des souvenirs agréables de ces fêtes réjouissantes.

Les derniers remerciements sont pour "Mic", qui était toujours avec moi durant la dernière ligne droite du marathon de rédaction. Ses corrections linguistiques de l'écrit comme de l'oral ainsi que ses suggestions fructueuses m'ont permis d'obtenir une soutenance de meilleure qualité.

CHAPITRE 1

INTRODUCTION

1.1 Motivation et objectifs de recherche

La survie des réseaux de transport est un sujet intéressant pour l'industrie ainsi que pour la recherche académique. En effet, les *pannes* (*failures* en anglais), incluant les coupures de câbles et les pannes dues à une défaillance des équipements réseaux, surviennent encore fréquemment. La croissance des services Internet et la dépendance socio-économique grandissante vis-à-vis de ces services exigent une disponibilité constante, impliquant la survie de ces réseaux lors des pannes. L'introduction de la technologie DWDM (Dense Wavelength Division Multiplexing) a équipé les réseaux de transport avec une infrastructure de très grande capacité puisque chaque fibre peut transporter jusqu'à 100 terabits/seconde. Par conséquent, une coupure de câble peut causer d'importantes pertes de données. Dans les années 1970, les services interrompus à cause de pannes sont manuellement repris en redirigeant les trafics endommagés sur d'autres chemins, appelés chemins de protection (*backup path* en anglais). Cependant, le temps nécessaire à l'établissement d'un chemin de protection et à la redirection du trafic de façon manuelle est trop important : il y a donc un réel besoin pour une reprise automatisée de service. Celle-ci consiste à chercher automatiquement un chemin alternatif, qui sera le chemin de protection, pour chaque chemin d'opération (*working path* en anglais).

Les méthodes automatisées pour reprendre les services dans le cas de pannes dans les réseaux sont divisées suivant deux familles : le paradigme pro-actif et le paradigme réactif. Dans cette thèse, le terme *protection* désignera le paradigme pro-actif dans lequel les chemins de protection sont recherchés et réservés au même moment que les chemins d'opération, c'est-à-dire avant que la panne se produise. Le terme *restauration* désignera le paradigme réactif dans lequel les chemins de protection sont recherchés une fois la panne survenue. Évidemment, la protection

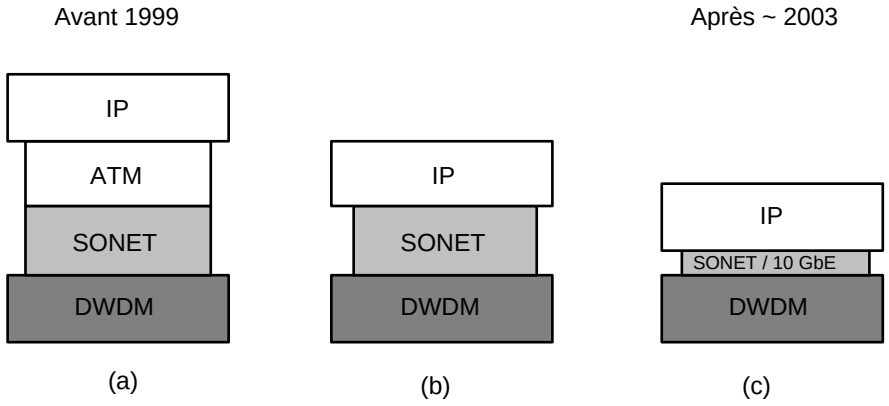

FIG. 1.1 – Les modèles IP sur DWDM

fournit une meilleure qualité de service car les chemins de protection sont prêts à être utilisés lorsqu'une panne intervient. Par contre, pour la restauration, il n'y a aucune garantie quant à l'existence d'un tel chemin, à cause des contraintes liées à la topologie du réseau ou des contraintes de ressources. Pour cette raison, la protection est intensivement déployée et étudiée dans le domaine des réseaux de transport.

Afin d'économiser la bande passante utilisée dans la protection, la *protection partagée*, dans laquelle les chemins de protection peuvent partager de la bande passante entre eux, a été proposée (voir la section 2.1.4 pour plus de détails). Compte tenu de ses avantages en termes de ressources et de qualité de service, nous considérons seulement ce type de protection dans les travaux décrits dans cette thèse.

Puisqu'une panne d'équipement ou une coupure de fibre optique cause une perte importante de données, la protection concerne principalement les réseaux à base de fibres optiques. Plusieurs modèles IP sur DWDM sont proposés pour ces réseaux : le modèle traditionnel IP/ATM/SONET/DWDM, le modèle réduit IP&MPLS/SONET/DWDM et le modèle IP&GMPLS/DWDM avec une couche SONET plus mince ou une couche 10 Gigabit Ethernet [Gro03], [ZM04b, p.212] (Figure 1.1). La protection peut être traitée indépendamment sur les différentes couches : IP, MPLS, GMPLS, ATM, SONET ou DWDM.

La protection appliquée à une couche permet d'ignorer les protections des autres couches. Au niveau de la couche IP, le problème de protection ne se pose pas réellement à cause de sa flexibilité de routage. Le routage IP se base sur les tables de routage qui sont mises à jour dynamiquement grâce aux échanges fréquents de messages entre les routeurs IP. Lors d'une panne, les chemins de routage traversant la région incriminée seront exclus des tables de routage. Une fois la convergence établie dans les tables de routage, les trafics endommagés sont redirigés sur d'autres chemins.

Une grande partie des études se concentre sur les protections des couches MPLS, GMPLS, ATM et SONET. Les protocoles de ces couches appartiennent à la classe des protocoles de commutation de circuits ou de circuits virtuels, alors qu'ils ont des caractéristiques de routage similaires. Pour cette raison, les technologies de protection de ces couches s'appliquent à une couche comme à l'autre. La protection au niveau des couches MPLS, GMPLS, ATM et SONET est plus rapide que celle de la couche IP : d'une part, parce que sauver une connexion sur l'une de ces couches permet de sauver plusieurs connections au niveau de la couche IP, d'autre part, parce que la convergence au niveau de la couche IP est lente. De façon similaire, la couche DWDM offre une protection encore plus rapide. Plusieurs études portant sur la protection de cette couche ont déjà été réalisées dans le contexte du trafic statique, [RS02, p.560], mais très peu dans le contexte du trafic dynamique : en effet, le routage et l'affectation dynamique de longueurs d'onde est déjà un problème complexe en soi. Avec la croissance des services Internet, le trafic devient de plus en plus dynamique et les études dans le contexte du trafic statique perdent leur pertinence, sauf pour la planification. Nous nous plaçons donc dans le contexte du trafic dynamique et nous nous intéressons plus spécialement à la protection au niveau des couches MPLS, GMPLS, ATM et SONET : elle répond au besoin de plusieurs modèles de réseaux et elle est rapide en pratique. Pour MPLS, GMPLS et ATM, nous nous limiterons au cas des connexions à bande passante garantie, ce qui implique que la bande passante demandée pour une connexion doit être exactement fournie tout le long du chemin et en tout temps. Cette limitation est une demande

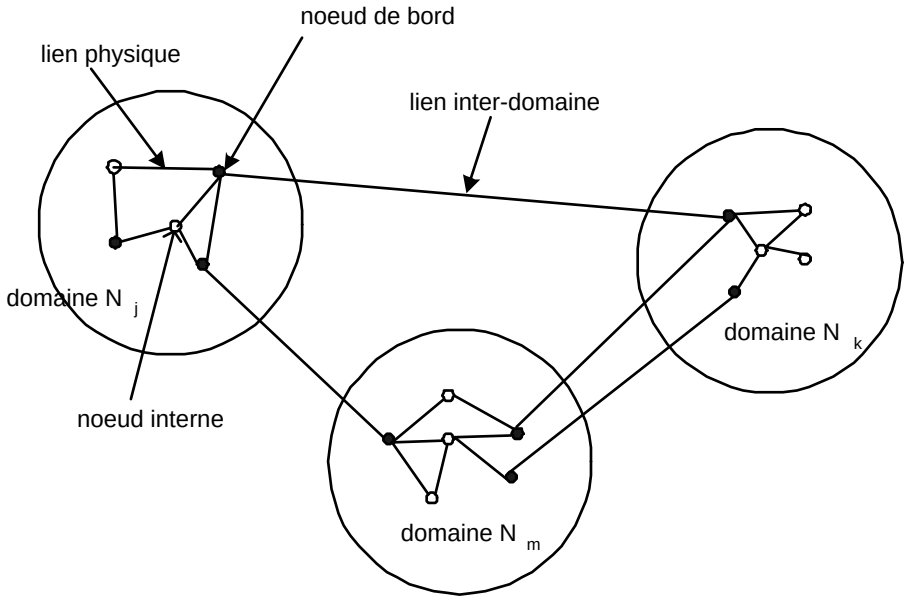

FIG. 1.2 – Un exemple du réseau multidomaines.

de qualité de service courante.

Étant donné la grande capacité des fibres optiques, elles sont souvent utilisées pour déployer des réseaux dorsaux (*backbone network*) comme les réseaux des grands fournisseurs de services, de provinces ou de pays, ainsi que pour les interconnexions entre eux. Un réseau comprenant plusieurs autres réseaux inter-connectés mais administrativement indépendants est appelé réseau multidomaines (voir la figure 1.2). Une panne dans un réseau multidomaines aura un impact à une bien plus grande échelle que celui des réseaux le composant en raison des connexions qui traversent plusieurs régions. Récemment, le tremblement de terre à Taiwan du 26 décembre 2006 a coupé une série de fibres optiques critiques et a causé une interruption sévère des télécommunications à travers tout l'est de l'Asie et a affecté en Chine plus de 10,000 noms de domaines (*Reuters* [Reu07]). Le plus grand fournisseur de services de télécommunications de Taiwan, *Chunghwa Telecom*, confirme que 98% de ses communications avec la Malaisie, Singapour, la Thailande et Hong Kong ont été coupées (*The Register* [The06]). Le système de fibres optiques *Trans Pacifique* qui connecte l'Asie avec les États Unis était presque paralysé. Environ

70% des communications entre le Vietnam et les États-Unis ont été coupées (*Tuoi Tre online* [Tuo06]). Les communications entre la Chine et l'Europe ont également été interrompues. La Corée et le Japon ont aussi été affectés car plusieurs de leurs connexions traversent les régions touchées par le tremblement de terre (*International Herald Tribune* [Int06]).

Il est donc très important de bien garder fonctionnels les réseaux multidomaines lors des pannes, sinon les impacts négatifs se multiplieront et se propageront à travers les domaines étant donné leurs interconnexions. Nous allons voir dans la section 3.2 que peu d'efforts ont été consacrés à la recherche de solution de protection des réseaux multidomaines. Par ailleurs, la plupart des autres solutions existantes ne sont pas applicables aux réseaux multidomaines puisque leurs calculs complexes de coût de protection exigent des informations complètes et globales du réseau tandis que la contrainte d'extensibilité du réseau multidomaines ne les permet pas (voir la section 2.2 pour cette contrainte, qui est référée en anglais par *scalability constraint*).

Pour toutes ces raisons, nous nous intéressons au problème du routage pour différents modèles de protection partagée dans le cas des réseaux multidomaines de transport. Nos études porteront principalement sur la protection dans le contexte du trafic dynamique au niveau des couches SONET, MPLS, GMPLS ou ATM en particulier. Les résultats de ces études s'appliquent également à tous les protocoles de commutation de circuits, avec bande passante garantie. Ils peuvent également être appliqués à la couche DWDM si un MSPP (*Multiservice Provisioning Platform*) [Cis03], [Muk06] est installé à chaque noeud. La contrainte de la continuité de longueur d'onde est relaxée grâce à la présence du traitement optique-électrique-optique dans un MSPP. Nous resterons dans le contexte d'une panne simple où le réseau subit une panne à la fois (voir la section 2.1.1). Notre travail est divisé selon deux axes.

D'abord, nous allons travailler avec un schéma simple de protection soit de protection partagée par *chemins* (voir la description détaillée de ce type de protection dans 2.1.3) pour les réseaux multidomaines. Nous nous intéressons au problème de

routage dans un contexte de trafic dynamique. Chaque demande de connexion est routée séparément avec l'objectif de minimiser la bande passante totale requise. Malgré les diverses solutions déjà proposées pour les réseaux constitués d'un seul domaine [TC03], [KL03], [QX01], [XQX02], elles ne sont pas appropriées pour les réseaux multidomaines à cause de la contrainte d'extensibilité. Quant à la solution plus spécialisée aux réseaux multidomaines proposée dans [DLM+04], elle ne tient pas compte de la possibilité de partage dans le routage. Nous proposons donc un algorithme de routage qui, à la fois, respecte la contrainte d'extensibilité et favorise les possibilités de partage durant le routage.

Dans le deuxième axe d'études, nous travaillons avec la protection partagée avec des *segments* se chevauchant (voir la description détaillée de ce type de protection dans 2.1.3). Deux raisons nous amènent à travailler sur ce type de protection : la possibilité de protéger tous les noeuds et sa rapidité de restauration. Le problème est de trouver de nouveau un routage dans un contexte de trafic dynamique. De même que pour le cas de la protection partagée par chemins, la plupart des travaux sur le sujet se limitent aux réseaux d'un domaine simple [HTC04] [RKM02], [HM03], [HM02], [XXQ02] et [CGYL07]. Les solutions pour les réseaux multidomaines [OMZ01], [ASL+02] sont incomplètes ou servent à un type de réseau particulier. Nous proposons donc un ensemble de solutions de routage pour la protection partagée avec des segments se chevauchant pour les réseaux multidomaines généraux.

Dans le cadre des études, nous nous préoccupons non seulement de la protection des liens mais également de celles des noeuds.

1.2 Contributions et organisation de la thèse

Nos recherches concernant les points évoqués ci-dessus ont abouti à plusieurs résultats que nous avons soumis sous forme d'articles dans des revues et des conférences internationales avec arbitrage. Plusieurs de ces articles ont même déjà été publiés. La thèse est constituée de quatre articles sélectionnés parmi les huit que

nous avons produits. Les articles non retenus contiennent des résultats préliminaires, qui ont été améliorés dans les quatre articles choisis. Chaque chapitre présentera un article. Le chapitre 2 a été ajouté pour donner les connaissances de base sur la protection et les réseaux multidomaines. Le chapitre 3 parcourt les solutions existantes de la littérature en détaillant celles sur lesquelles nous nous basons.

Dans le chapitre 4, nous proposons une solution de routage pour la protection partagée par chemins (*Shared Path Protection*-SPP) pour les réseaux multidomaines. Afin de respecter la contrainte d'extensibilité, nous proposons une technologie d'agrégation qui transforme le réseau multidomaines en un réseau d'un domaine simple. Nous proposons également un calcul approximatif du coût du chemin d'opération et du chemin de protection dans ce réseau agrégé, ainsi que dans le réseau initial. Ce calcul permet d'éviter les calculs complexes présentés dans les solutions existantes. Il est à noter que ces calculs complexes contraignent les solutions dans les réseaux d'un domaine simple. Un routage, appelé routage inter-domaine, est réalisé dans le réseau agrégé avec des coûts approximatifs pour déterminer les chemins des noeuds de bord (c'est-à-dire les noeuds d'accès des domaines, voir la figure 1.2) à emprunter. Un autre routage, appelé routage intra-domaine, est effectué plus tard dans chaque domaine initial afin de compléter ces chemins avec des noeuds internes et des liens physiques. En plus de l'algorithme de routage, nous décrivons également les informations de routage à échanger entre les domaines ainsi que la manière dont elles devraient être échangées. En comparaison avec les solutions existantes dans le cas de la protection partagée par chemins, cette solution offre les avantages suivants :

1. elle respecte la contrainte d'extensibilité surtout par rapport aux approches dans lesquelles on ne considère qu'un domaine simple, tandis que la qualité des solutions est proche de celles de ces dernières ;
2. elle tient compte de la possibilité de partage de la bande passante durant son routage alors que les solutions actuelles de protection par chemins pour les réseaux multidomaines ne la considèrent pas.

Le contenu de ce chapitre a été publié sous forme d'article dans [TT06] :

D. L. Truong et B. Thiongane, "Dynamic routing for Shared Path Protection in Multidomain Optical Mesh Network", *OSA Journal of Optical Networking*, vol. 5, no. 1, janvier 2006, pages 58-74.

Le chapitre 5 propose une étape de ré-optimisation servant à améliorer la qualité de la solution proposée dans le chapitre 4. L'idée vient du fait que si l'on peut considérer toutes les demandes simultanément comme lors du routage statique, on pourrait mieux favoriser le partage entre les chemins de protection. Par conséquent, dans cette étape de ré-optimisation, nous remettons en cause un ensemble de chemins de protection antérieurement réservés et nous en cherchons simultanément de nouveaux, avec l'objectif de minimiser la capacité totale de protection. Nous proposons trois modèles de ré-optimisation que nous décrivons ci-dessous.

1. Re-routage global : tous les chemins de protection de bout-en-bout sont remis en cause. Ce modèle offre une meilleure compression de la capacité de protection mais il n'est pas extensible pour les réseaux multidomaines.
2. Re-routage local : les parties des chemins de protection internes à chaque domaine sont remises en cause et recalculées ensemble afin de minimiser la capacité de protection totale du domaine. Ce modèle est extensible pour les réseaux multidomaines.
3. Re-routage local avec le moins d'effort : ce modèle minimise l'instabilité causée par la ré-optimisation du modèle de re-routage local tout en maintenant la qualité de celle-ci.

Le contenu de ce chapitre a été publié dans l'article [JT06] :

B. Jaumard et D. L. Truong, "Backup Path Re-optimizations for Shared Path Protection in Multi-domain Networks", *IEEE/ Globecom 2006*, San Francisco, USA, 27 novembre - 1 décembre 2006.

Le chapitre 6 présente notre approche du routage pour la protection partagée avec des segments se chevauchant, notée OSSP. Nous nous basons aussi sur la technologie d'agrégation du chapitre 4. Un autre calcul approximatif de coût

est également utilisé. Le routage se compose aussi de deux étapes : l'étape inter-domaine conçue spécialement pour OSSP et l'étape intra-domaine similaire à celle du chapitre 4. Un algorithme glouton et un algorithme de programmation dynamique sont proposés pour l'étape inter-domaine. De nombreuses expérimentations sont réalisées. Elles montrent que cette approche fournit des résultats comparables avec ceux de la solution optimale de OSSP pour un domaine simple en termes d'utilisation de ressources. Évidemment, notre solution est extensible pour les réseaux multidomaines tandis que la solution optimale ne l'est pas. D'autres comparaisons avec le schéma de protection par chemins du chapitre 4 sont également discutées. Le contenu de ce chapitre a été présenté dans l'article [TJ07b] :

D. L Truong et B. Jaumard, "Using Topology Aggregation for Efficient Segment Shared Protection Solutions in Multi-domain networks", *IEEE Journal of Selected Area in Communications/Optical Communications and Networking series*, 2007 (accepté pour publication).

Il est à noter que les résultats préliminaires présentés dans cet article ont été publiés dans l'article [TJ06] :

D. L. Truong et B. Jaumard, "Overlapped Segment Shared Protection in Multi-domain Optical Networks", *dans Proc. IEEE/ Asia-Pacific Optical Communication*, Gwangju, Korea, 3-7 septembre 2006, pages 63541K-1-63541K-10.

Dans le chapitre 7, nous proposons une approche tout à fait nouvelle pour OSSP par rapport aux points suivants. D'une part, alors que dans les approches proposées précédemment, la bande passante de protection peut être partagée sur tous les liens afin de profiter du maximum de partages possibles, dans cette nouvelle approche, nous n'acceptons que les partages entre les segments de protection qui traversent des chemins internes identiques au sein d'un domaine. Malgré le fait que cette approche ignore le bénéfice du partage entre les segments ayant seulement quelques liens communs, il permet de produire une solution plus extensible que celles proposées précédemment. La principale raison étant qu'on ne doit pas aller jusqu'au niveau des liens physiques pour calculer la quantité de bande passante utilisée par un segment de protection comme dans les solutions pour un domaine simple ou faire

des approximations comme dans nos travaux précédents. On peut se contenter de rester au niveau du réseau agrégé tout en obtenant un calcul exact. Ce calcul exact amène à une solution de routage plus contrôlable que si l'on utilise des approximations. D'autre part, une banque de chemins internes potentiels que les segments peuvent emprunter est pré-sélectionnée dans chaque domaine. La pré-sélection est faite en fonction de plusieurs critères qui visent à promouvoir le partage et à réduire l'utilisation des ressources. Elle fait l'objet d'un modèle mathématique. Pour le résoudre de façon exacte, nous proposons un modèle de programmation linéaire en nombres entiers. Pour une résolution en temps raisonnable, nous proposons un algorithme glouton efficace. Puisque la pré-sélection est soigneusement définie, elle ne devrait pas être vue comme une contrainte imposée sur les choix des chemins internes, mais plutôt comme un outil d'orientation vers les meilleurs choix parmi des routes possibles. Les résultats de cette nouvelle approche surpassent ceux de l'approche proposée dans le chapitre 6. En plus, la solution proposée pourrait être appliquée à la couche DWDM avec la contrainte de continuité de longueur d'onde respectée à l'intérieur des domaines. Les conversions de longueur d'onde sont uniquement possibles aux noeuds de bord. Le contenu de ce chapitre a été soumis à un journal pour publication [TJ07a] :

D. L. Truong et B. Jaumard, "A Novel Approach for Segment Shared Protection in Multi-domain Networks".

Le tableau 1.1 classifie les approches proposées dans les articles que nous incluons dans la thèse. Puisque nous avons publié nos articles indépendamment, il y aura quelques redondances dans les sections d'introduction des différents articles présentés. Nous n'enlèverons pas ces parties communes afin de conserver les articles conformes aux versions originales soumises pour publication ou déjà publiées. Pour la même raison, nous avons conservé la rédaction en anglais des articles.

Enfin, le chapitre 8 conclut la thèse.

Revue/Conf.	Modèle de protection	Calcul de coût	Échelle de partage	Algorithme	État
JON	chemin	A	lien	H	publié
Globecom	chemin	E	lien	O, H	publié
JSAC/OCN	segment	A	lien	H	accepté
	segment	E	chemin interne	O, H	soumis

O/H : Optimisation ou Heuristique.
E/A : Exact ou Approximatif.

TAB. 1.1 – Classification des approches proposées dans les articles inclus dans la thèse

1.3 Articles de conférences et de revues rédigés durant la thèse

Les articles de conférences et de revues produits durant la thèse sont listés ci-dessous en ordre chronologique. Ceux qui ont été inclus dans la thèse sont indiqués avec une astérisque (*) :

[1] D. L. Truong, O. Cherkaoui, H. Elbiaze, N. Rico et M. Aboulhamid, "A Policy-based approach for User controlled Lightpath Provisioning", *dans Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS)*, vol. 1, Korea, avr. 2004, page 859-873.

[2] B. Thiongane et D. L. Truong, "Shared Path Protection in Multi-domain Optical Mesh Networks" , *dans Proc. IASTED/Communication and Computer Network*, Marina del Rey, CA, USA, 24-26 oct. 2005, pages 138-145. (Les noms des auteurs sont listés selon l'ordre alphabétique).

[3*] D. L. Truong et B. Thiongane, "Dynamic routing for Shared Path Protection in Multidomain Optical Mesh Network", *OSA Journal of Optical Networking*, vol. 5, no. 1, janv. 2006, pages 58-74.

[4*] B. Jaumard et D. L. Truong, "Backup Path Re-optimizations for Shared Path Protection in Multi-domain Networks", *IEEE/ Globecom 2006*, San Francisco, USA, 27 nov. - 1 déc. 2006.

[5] D. L. Truong et B. Jaumard, "Overlapped Segment Shared Protection in

Multi-domain Optical Networks", *dans Proc. IEEE/ Asia-Pacific Optical Communication*, Gwangju, Korea, 3-7 sept. 2006, pages 63541K-1-63541K-10.

[6*] D. L. Truong et B. Jaumard, "Using Topology Aggregation for Efficient Segment Shared Protection Solutions in Multi-domain networks", *IEEE Journal of Selected Area in Communications/Optical Communications and Networking series*, vol. 25, no. 9, déc. 2007 (à paraître).

[7*] D. L. Truong et B. Jaumard, "A Novel Approach for Overlapping Segment Shared Protection in Multi-domain Networks", 2007 (soumis pour publication).

[8] B. Thiongane et D. L. Truong,"Shared Path Protection for bandwidth guaranteed connections in Multi-domain Optical Mesh Networks", *International Journal of Computers and Applications*, 2007 (accepté pour publication).

CHAPITRE 2

GÉNÉRALITÉS SUR LA PROTECTION ET LES RÉSEAUX MULTIDOMAINES

Dans ce chapitre, nous allons donner une description des connaissances de base sur la protection et les réseaux multidomaines, les hypothèses générales de travail du problème de protection ainsi que les définitions des termes que nous utiliserons dans la suite de la thèse.

2.1 Généralités sur la protection

La protection peut être divisée chronologiquement en trois phases.

1. Routage : dans cette phase, les chemins d'opération ainsi que leur chemin de protection sont recherchés. Un chemin d'opération et son chemin de protection ne doivent pas être simultanément endommagés lors d'une panne sinon le chemin de protection ne sera plus disponible pour remplacer le chemin d'opération associé, qui lui est déjà en panne. À l'issue de la recherche, les chemins d'opération doivent être établis et leurs chemins de protection doivent être réservés. Selon la caractéristique de la protection, cette phase devrait être réalisée bien avant la panne.
2. Notification de panne : lorsqu'un équipement ou une fibre tombe en panne, celle-ci est immédiatement détectée et le noeud qui contrôle la mise en place du (des) chemin(s) de protection en est informé.
3. Activation des chemins de protection : les chemins de protection sont configurés (s'ils ne sont pas pré-configurés) à partir des ressources réservées et les trafics endommagés sont redirigés vers ces nouveaux chemins de protection.

La deuxième et la troisième phases sont réalisées après que la panne soit survenue. Comme elles servent à restaurer le service endommagé, l'intégration de ces deux

phases est référée dans plusieurs documents par le terme *restauration*. Pourtant on ne devrait pas la confondre avec le paradigme de restauration. La durée de la restauration est calculée par :

$$T_{restauration} = T_{notification} + T_{configuration}. \tag{2.1}$$

Cette durée correspond aussi au temps où la connexion affectée par la panne est hors service, elle doit être très limitée pour ne pas interrompre les services des couches supérieures. Le temps de notification est le temps nécessaire pour détecter la panne (noté t_d) plus le temps requis pour propager le message d'avertissement le long du chemin d'opération. Le temps de configuration comprend le temps pour transférer la demande de configuration le long du chemin de protection et le temps pour configurer une commutation (noté t_c) si les commutations sont réalisées en cascade. Alors, si ℓ^W et ℓ^B sont respectivement les longueurs des chemins d'opération et de protection, t^W et t^B les temps de traitement d'un message sur le chemin d'opération et de protection respectivement, le temps de restauration est calculé par la formule suivante :

$$T_{restauration} = t_d + t_c + \ell^W \times t^W + \ell^B \times t^B. \tag{2.2}$$

Les deux premiers termes sont dominés par les deux derniers. On voit clairement que plus les chemins d'opération et de protection sont courts, plus la restauration est rapide.

Il est souvent affirmé à tort que le temps de restauration doit rester inférieur à 50 ms. On suspecte que ce délai de 50 ms est issu de la concurrence entre la protection basée sur des anneaux, dont la durée de restauration est de moins de 50 ms, et la protection maillée, dont la durée de restauration est d'environ 200 ms ; tout ceci afin d'exclure l'utilisation de la protection maillée [Gro03]. Une étude [Sch01] montre que les services de la voix, SNA, ATM, X.25, SS7, DS1, les données 56 kb/s, la vidéo digitale NTC, ou les services d'accès au SONET OC-12 et OC-48 restent fonctionnels durant une interruption de connexion de 200 ms. En plus, à

l'exception SS7, les autres services peuvent même survivre avec une interruption de 2 à 5 secondes. Par conséquent, une durée de 2 secondes peut être considérée comme un seuil de restauration satisfaisant [Gro03].

Revenons aux trois phases de la protection. La phase de routage est la plus complexe et elle influence également la durée de restauration : elle détermine les longueurs des chemins d'opération et de protection. Par conséquent, nous accentuerons nos efforts sur cette phase, les deux autres phases seront mentionnées brièvement dans chacun de nos articles.

2.1.1 Panne simple

Les statistiques de Telcordia dans [TN94] ont montré que le taux de coupures des fibres optiques est de 4.39 coupures/année/1000 miles de fibres tandis que les pannes des équipements causent une interruption de service de 6.5 minutes/année. Le temps moyen de réparation d'une coupure de fibre et d'un équipement est respectivement de 12 et de 2 heures. La probabilité qu'une seconde coupure ou qu'une seconde panne d'équipements se produise pendant les réparations est donc très faible. Ainsi, on suppose souvent qu'il y a seulement des pannes simples dans le réseau. La panne aura le temps d'être réparée avant qu'une seconde panne survienne. Cette hypothèse permet de simplifier les protections tout en satisfaisant les exigences pratiques. Cette hypothèse, appelée *contexte d'une panne simple*, se retrouve en anglais dans nos articles sous le libellé *single failure context*.

2.1.2 Anneau auto-réparateur versus protection maillée

L'anneau auto-réparateur (*self-healing ring*) est un modèle classique de protection souvent utilisée pour les réseaux SONET/SDH. UPSR (Unidirectional Path-Switched Ring) et BLSR (Bidirectional Line Switched Ring) sont deux exemples de ce modèle. Dans UPSR, un anneau d'opération transfère le trafic dans une seule direction. Un autre anneau est dédié pour la protection. Les deux anneaux partagent les mêmes câbles (figure 2.1a). Lors d'une panne, le trafic sera transféré uniquement

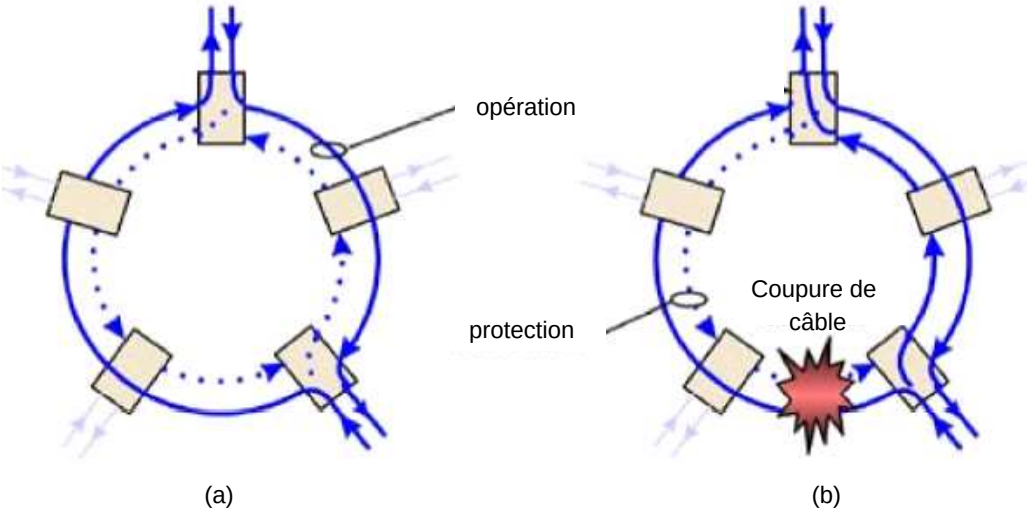

FIG. 2.1 – UPSR (Figure extraite de [Gro03])

sur l'anneau de protection en direction opposée. La redirection du trafic est faite aux points d'entrée et de sortie du trafic des anneaux (figure 2.1b).

Dans un anneau BLSR 4 fibres, deux anneaux d'opération transfèrent le trafic dans les deux directions et deux autres anneaux de protection sont nécessaires. Les anneaux partagent les mêmes câbles (figure 2.2a). Lors d'une panne, le trafic est transféré partiellement sur l'anneau d'opération puis dirigé sur l'anneau de protection autour du lien endommagé (figure 2.2b). BLSR 2 fibres peut être considéré comme un BLSR 4 fibres dans lequel les fibres de protection sont incorporées dans les fibres d'opération. Deux anneaux sont utilisés dont chacun transfère le trafic d'opération à la moitié de la capacité et réserve le reste pour la protection. En réalité, BLSR 2 fibres est beaucoup plus utilisé que BRSL 4 fibres.

Le modèle d'anneau est connu pour sa restauration rapide grâce à des anneaux de protection pré-configurés, c'est-à-dire, $T_{configuration} = 0$, et pour son implémentation simple dans les petits réseaux. Cependant, l'implémentation efficace des anneaux devient compliquée dans des réseaux de grande taille, surtout en comparaison avec les modèles de protection maillée. De plus, la technique de configuration automatique des connexions est beaucoup plus mature pour la topologie maillée que pour celle en anneau. Il faut noter aussi que la protection maillée permet de ser-

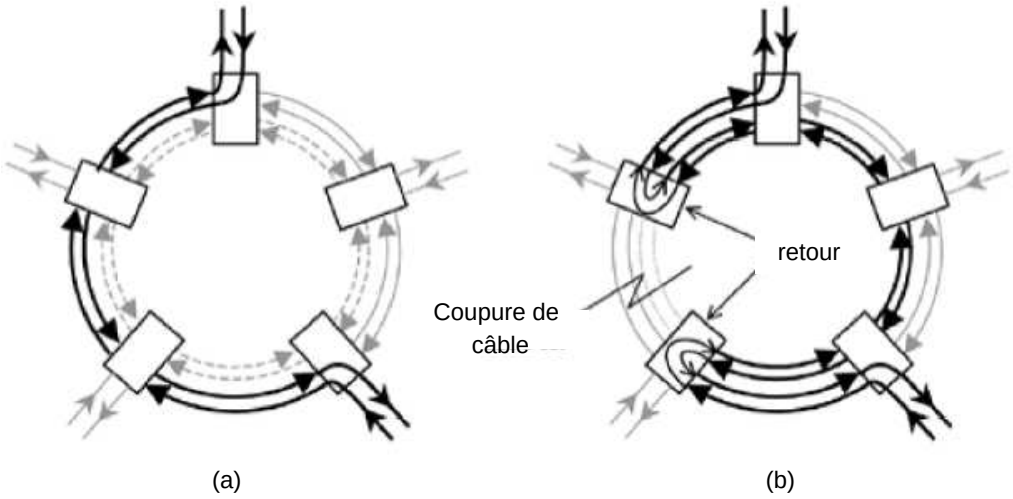

FIG. 2.2 – BLSR 4 fibres (Figure extraite de [Gro03])

vir un plus grand nombre de demandes que la protection en anneau à capacité équivalente. Pour cette raison, un grand nombre d'études récentes portent sur la protection maillée. C'est aussi le cas de nos études dans cette thèse. Nous allons voir dans la section suivante les modèles de protection maillée de base.

2.1.3 Protections maillées : par liens, par segments, par chemins

Les modèles de protection maillée sont classifiés selon l'échelle de re-routage du trafic lors des pannes. Dans la protection par liens (*link-based protection*), chaque lien du chemin d'opération est protégé par un segment de protection distinct (figure 2.3a). Lors d'une panne sur un lien du chemin d'opération, l'existence de la panne est notifiée aux deux noeuds extrêmes du lien et le segment de protection de ce lien est activé. Le trafic est encore acheminé sur les liens non-affectés du chemin d'opération et détourné autour du lien affecté sur le segment de protection [ZM04a]. Dans la protection par chemins, le chemin d'opération est protégé de bout-en-bout par un seul chemin de protection[1] (figure 2.3b). Indépendamment

[1]Nous utilisons le terme protection par chemins lors que nous évoquons la question de protection au niveau d'un réseau. Il est important de noter que dans un tel schéma de protection, chaque chemin d'opération est protégé par un seul chemin, son chemin de protection.

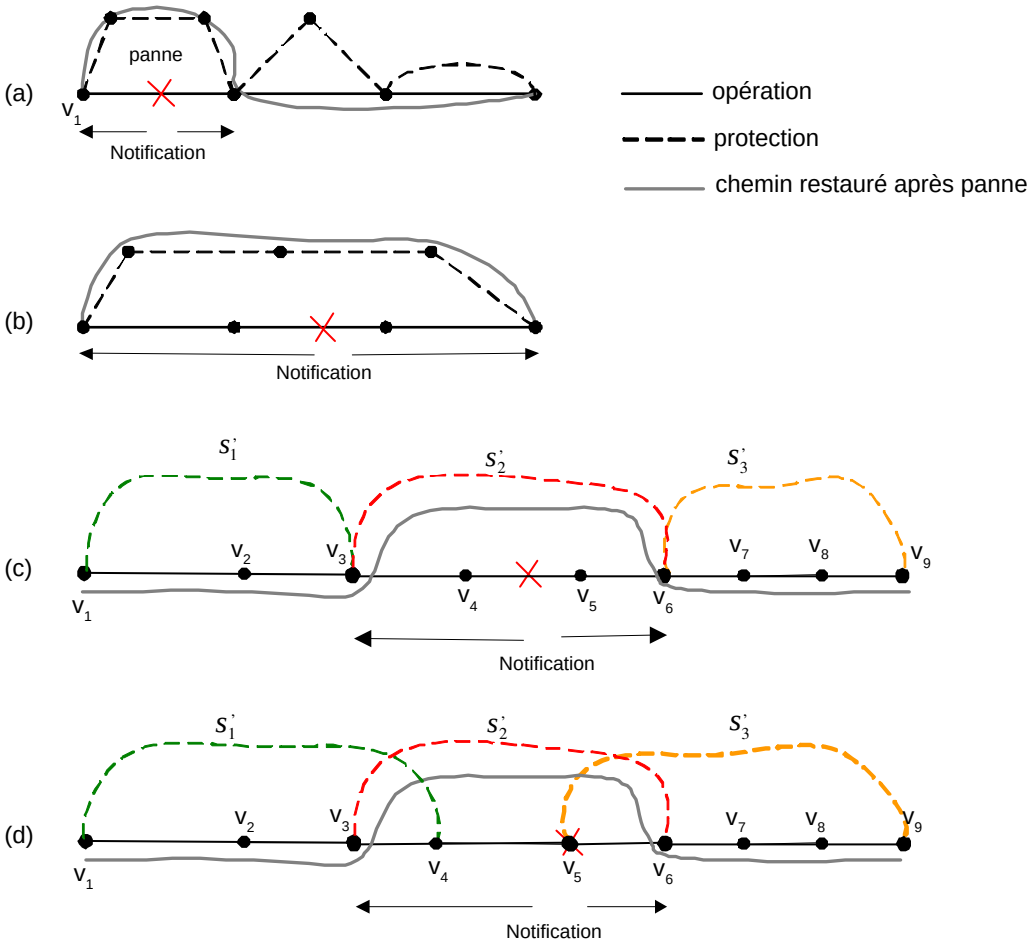

FIG. 2.3 – Les modèles de la protection maillée : protection par liens (a), par chemins (b), par segments (c) et protection avec des segments se chevauchant (d).

de la panne sur le chemin d'opération, des messages de notification sont envoyés vers la source et la destination, demandant l'activation du chemin de protection. Le chemin de protection remplace ensuite le chemin d'opération de bout-en-bout pour transférer le trafic. En comparant les deux modèles, la protection par liens est généralement plus rapide grâce à sa courte notification et son court segment de protection à configurer. Pourtant, ce modèle a tendance à utiliser plus de ressources de protection car ses segments de protection traversent au total un plus grand nombre de liens. Il faut noter aussi que la protection par liens ne permet pas de protéger les noeuds puisqu'ils constituent des points communs entre le chemin d'opération et les segments de protection. Au contraire, la protection à l'aide d'un chemin est capable de protéger tous les noeuds, excepté les noeuds source et destination qui ne peuvent être protégés par aucun modèle de protection topologique en général. En pratique, ces noeuds source et destination sont protégés par dédoublement de leur équipement, ce qui est totalement transparent pour notre travail.

Le modèle de protection par segments est une solution intermédiaire entre les modèles par liens et par chemins. Un segment est un ensemble de liens continus entre deux noeuds définis comme la tête et la queue du segment. Le chemin d'opération est divisé en plusieurs segments. Chacun est protégé par un segment de protection distinct. Dans le modèle classique de protection par segments (figure 2.3c), les segments d'opération se concatènent sans chevauchement. Les modèles de protection par liens et par chemins sont donc deux cas particuliers du modèle de protection par segments. Tout comme la protection par liens, la protection par segments ne permet pas de protéger les têtes et les queues des segments. Pourtant elle permet de restaurer les pannes plus rapidement qu'avec le modèle de protection par chemins.

La protection avec des segments se chevauchant est introduite dans [GMM00] et développée avec plusieurs variantes dans les travaux subséquents [HM02], [XXQ02], [HTC04], etc. Hormis les avantages qu'elle hérite de la protection typique par segments, elle permet en plus de protéger tous les noeuds grâce aux chevauchements des segments. Une tête ou une queue d'un segment est protégée au moins par un

autre segment de protection. La figure 2.3d montre un exemple de cette protection. Le noeud v_5 est la tête du segment $v_5 - v_6 - v_7 - v_8 - v_9$ alors il n'est pas protégé par le segment s'_3. Pourtant, v_5 est aussi un noeud intermédiaire du segment $v_3 - v_4 - v_5 - v_6$ alors il est protégé par le segment protection s'_2.

2.1.4 Protection dédiée versus protection partagée

Dans la protection dédiée, les ressources sont réservées exclusivement pour un chemin ou un segment de protection. Ainsi les chemins et segments de protection peuvent être pré-configurés : la restauration est très rapide. Les protections dédiées typiques sont : i) protection 1+1 dans laquelle le trafic est transmis sur le chemin d'opération et reproduit en même temps sur le chemin de protection, et ii) protection 1:1 dans laquelle le trafic sera transmis sur le chemin de protection uniquement en cas de panne. La protection dédiée demande donc 100% des ressources en service pour assurer la protection. La meilleure utilisation de ressources est seulement de 50% [ZM04b].

À l'inverse de la protection dédiée, la protection partagée accepte que, les chemins/segments de protections contre différentes pannes, partagent les ressources sur les liens, les noeuds, les segments ou les chemins communs (voir [HM04c] pour plus de détails). La protection partagée est attrayant puisqu'elle exige moins de ressources de protection par rapport à la protection dédiée. Comme les ressources de protection peuvent être utilisées par plusieurs chemins/segments de protection, ces derniers ne peuvent être configurés qu'après l'apparition de la panne, quand on sait exactement quel chemin ou quel segment activer. La restauration est donc plus lente que la protection dédiée. Pour définir une protection qui est capable de restaurer toutes les connexions lors des pannes, la protection partagée est souvent regardée dans le contexte d'un nombre limité de pannes simultanées, par exemple une panne simple. Dans le cas d'une panne simple, la condition de partage est énoncée de la manière suivante : les chemins/segments de protection peuvent partager leurs ressources si leurs chemins/segments d'opération correspondants ne sont pas endommagés simultanément lors d'une panne simple.

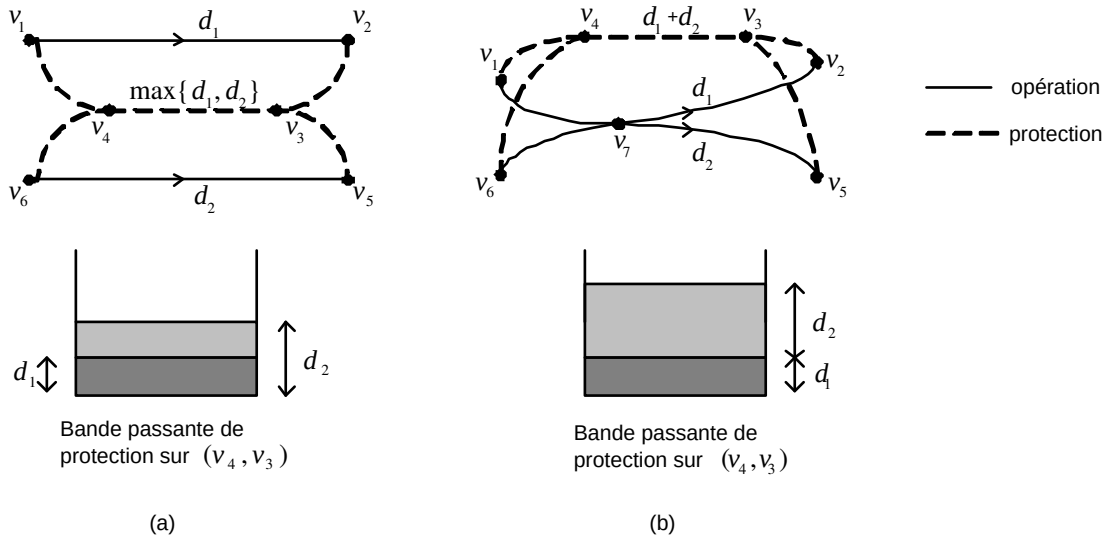

FIG. 2.4 – Exemples des chemins de protection qui partagent des ressources (a) et qui ne le peuvent pas (b).

La figure 2.4 montre un premier exemple d'un cas où deux chemins de protection partagent des ressources, cas (a) ; et un second exemple du cas où ils ne peuvent pas partager, cas (b). On considère deux chemins d'opération : de v_1 à v_2 avec la bande passante d_1 et de v_6 à v_5 avec la bande passante d_2. Dans le cas (a), les chemins d'opération sont disjoints. Par conséquent, ils ne peuvent pas être endommagés simultanément lors d'une panne simple, leurs chemins de protection peuvent partager la bande passante sur leur lien commun (v_4, v_3). La bande passante totale à réserver sur (v_4, v_3) pour ces deux chemins est $\max\{d_1, d_2\}$. Dans le cas (b), les deux chemins d'opération s'entrecroisent en v_7 ; en conséquence, leurs chemins de protection ne peuvent pas partager de bande passante. Chacun de ces deux chemins doit réserver séparément de la bande passante sur (v_4, v_3) conduisant à une bande passante totale $d_1 + d_2$ qui est supérieure à celle du cas (a).

S'inspirant de cette économie des ressources, nous nous intéressons seulement à la protection partagée. Quand le partage est incorporé dans une protection par chemins ou par segments, on parlera de protection partagée par chemins SPP (*Shared Path Protection*) ou de protection partagée par segments SSP (*Shared Segment*

Protection) ou encore de protection partagée avec des segments se chevauchant OSSP (*Overlapped Segment Shared Protection*) .

2.1.5 Routage pour la protection

Comme nous l'avons mentionné précédemment, nous nous intéressons plus particulièrement à la protection des couches MPLS, GMPLS, ATM et SONET qui appartiennent à la classe des protocoles de commutation de circuits ou de circuits virtuels avec une bande passante garantie. La nature de la commutation de circuits ainsi que la demande de bande passante garantie imposent que l'acheminement de chacune des connexions doit être déterminé et les informations associées doivent être transmises à la source avant que le trafic soit mis sur la connexion. Le routage est donc un routage à la source (*source routing* en anglais). Ce routage nécessite que le point de calcul du réseau ait des informations complètes sur l'ensemble du réseau afin de pouvoir déterminer les chemins optimaux. Ceci est très différent du cas du protocole IP dans lequel l'acheminement se fait localement de noeud en noeud durant la propagation du trafic dans le réseau.

Le routage dans le cas de la protection dédiée est relativement simple à cause de la symétrie entre le chemin d'opération et le chemin de protection. Le problème est de chercher deux chemins disjoints pour chaque paire de source/destination. Plusieurs algorithmes traditionnels [Suu74], [ST84], [Bha99] et d'autres plus récents [Kle96], [XCX+04] pourraient être utilisés pour résoudre ce dernier problème. Le routage pour la protection partagée est plus complexe. Il est caractérisé par les critères suivants [ZM04b].

- Un chemin d'opération et son chemin de protection ne doivent pas être endommagés simultanément lors d'une panne, sinon une panne simple pourrait interrompre les deux chemins menant à un échec de la restauration. Cette condition est généralement interprétée par la disjonction des deux chemins. Elle s'applique également pour la protection dédiée.
- Deux chemins de protection dont les chemins d'opération peuvent être endommagés simultanément lors d'une panne, ne doivent pas partager de res-

sources. En effet, une panne simple interrompant leurs chemins d'opération simultanément, force l'activation de leurs chemins de protection au même moment, alors que les ressources partagées ne sont suffisantes seulement pour l'un d'entre eux.

– Les ressources de protection réservées devraient être partagées autant que possible pour augmenter l'efficacité d'utilisation de ces ressources dans le réseau. Ce partage devrait être favorisé durant le routage.

Le second critère montre que le niveau de partage des ressources entre les chemins de protection dépend fortement de leurs chemins d'opération. Ce critère rend le problème encore plus complexe et les informations détaillées et complètes d'allocation de ressources sont exigées. Nous allons voir dans la section 2.2 que ceci est un obstacle à surmonter pour le routage de la protection partagée dans les réseaux multidomaines.

Plusieurs approches ont été proposées pour résoudre le problème de routage dans le cas de la protection partagée. La figure 2.5 montre une classification de la protection partagée selon son routage. Un grand nombre des algorithmes de routage prenant en compte la protection sont consacrés au routage statique, par exemple les travaux dans [MK98], [GS98], [LR01], [Liu01], [Bou05] etc. Dans le routage statique, les capacités des ressources d'opération et de protection sont planifiées de façon optimale pour l'ensemble des demandes futures de connexions étant donné leurs profils. Ce problème est aussi connu sous le nom du problème de conception de réseaux capables de survie (*survivable network design* en anglais). Parmi les travaux dans ce domaine, il faut citer la protection à l'aide de p-cycles qui est proposée pour la première fois dans [GS98] et développée dans [SG03], [SGA02], etc. Un avantage du routage statique est l'optimalité des chemins d'opération ainsi que ceux de protection pour un profil de demandes donné. Il offre alors une meilleure utilisation des ressources tant que le trafic futur adhère au profil. Quand les demandes ne correspondent plus au profil du trafic, la qualité des solutions de routage statique se détériore. Le routage statique convient donc seulement pour la conception des réseaux de petite taille dont les demandes de trafic sont stables [HM02].

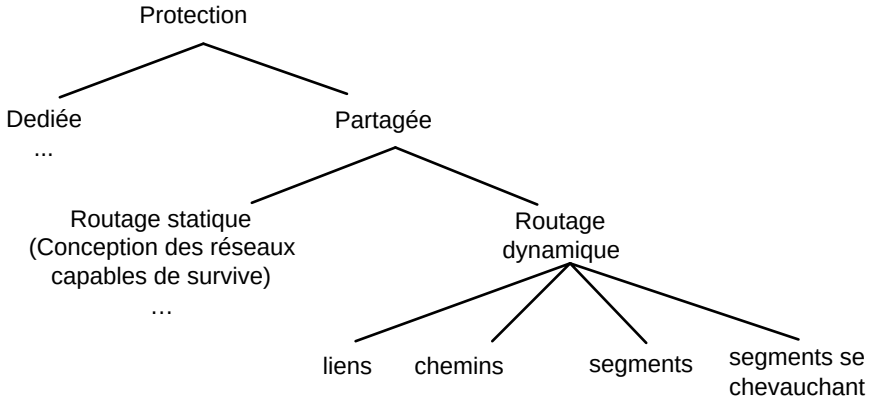

FIG. 2.5 – Classification des méthodes de protection.

Cependant, les demandes de connexions aujourd'hui changent dynamiquement. Il est très difficile, voire improbable, de trouver un profil de demandes qui peut être valable jour après jour. Le routage dynamique surmonte cette faiblesse en plaçant chacune des demandes à son arrivée et sans connaissance sur les demandes futures. Pour une demande courante, les chemins d'opération et de protection sont conçus de façon optimale, souvent avec l'objectif de minimiser la bande passante utilisée par ces chemins. L'avantage du routage dynamique est qu'il n'exige aucune connaissance sur le trafic futur. Le routage dynamique convient donc aux réseaux de grande taille associés à un trafic imprévisible. Cela nous a amené à seulement considérer le cas du routage dynamique dans cette thèse.

2.2 Réseaux multidomaines

Les réseaux de communication sont gérés par différentes organisations. Un réseau, sous une gestion administrative unique et sur lequel un ensemble de protocoles sont installés de façon homogène, est considéré comme un domaine simple. Des exemples de réseaux d'un domaine simple sont : le réseau d'une université (ex : réseau de l'Université de Montréal, de McGill, etc.), le réseau dorsal d'une organisation (ex : le réseau informatique du RISQ [ris07]) le réseau dorsal d'un pays (ex : Canet [can05], NSF [nsf05]). On le réfère également par le terme de

système autonome (*Autonomous System*). Comme le réseau d'un domaine simple est sous une gestion unique, sa topologie ainsi que ses informations d'allocation de ressources sont souvent disponibles à tous les noeuds du réseau. Ceci donne à ces noeuds une vision complète du réseau pour pouvoir réaliser les routages. La disponibilité des informations est obtenue par des échanges et des mises à jour régulières des informations entre les noeuds du réseau selon un protocole de routage interne IGP (*Interior Gateway Protocol*) tel que OSPF (*Open Shortest Path First*) ou ISIS (*Intermediate System to Intermediate System*).

Un réseau multidomaines (voir la figure 1.2 dans le chapitre 1) est défini comme une interconnexion de réseaux d'un domaine simple [BSO02]. Un noeud d'un domaine qui n'a aucun lien avec les noeuds d'un autre domaine est appelé noeud interne (*internal node*). Les noeuds qui connectent un domaine avec d'autres domaines sont appelés les noeuds de bord (*border nodes*). Le lien connectant deux noeuds de bord appartenant à deux domaines différents est appelé un lien inter-domaine (*inter-domain link*). Il fait le pont entre ces deux domaines. Le lien connectant deux noeuds internes d'un même domaine est un lien physique (*physical link*). Pour des questions d'administration, de sécurité et surtout d'extensibilité, les informations de routage d'un domaine ne sont pas toutes publiées à l'extérieur du domaine, mais seulement des agrégations de celles-ci [LRVB04] à travers la communication entre les noeuds de bord. Cette restriction d'échanges d'informations entre les domaines est référée dans cette thèse par le terme *contrainte d'extensibilité* et dans nos articles en anglais par *scalability constraint* ou *scalability requirement*. Les protocoles EGP (*Exterior Gateway Protocol*) tels que BGP (*Border Gateway Protocol*) pourraient être utilisés pour échanger les informations agrégées.

La conséquence de la contrainte d'extensibilité se traduit par le fait qu'un noeud donné du réseau ne connaît ni la topologie globale ni les allocations de ressources détaillées de tous les liens du réseau. Ceci impose des difficultés additionnelles au problème de routage dans un réseau multidomaines, surtout un routage pour la protection partagée qui, comme on l'a vu dans la section 2.1.5, demande des informations d'allocation de ressources globales et détaillées. Toutefois, chaque do-

maine détient une vision agrégée du réseau multidomaines grâce aux échanges des informations agrégées entre domaines, et une vue complète de lui-même grâce aux échanges fréquents d'information de routage interne.

CHAPITRE 3

ÉTAT DE L'ART

Ce chapitre fait une revue de la littérature sur les approches de protection par chemins et par segments s'adressant aux réseaux d'un domaine simple ainsi qu'aux réseaux multi-domaines.

3.1 Protection d'un domaine simple

Dans cette section, nous présentons les solutions existantes de protection par chemins (SPP) et celles avec des segments se chevauchant (OSSP) dans le cadre des réseaux composés d'un seul domaine. Nous commençons par la protection par chemins et poursuivons avec la protection par segments.

3.1.1 Protection par chemins

Le routage, dans le cas de la protection par chemins, consiste à chercher un chemin d'opération ainsi qu'un chemin de protection tels que ces deux chemins soient disjoints, en termes de liens et de noeuds, et que la capacité totale qui leur est allouée soit minimale. Étant donné une demande de connexion, le chemin d'opération occupe constamment la bande passante demandée sur tous ses liens. Par contre, le chemin de protection occupe une fraction variable de la bande passante sur ses propres liens, en raison de différents partages avec les chemins de protection alloués antérieurement à d'autres chemins d'opération. Le coût de protection d'un lien est donc différent de son coût d'opération. Le problème de routage SPP est vu comme un problème de recherche de deux chemins disjoints avec deux coûts de liens. La complexité de ce problème a été étudiée et classifiée NP-difficile [LMDL92]. Le problème équivalent dans le cadre des réseaux DWDM dans lequel, si les chemins de protection partagent la bande passante, ils partagent une longueur d'onde entière, est également classifié NP-complet [OZZ$^+$04].

Travail	Granularité de partage	Conv.	Info.	Résumé de la contribution
SCI, SPI de Kodialam *et al* [KL00] [KL03]	S	O	C, P	Solution exacte basée sur PNE. Recherche conjointe des chemins d'opération et de protection. Approximation de coût dans SPI.
Travail de Su *et al.* [SS01]	W	O	C	Solution exacte basée sur PNE et une heuristique à deux étapes.
Travail de Xin *et al.* dans [XYDQ01]	W	O	C	K itérations de l'algorithme à deux étapes pour trouver K candidats pour les chemins d'opération et de protection. Choisir la paire correspondant au coût le plus faible.
DPIM de Qiao *et al.* [QX01]	S	O	C, P	Gestion distribuée des informations. Meilleure approximation du coût que celle de SPI.
APF-BPC de Xu *et al.* [XQX02]	S	O	P	Le chemin d'opération est recherché en considérant le coût potentiel de son chemin de protection.
ITSA de Tapocal *et al.* [TC03]	S	O	P	K itérations de l'algorithme à deux étapes de la même façon que dans [XYDQ01] mais la granularité de partage est plus fine.
Travail de Xiong *et al.* [XXQ03a]	S	O	C, P	Amélioration de SCI et DPIM en ajoutant des pondérations dans les fonctions objectives.

Con. : Conversion de longueurs d'onde.
O/N : Oui ou Non.
Info. : Quantité des informations demandées.
C/P : Information Complète ou Partielle demandée.
W/S : Toute une longueur d'onde (W) ou une portion de longueur d'onde (S).

TAB. 3.1 – Travaux existants sur le routage dynamique pour la protection par chemins

Travail	Granularité de partage	Conv.	Info.	Résumé de la contribution
Travail de Li *et al.* dans [LWKD03]	S	O	C, P	Algorithme à deux étapes. Calculs de coûts similaires à ceux de SPI et SCI. Signalisation distribuée.
Bouillet et al [BL04]	W	O	P	Une approche stochastique. K-plus court chemins.
CAFES-OPT de Ou *et al.* [OZZ+04]	W	O	C	Algorithme à deux étapes afin de trouver les premiers chemins d'opération et de protection et les optimiser ensuite en fixant un et recalculant l'autre en alternance.

TAB. 3.2 – Travaux existants sur le routage dynamique pour la protection par chemins (suite)

Les tableaux 3.1 et 3.2 listent les solutions de routage dynamique de la protection par chemins dans le cadre des réseaux d'un seul domaine. L'objectif de ces solutions est de minimiser la capacité totale en termes de bande passante ou en termes d'unités de longueur d'onde de la demande courante. Pour un algorithme proposé dans [SS01] et ceux dans [LWKD03] et [BL04], un chemin d'opération et son chemin de protection sont calculés séparément lors de deux étapes distinctes. Le chemin d'opération est calculé d'abord, et correspond souvent au plus court chemin. Le chemin de protection est généralement calculé ensuite avec un calcul de plus court chemin sur les ressources résiduelles, et la bande passante utilisée définit le coût de protection. Le coût de protection d'un lien dépend de sa quantité de bande passante de protection partageable pour le chemin de protection, ce qui dépend, à son tour, du chemin d'opération lui-même. Puisque les chemins d'opération et de protection sont calculés séparément, les algorithmes de routage sont donc classifiés comme "algorithmes à deux étapes" (*two step algorithms*). Leurs temps de calcul, tout à fait acceptables, sacrifient cependant la qualité de leurs solutions. En effet, la paire des chemins d'opération et de protection trouvée par ces algorithmes

est en général sous-optimale. Étant donné un chemin d'opération, ces algorithmes pourraient trouver un chemin de protection optimal de plus faible coût. Cependant, en changeant le chemin d'opération, il peut exister un autre chemin de protection dont le coût est encore plus faible en raison de nouvelles possibilités de partage. Cela pourrait entraîner un coût total plus faible, comparé à celui obtenu par un algorithme à deux étapes.

Dans [XYDQ01], [TC03], les algorithmes à deux étapes sont appliqués un certain nombre de fois afin de trouver plusieurs candidats pour les chemins d'opération et de protection, pour finalement choisir la paire la moins coûteuse. Ces algorithmes permettent de mieux s'approcher de la solution optimale. D'autres travaux tels que [KL00], [KL03], [SS01], [XQX02], [XXQ03a] et [OZZ+04] calculent le chemin d'opération conjointement avec son chemin de protection, ce qui permet d'améliorer les résultats, voire d'obtenir la solution optimale. Dans [KL00] et [KL03], une formulation de programmation linéaire en nombres entiers - PNE (*Integer Linear Programming* - ILP) exacte est proposée pour les réseaux à bande passante garantie où la bande passante de protection pourrait être partagée pour toutes les granularités. Les auteurs de [SS01] proposent une autre solution exacte, basée également sur une formulation PNE, dans le cadre des réseaux DWDM dans laquelle la granularité de partage correspond à une longueur d'onde entière.

Parmi les travaux existants précédemment cités, [KL00] est le pionnier. En plus de la solution exacte proposée, les calculs de coûts exacts sont très bien définis dans son modèle avec information complète SCI (*Sharing with Complete Routing Information*). Ultérieurement, plusieurs travaux se basent sur ces calculs de coûts. Nous nous en inspirons également. En raison de leurs rôles importants, nous détaillerons ces calculs dans la prochaine section.

3.1.2 Modèle de partage avec information complète

Les calculs de coûts du SCI sont présentés dans [KL00], [KL03] et re-expliqués plus clairement dans [HTC04]. Afin de les présenter, considérons une demande courante d'une bande passante d entre un noeud source s et un noeud destination

p	le chemin d'opération à trouver pour la demande courante.
p'	le chemin de protection à trouver pour la demande courante.
c_ℓ^{res}	la capacité résiduelle du lien ℓ.
$B_{\ell'}$	la capacité de protection du lien ℓ'.
$\Phi_{\ell'}^{\ell}$	l'ensemble des demandes utilisant le lien ℓ pour leur chemin d'opération et le lien ℓ' pour leur chemin de protection.
$B_{\ell'}^{\ell}$	la bande passante totale des demandes dans $\Phi_{\ell'}^{\ell}$.
$b_{\ell'}^{\ell}$	la bande passante de protection à réserver sur le lien ℓ' pour le chemin de protection de la demande courante.
x_ℓ	la variable binaire qui annonce le passage du chemin d'opération p par un lien ℓ. Elle est fixée à 1 si p traverse ℓ et 0 sinon.
$y_{\ell'}$	la variable binaire qui annonce le passage du chemin de protection p' par un lien ℓ'. Elle est fixée à 1 si p' traverse ℓ' et 0 sinon.

TAB. 3.3 – Notations

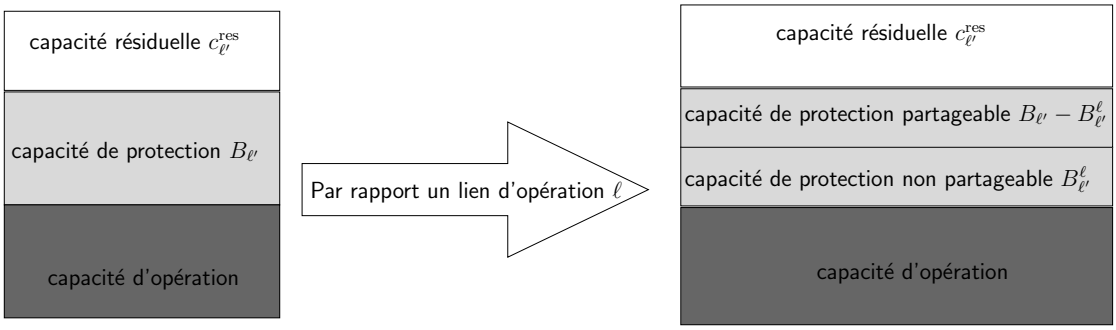

FIG. 3.1 – Structure de la bande passante sur un lien

t. Cette demande intervient dans le réseau où certains chemins d'opération et de protection sont déjà présents. Le problème tel qu'énoncé dans [KL00] est de trouver un chemin d'opération et son chemin de protection minimisant la bande passante totale à allouer aux deux chemins. Le tableau 3.3 introduit les notations.

Une demande k de l'ensemble $\Phi_{\ell'}^{\ell}$ est représentée par un triplet (s_k, t_k, d_k) où s_k, t_k, d_k sont la source, la destination et la bande passante requise par la demande. Alors :

$$B_{\ell'}^{\ell} = \sum_{k \in \Phi_{\ell'}^{\ell}} d_k. \tag{3.1}$$

La figure 3.1 montre la structure de la bande passante sur un lien ℓ'. La bande

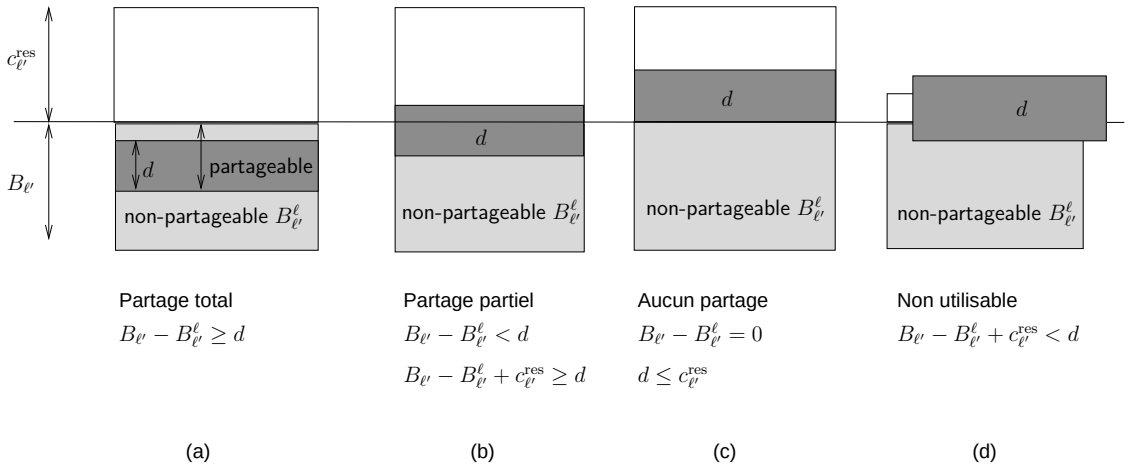

FIG. 3.2 – Bande passante de protection nécessaire sur un lien

passante totale sur ℓ' est composée de : la capacité d'opération, la capacité de protection $B_{\ell'}$ et la capacité résiduelle $c^{\text{res}}_{\ell'}$. Pour un lien d'opération ℓ donné, la capacité de protection $B_{\ell'}$ se divise en deux parties :

- $B^{\ell}_{\ell'}$: la bande passante sur ℓ' qui ne peut pas être utilisée par le chemin de protection p' de la demande courante si le lien ℓ est utilisé par son chemin d'opération p. Ceci est dû à la condition de partage. Dans le cas contraire, lors d'une panne simple sur ℓ, p' et tous les chemins de protection des demandes appartenant à l'ensemble $\Phi^{\ell}_{\ell'}$ seront simultanément activés pour remplacer leur chemin d'opération. Par conséquent, $B^{\ell}_{\ell'}$ sera requise pour p' ainsi que pour les demandes en $\Phi^{\ell}_{\ell'}$.
- $B_{\ell'} - B^{\ell}_{\ell'}$: la bande passante de protection de ℓ' qui est partageable avec le chemin de protection p' de la demande courante, dans le cas où le lien ℓ est utilisé par son chemin d'opération p.

Le chemin de protection p' peut traverser le lien ℓ' dans l'une des trois situations suivantes, illustrées dans la figure 3.2 :

- Le lien ℓ' détient suffisamment de capacité de protection partageable pour la demande courante : $B_{\ell'} - B^{\ell}_{\ell'} \geq d$, voir la figure 3.2a. Le chemin de protection p' peut emprunter ℓ' sans bande passante supplémentaire à réserver sur ce lien. Autrement dit, le coût de p' sur ℓ' est 0.

– Le lien ℓ' ne détient pas suffisamment de capacité de protection partageable pour la demande courante. Dans ce cas, p' utilise partiellement la bande passante de la capacité résiduelle du lien ℓ' : $B_{\ell'} - B^{\ell}_{\ell'} < d \leq B_{\ell'} - B^{\ell}_{\ell'} + c^{\text{res}}_{\ell'}$, voir la figure 3.2b et 3.2c. Le chemin de protection p' peut emprunter ℓ' avec un coût équivalent à la bande passante additionnelle venue de la capacité résiduelle : $d + B^{\ell}_{\ell'} - B_{\ell'}$.

– Le lien ℓ' ne détient pas suffisamment ni de capacité de protection partageable, ni de capacité résiduelle : $B_{\ell'} - B^{\ell}_{\ell'} + c^{\text{res}}_{\ell'} < d$, voir la figure 3.2d. Le chemin de protection p' ne peut donc pas emprunter ce lien. Autrement dit le coût de p' sur le lien ℓ' est ∞.

En résumé, le coût $b^{\ell}_{\ell'}$ du chemin de protection de p' d'un lien ℓ', est donc définie par :

$$b^{\ell}_{\ell'} = \begin{cases} 0 & \text{si } B_{\ell'} - B^{\ell}_{\ell'} \geq d \text{ et } \ell \neq \ell' \\ d + B^{\ell}_{\ell'} - B_{\ell'} & \text{si } B_{\ell'} - B^{\ell}_{\ell'} < d \leq B_{\ell'} - B^{\ell}_{\ell'} + c^{\text{res}}_{\ell'} \text{ et } \ell \neq \ell' \\ \infty & \text{sinon.} \end{cases} \tag{3.2}$$

Le coût du chemin d'opération d'un lien est simplement égal à la bande passante demandée car les chemins d'opération ne partagent pas de bande passante entre eux. La fonction objectif du routage, la minimisation de la bande passante utilisée par les chemins d'opération et de protection de la demande courante, est définie par :

$$\min \left(d \times \sum_{\ell} x_{\ell} + \sum_{\ell'} b^{\ell}_{\ell'} \times y_{\ell'} \right). \tag{3.3}$$

Pour résoudre ce problème, les auteurs de [KL00] ont proposé une solution exacte SCI en se basant sur la programmation linéaire. Afin de linéariser la fonction objectif, la variable $z_{\ell'}$ est introduite. Le modèle est comme suit :

Objectif :

$$\min \left(d \times \sum_{\ell} x_{\ell} + \sum_{\ell'} z_{\ell'} \right) \tag{3.4}$$

Contraintes :

$$\sum_{e\in\Gamma^+(v_i)} x_\ell - \sum_{e\in\Gamma^-(v_i)} x_\ell = \begin{cases} 1 & \text{si } v_i = s \\ 0 & \text{si } v_i \neq s,t \\ -1 & \text{si } v_i = t \end{cases} \qquad \forall v_i \qquad (3.5)$$

$$\sum_{e\in\Gamma^+(v_i)} y_\ell - \sum_{e\in\Gamma^-(v_i)} y_\ell = \begin{cases} 1 & \text{si } v_i = s \\ 0 & \text{si } v_i \neq s,t \\ -1 & \text{si } v_i = t \end{cases} \qquad \forall v_i \qquad (3.6)$$

$$z_{\ell'} \geq b^\ell_{\ell'} \times (x_\ell + y_{\ell'} - 1) \qquad \forall \ell, \ell' \qquad (3.7)$$

$$z_{\ell'} \geq 0 \qquad \forall \ell' \qquad (3.8)$$

où $\Gamma^+(v_i)$ (respectivement $\Gamma^-(v_i)$) est l'ensemble des arcs qui sortent du (resp. qui entrent dans le) noeud v_i.

La résolution de ce modèle fournit une solution exacte optimale du routage dynamique de l'OSSP. On peut observer que les coûts de protection $b^\ell_{\ell'}$ pour toutes les paires de liens ℓ, ℓ' du réseau sont des paramètres cruciaux du modèle. Les valeurs de ces coûts doivent être déterminées et mises à la disposition du noeud de calcul qui résoudra le modèle. Selon l'équation (3.2), le coût $b^\ell_{\ell'}$ dépend de $B^\ell_{\ell'}$ et $B_{\ell'}$, alors les valeurs de $B^\ell_{\ell'}$ et $B_{\ell'}$ pour toutes les paires ℓ, ℓ' devraient être également mises à la disposition du noeud de calcul. Puisque $B^\ell_{\ell'}$ est défini comme étant la bande passante totale demandée par les chemins dans l'ensemble $\Phi^\ell_{\ell'}$, le noeud de calcul doit savoir, afin de calculer $B^\ell_{\ell'}$, quels sont les chemins de protection qui passent par le lien ℓ' et quels sont leurs chemins d'opération. En d'autres termes, le noeud de calcul devrait garder en mémoire l'historique des allocations de bande passante d'opération et de protection de chaque lien du réseau. Cette demande d'information complète et globale ne peut évidemment pas être satisfaite dans un réseau multidomaines sinon on ne pourrait pas satisfaire la contrainte d'extensibilité. La solution est donc restreinte aux réseaux d'un domaine simple. De même, l'application de solutions utilisant des calculs de coût similaires est limitée aux réseaux

d'un domaine simple.

Bien que nos calculs de coût soient inspirés de ceux présentés ci-dessus, plusieurs changements spécifiques à chacun des problèmes abordés dans les chapitres qui suivront, ont été apportés pour éliminer la nécessité d'informations complètes et globales, afin de satisfaire la contrainte d'extensibilité du réseau multidomaines.

3.1.3 Protection par segments

Les tableaux 3.4, 3.5, 3.6 listent différentes approches de protection par segments dans le cadre de réseaux d'un domaine simple. La plupart des travaux, voir par exemple SLSP [HM02], OPDA [HM03] ou OSHLA [HM04a], PROMISE [XXQ03b], SHALL [LYL06], ainsi que [RKM02], [ORM05], [YZW06] et [CGYL07] réalisent un routage à deux étapes, c'est-à-dire que le chemin d'opération est recherché d'abord et après l'avoir fixé, les segments de protection sont déterminés. Ceci ressemble à l'algorithme à deux étapes dans la protection par chemins. Dans la majorité des cas, le chemin d'opération est le plus court chemin sur lequel la capacité résiduelle reste suffisante. Dans SLSP [HM02], les segments de protection sont calculés de façon séquentielle. Au contraire, dans [HM03], [XXQ03b], [TH04], [ORM05] [LYL06], ils sont déterminés conjointement ce qui permet d'améliorer la qualité des solutions. Les auteurs de [HTC04] ont proposé une solution optimale que nous avons notée SLSP-ILP. Cette solution se base sur un modèle PNE recherchant conjointement les segments d'opération et ceux de protection avec la possibilité de partage. Par contre, le modèle ne peut qu'être utilisé dans les réseaux de petite taille à cause du temps de calcul qui croît rapidement en fonction de la taille du réseau.

Les auteurs : Ho *et al.* [HM02], [HM04a], [HTC04], Tapocal *et al.* [TH04], [THVC05], Xu *et al.*, [XXQ03b] Luo *et al.* [LYL06], Lu *et al.* [LLWL06] et Cao *et al.* [CGYL07] se basent tous sur des calculs de coût de protection identiques ou similaires à celui de l'équation (3.2) de la section 3.1.2. Ils nécessitent donc que les informations soient, soit complètes, soit partielles, mais toujours globales, c'est-à-dire avec les informations d'allocation de ressources de tous les liens du réseau. Ces travaux sont donc restreints aux réseaux d'un domaine simple. Au contraire, le

Travail	Granularité de partage	Conv.	Info.	Résumé de la contribution
SLSP [HM02] de Ho *et al.*	non précisé	O	C	Routage à deux étapes. Le chemin d'opération est divisé en segments de longueurs égales, puis les segments de protection sont calculés individuellement.
[RKM02] de Ranjith *et al.*	S	O	N	Routage à deux étapes. Le chemin d'opération est divisé conjointement avec la recherche des segments de protection, *sans* tenir compte de la possibilité de partage de la bande passante de protection.
SLSP-O (ou OPDA) [HM03] ou OSHLA [HM04a] de Ho *et al.*	W	O	C	Routage à deux étapes. En fixant le chemin d'opération, les segments de protection sont recherchés de façon optimale en considérant tous les cycles de protection possibles. Un cycle de protection est formé par une paire de segments d'opération et de protection.
PROMISE de Xu *et al.* [XXQ03b]	S	O	C, P	Routage à deux étapes. Solution PNE exacte et programmation dynamique pour la recherche conjointe des segments de protection.

Conv. : Conversion de longueur d'onde.
O/N : Oui ou Non.
Info. : Quantité des informations demandée.
C/P/N : Information Complète ou Partielle ou aucune information (N) demandée.
W/S : Toute une longueur d'onde (W) ou une portion de longueur d'onde (S).

TAB. 3.4 – Travaux existants sur le routage dynamique pour la protection par segments

Travail	Granularité de partage	Conv.	Info.	Résumé de la contribution
CDR de Ho *et al.* [HTC04]	S, W	O	C	Solution heuristique. Les choix des entêtes et des queues des segments sont prédéfinis le long de K plus courts chemins. ITSA [TC03] pour chaque paire entête-queue pour trouver leurs segments d'opération et de protection.
SLSP-ILP de Ho *et al.* [HTC04]	S, W	O	C	Solution optimale. Modèle PNE pour la recherche conjointe du chemin d'opération et de ses segments de protection.
Travail de Tapocal *et al.* [TH04]	S, W	O	C	Preuve de la NP-complétude du problème OSSP. Intégration de la contrainte du temps de restauration dans la solution optimale SLSP-ILP.
SLSP-R de Ho *et al.* [HM04b]	non précisé	O	C	SLSP avec la ré-allocation de la capacité de protection en remettant en cause des groupes de segments de protection.
Travail de Ou *et al.* [ORM05]	W	O	C	Routage à deux étapes. Amélioration de [RKM02] : les segments de protection sont déterminés en tenant compte de la possibilité de partage. Les bandes passantes partageables sont calculées en supposant que tous les segments d'opération sont initialement le chemin d'opération entier et puis rajustées après que les segments d'opération ont été déterminés.

TAB. 3.5 – Travaux existants sur le routage dynamique pour la protection par segments (suite)

Travail	Granularité de partage	Conv.	Info.	Résumé de la contribution
Travail de Tapocal *et al.* [THVC05]	S, W	O	C	Solution heuristique avec prise en compte de la contrainte du temps de restauration. Une solution préliminaire sans la contrainte du temps de restauration est d'abord calculée, elle est ensuite rajustée pour la satisfaire.
SHALL de Luo *et al.* [LYL06]	S	O	C, P	Routage à deux étapes. Les segments ne doivent pas se chevaucher. Solution heuristique qui recherche conjointement des segments de protection.
Travail de Yong *et al.* [YZW06]	W	O	C	Routage à deux étapes. Une fois que le chemin d'opération est déterminé, tous ses segments d'opération et leurs segments de protection possibles sont identifiés. L'ensemble des segments de protection valides est recherché à partir de ces segments potentiels en utilisant une transformation sur un graphe.
ASSP de Lu *et al.* [LLWL06]	W	O	C	SSP pour les connexions multicasts.
Travail de Cao *et al.* [CGYL07]	S,W	O	C	Routage à deux étapes. Recherche récursive des segments de protection. Pour chaque segment de protection, toutes les entêtes et queues possibles sont essayées, le choix final ne se base pas sur les coûts de protection. Les segments ne doivent pas se chevaucher. Restriction de la longueur du segment de protection.

TAB. 3.6 – Travaux existants sur le routage dynamique pour la protection par segments (suite)

travail de Ranjith *et al.* dans [RKM02] n'utilise aucune information d'allocation de ressources. Malheureusement, cela mène au fait qu'il est incapable de tenir compte de la possibilité de partage dans son algorithme de routage. En effet, le travail de Ou *et al.* dans [ORM05] est le résultat du routage de [RKM02] avec la possibilité de partage considérée. De nouveau, [ORM05] doit utiliser des informations complètes pour calculer le coût de protection, ce qui restreint également l'utilisation de son approche aux réseaux d'un domaine simple.

3.2 Protection multidomaines

Bien que la protection pour les réseaux multidomaines est très importante, elle n'a pas encore été l'objet de beaucoup d'études. Le tableau 3.2 montre les travaux réalisés dans la littérature sur ce domaine d'études.

Les auteurs de [DLM+04] et [RMD04] proposent une protection dédiée par chemins pour les réseaux multidomaines MPLS. Les autres travaux se consacrent à la protection par segments. Intuitivement, on pense que la protection multidomaines pourrait s'effectuer à l'aide de plusieurs protections d'un domaine simple, où chaque protection recouvre une partie du chemin d'opération dans un domaine tel que proposé dans [OMZ01]. Pourtant, une telle protection laisse les liens inter-domaines et les noeuds de bord non-protégés. Le travail dans [ASL+02] essaie de surmonter cette faiblesse en ajoutant une restauration au cas où un des noeuds de bord tombe en panne. Dans un tel cas, le chemin d'opération complet est recherché et rétabli une fois que la panne est survenue. Évidemment, il n'y a aucune garantie de l'existence d'un tel chemin, avec de plus, un temps hors service supplémentaire causé par le délai de sa recherche. Par conséquent, la qualité de la protection se dégrade. Mieux centré sur la protection, le travail dans [MKAM04] propose une protection avec segments se chevauchant. Cependant, le réseau mentionné dans [MKAM04] est un réseau multidomaines spécial où les domaines sont tous reliés à un domaine fédérateur. Une connexion part du domaine source, passe par quelques liens du domaine fédérateur et arrive directement au domaine destination (voir la figure

Travail	Mode	Partage	Résumé de la contribution.
Travail de [OMZ01]	S	N	Protection par segments sans chevauchement. Les noeuds de bord sont les entêtes et les queues des segments.
Travail de [ASL+02]	S	P	Amélioration du travail de [OMZ01]. Protection par segments sans chevauchement. Un nouveau chemin d'opération est recherché lorsqu'un de ses noeuds de bord tombe en panne.
Travail de D'achille *et al.* dans [DLM+04] et de Ricciato *et al.* dans [RMD04]	P	N	Proposé pour les réseaux MPLS. Routage à deux étapes. Le chemin d'opération est le plus court chemin.
Travail de [MKAM04]	O	P	OSSP pour les réseaux multidomaines spéciaux où il n'y a pas de domaines de transit.
Travail de [HC02]	S		Routage statique : pré-planification de la capacité de protection. Les domaines sont déterminés par l'algorithme mais pas comme une contrainte administrative. Utilisation d'un cycle Hamiltonien.

Partage : Partage de la bande passante de protection P, et sans partage N
Mode : protection par chemins (Path), par Segments et avec des segments se chevauchant (Overlapping segment)

TAB. 3.7 – Travaux existants sur le routage dynamique pour la protection dans les réseaux multidomaines

3.3a). Le cas où une connexion emprunte des liens internes d'un domaine qui est différent de ces trois domaines pour transiter n'est pas considéré. Alors, le problème de routage peut être résolu principalement dans le domaine fédérateur avec des informations complètes. Nous constatons qu'en pratique, les domaines sont reliés entre eux sans domaine fédérateur ; un réseau est souvent connecté avec un autre distant par un ou plusieurs réseaux intermédiaires. Par exemple dans la figure 3.3b, N_2 et N_3 sont des domaines de transit entre N_1 et N_2. Une connexion pourrait emprunter plusieurs liens internes des domaines de transit. Dans la figure 3.3b, le chemin entre la source et la destination utilisent des liens internes des domaines N_1 et N_2. Puisque chacun de ces domaines gère uniquement ses propres informations de routage, le routage dans le réseau générique est plus difficile à réaliser que dans le réseau étudié dans [MKAM04].

Malgré les autres travaux s'adressant au routage dynamique, le travail dans [HC02] propose plutôt un routage statique pour planifier la capacité d'opération et de protection du réseau. Il construit plusieurs cycles Hamiltoniens, chacun protège des liens de son cycle. Une région formée par un cycle Hamiltonien est appelée dans ce travail un "domaine", avec une définition qui diffère de la notion habituelle de domaine administratif dans les réseaux multidomaines. Le réseau qui fait l'objet de ce travail est toujours un réseau d'un domaine simple.

3.3 Synthèse

Dans ce chapitre, nous avons parcouru les solutions de routage dynamique pour la protection partagée. Dans les sections 3.1.1 et 3.1.3, nous avons vu les approches de protection par chemins puis par segments dont l'application reste limitée aux réseaux d'un domaine simple. Dans la section 3.2, les approches de protection par chemins ainsi que par segments pour les réseaux multidomaines ont également été examinées. Malgré un certain nombre de solutions existantes, on peut constater qu'il n'y a pas encore de solution satisfaisante de protection partagée par chemins pour les réseaux multidomaines ni de solution satisfaisante de protection partagée

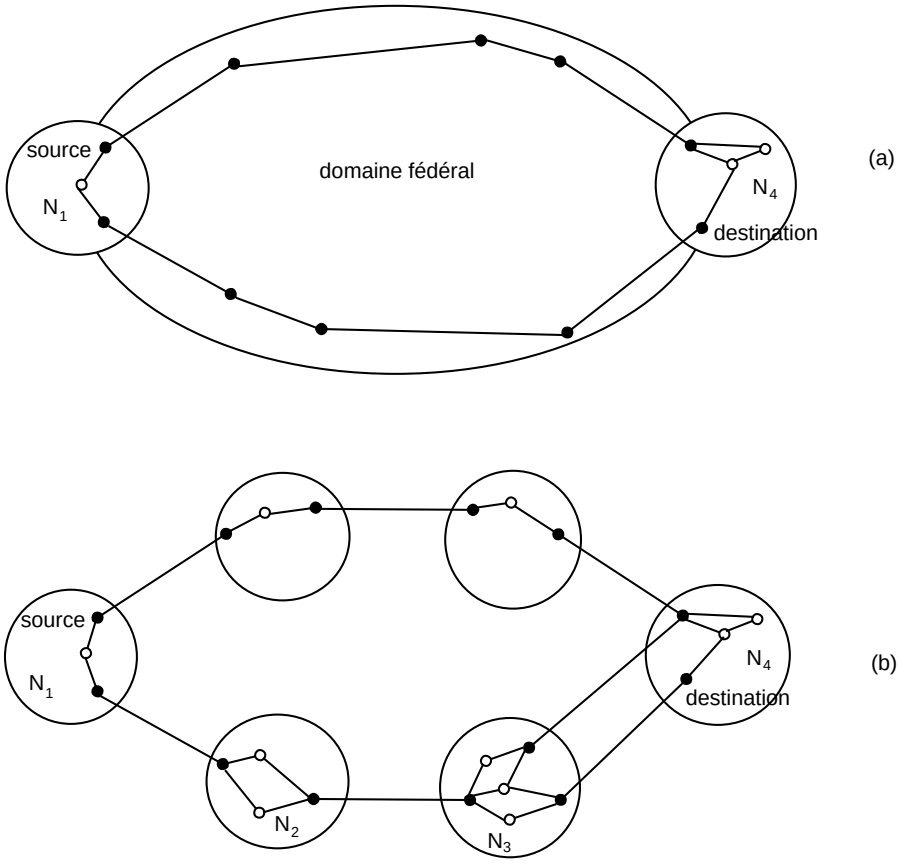

FIG. 3.3 – Le réseau multidomaines supposé dans [MKAM04] (a), et le réseau multidomaines générique

par segments dans le cas des réseaux multidomaines génériques, solution qui permettrait de protéger tous les liens et les noeuds du réseau. Nous consacrons donc cette thèse à la recherche de telles solutions.

CHAPITRE 4

DYNAMIC ROUTING FOR SHARED PATH PROTECTION IN MULTI-DOMAIN OPTICAL MESH NETWORKS

D.L. Truong and B. Thiongane

Abstract: The routing problem for shared path protection in multi-domain optical mesh networks is more difficult than that in single-domain mesh networks due to the lack of complete and global knowledge on the network topology and bandwidth allocation. To overcome this difficulty, we propose an aggregated network modeling by underestimation with a two-step routing strategy. In the first step, a rough routing solution is sketched in a virtual network which is the topology aggregation of the multi-domain network. A complete routing is then determined by solving routing problems within the original single-domain networks. The first step can be solved by either using an exact mathematical programming or an heuristic while the second step is always solved by heuristics. Computational results show the relevance of the aggregated network modeling. They also prove the scalability of the proposed routing for multi-domain networks and its efficiency in comparison with the optimal solution obtained by using the complete information scenario. In addition, we believe that short working paths lead to a higher possibility of sharing backup resources amongst backup paths. Our mathematical programming model minimizes the total requested resource and at the same time provides a short working path resulting an in additional overall resource saving.

Keywords: Multi-domain Network, Protection, Routing.

Status: Cet article a été publié dans le *Journal of Optical Networking* de l' *Optical Society of America*, vol. 5, no. 1, pages 58-74, janvier 2006.

4.1 Introduction

It has been recognized that Shared Path Protection (SPP) both protects against link and node failures and saves resources thanks to bandwidth sharing among backup lightpaths (see [RSM03]). In the single-failure scenario, two backup lightpaths could share bandwidth among them if their working lightpaths are link or node-disjoint, later called the *sharing condition*. SPP routing consists in finding a pair of working and backup lightpaths that satisfy the *sharing condition* and optimize a particular criteria such as requested bandwidth capacity, number of wavelength conversions, fiber link length, etc. This paper considers the problem of dynamic routing for SPP in multi-domain optical mesh networks while minimizing the total bandwidth required by the working and backup lightpaths. Since the node-disjoint condition is equivalent to the link-disjoint condition by splitting each node into two halves with a "virtual" directed link between them (see [QX01]), the focus will be on the link-disjoint condition. We assume that links are not bundled together and thus a failure affects at most one link (which is not the case in [BL04]). We assume also that every network node has OEO treatment so they can switch sub-wavelength and wavelength assignment is easy to handle.

There are some static (or off-line) SPP routing approaches proposed for single domain [GS98] or multi-domain [HC02] networks. Given a network with known topology, link capacities and future requested traffic, these works design fixed working and backup capacities for each link. Since network traffic changes unpredictably and frequently, a dynamic (on-line) routing without *a priori* knowledge of the network traffic is necessary.

Dynamic SPP routing identifies a pair of disjoint working and backup paths that minimally consume bandwidth according to the current network state, while satisfying the *sharing condition*. This problem has been proven NP-hard in [LMDL92]. An exact ILP-based solution called *Sharing with Complete Information* (SCI) has been proposed in [KL00], in which the total bandwidth consumed by the working and backup path is minimized. The ILP formulation requires detailed and

global information on the entire network topology and the bandwidth allocation history for each network link. The *Two-Step Approach* (TSA) [TC03] minimizes the working and backup bandwidth separately but computes it in the same way as SCI, leading to the same information requirement as SCI. In order to reduce the per-link information, *Sharing with Partial Information* (SPI) [KL03] and *Distributed Partial Information Management* (DPIM) [QX01] have been introduced. They overestimate the working and backup bandwidth consumption in comparison with SCI and apply the same ILP to minimize the total overestimated bandwidth. Later, *Active Path First-Backup Path Cost* (APF-BPC), a heuristic-based routing using partial information scenarios of DPIM and SPI, was proposed in [XQX02]. In all cases, the *global knowledge* (either partial or complete) on each link and the complete network topology are mandatory at the network ingress nodes.

In multi-domain networks, it is impractical to make this global information available at a node. A multi-domain network is an interconnection of several independent single-domain networks [BSO02] (Figure 4.1a). To support the scalability, the routing information should not be excessively and frequently exchanged throughout the multi-domain network [LRVB04]. The detailed connectivity and bandwidth allocation of a domain is limited within itself, and only aggregated information can be exposed to external domains. As a result, no node is aware either of the global multi-domain network topology or the bandwidth allocation on all network links. We call this constraint the *"scalability constraint"*. It makes the above listed solutions inapplicable to multi-domain networks.

A few works have been proposed on dynamic protection for multi-domain networks but none have been devoted to SPP. *No-sharing* path protection was proposed in [DLM$^+$04] while *no-sharing* segment protection was introduced in [OMZ01]. The latter was improved in [ASL$^+$02] to become segment-shared protection although no details on its routing model were described. In [MKAM04], a routing for segment shared protection was proposed where a lightpath is not allowed to pass through any domain. In real multi-domain network, lightpaths often pass through many domains. This is illustrated in Figure 4.1a where a lightpath from domain $\mathcal{N}_1$ to

$\mathcal{N}_3$ can pass through $\mathcal{N}_2$.

This paper deals with SPP routing in multi-domain networks without global infor-

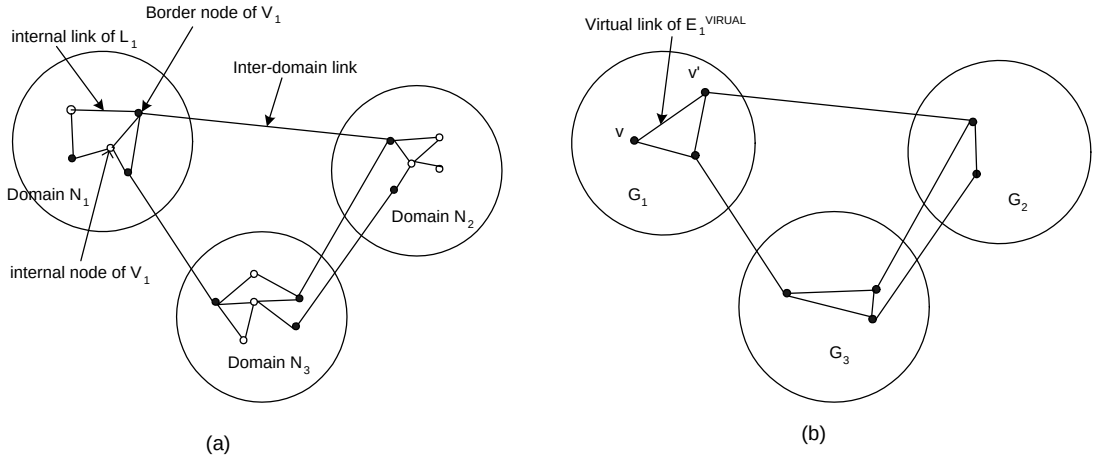

Figure 4.1: A multidomain network (a) and its *inter-domain network* (b) obtained by topology aggregation.

mation knowledge. Our main idea is to transform the original multi-domain routing problem into several single-domain routing problems which are solved separately by using adapted versions of existing single-domain SPP routings on underestimated information. We propose a two-step routing strategy. First, the multi-domain network is topologically aggregated to become a single-domain network, called *inter-domain network*, in which a rough routing is sketched out. A detailed routing is then determined within each original single-domain network. The use of aggregate information at the first step removes the global information requirement and thus preserves scalability. We propose two approaches to realize the routing strategy. The two approaches are compared through computational results. To evaluate the relevance of the aggregate information, the approaches are compared to SCI when the latter is executed on multi-domain networks. Also note that while DPIM and SPI try to reduce the amount of required per-link information, we concentrate on reducing the details of the information to be advertised from a domain and the frequency of information exchanged between domains as well as within domains.

The final objective is to respect the *scalability constraint*.

This paper is organized as follows: the next section introduces the notation and the two-step routing strategy. In Section 4.3, the cost functions are defined using aggregate information. The two approaches to realize the two-step routing strategy are presented in Section 4.4. Section 4.5 presents the routing signaling that co-ordinates the two routing steps and the routing information update. Section 4.6 shows the computational results on a multi-domain network built from real single-domain networks. Finally, Section 4.7 concludes the paper.

4.2 Notation and Two-step Routing Strategy

The multi-domain network is represented by a graph $\mathcal{N} = (V, L)$ composed of M connected single-domain networks $\mathcal{N}_i = (V_i, L_i)$, $i = 1, .., M$. The set V (resp. V_i) and L (resp. L_i) are respectively the set of nodes and the set of links of $\mathcal{N}$ (resp. $\mathcal{N}_i$). Each single-domain set of nodes V_i decomposes into the border nodes V_i^{BORDER} and the core nodes V_i^{CORE}. Moreover, note that L decomposes into L^{INTRA} and L^{INTER}. $L^{\text{INTER}} = \{(v, v') \in L : v \in V_i^{\text{BORDER}}, v' \in V_j^{\text{BORDER}} \neq V_i^{\text{BORDER}}\}$ is the set of inter-domain links where an inter-domain link connects two border nodes of two different domains. On the other hand, $L^{\text{INTRA}} = \bigcup_{i=1..M} L_i$ is the set of links within domains. A clique mesh topology aggregation will be applied to $\mathcal{N}_i$, $i = 1, .., M$, to obtain an aggregated graph $G_i = (V_i^{\text{BORDER}}, E_i^{\text{VIRTUAL}})$ containing only border nodes of $\mathcal{N}_i$ and the set of virtual links connecting all pairs of border nodes $E_i^{\text{VIRTUAL}} = \{(v, v') : v, v' \in V_i^{\text{BORDER}}\}$. The resulting network $G = (V^{\text{BORDER}}, E)$ is a compact *inter-domain network* (see an illustration on Figure 4.1b) where V^{BORDER} contains all border nodes of $\mathcal{N}$ and E contains all virtual links E^{VIRTUAL} and inter-domain links L^{INTER}:

$$V^{\text{BORDER}} = \bigcup_{i=1..M} V_i^{\text{BORDER}}$$

$$E^{\text{VIRTUAL}} = \bigcup_{i=1..M} E_i^{\text{VIRTUAL}}$$

$$E = E^{\text{VIRTUAL}} \cup L^{\text{INTER}}$$

We will denote by e an edge of G and ℓ a fiber link of $\mathcal{N}$. Thus an inter-domain link can be denoted by e or ℓ.

When e is a virtual link between v and $v' \in \mathcal{N}_i$, we define $\mathcal{P}_e$ as the set of physical paths within $\mathcal{N}_i$ between v and v', and $\mathcal{P}_e = \{e\}$ when e is an inter-domain link. An element of $\mathcal{P}_e$ is an instance of e. A link e will be associated with a link-state representing some routing information obtained from all elements of $\mathcal{P}_e$. This link-state will be disseminated to all multi-domain network border nodes. Thus, these border nodes have a common aggregated view of the multi-domain network. More details are given in Sections 4.3 and 4.5.

Let us consider a new request of bandwidth d from a source border node v_s to a destination border node v_d. The requested bandwidth will be routed over a single path. The following notation is introduced where roman letters are for the original network $\mathcal{N}$ meanwhile greek letters are for the aggregated network G.

p, p' are respectively the complete working and backup paths in $\mathcal{N}$ to find for the new request.

c_ℓ^{res} is the residual bandwidth capacity on physical link $\ell \in L$.

a_ℓ is the bandwidth the working path p will consume on physical link $\ell \in L$. Evidently, $a_\ell = d$ if there is sufficient residual capacity on ℓ.

$B_{\ell'}$ is the bandwidth reserved on physical link $\ell' \in L$ by existing backup paths.

$B_{\ell'}^{\ell}$ is the bandwidth reserved on physical link $\ell' \in L$ by existing backup paths that protect the working paths passing through link $\ell \in L$. This bandwidth is not sharable for protecting any new working path containing ℓ.

$B_{\max}^{\ell}$ is the maximal backup bandwidths reserved on a network link for protecting the working paths that pass through link $\ell \in L$. Indeed, $B_{\max}^{\ell} = \max\limits_{\ell' \in L} B_{\ell'}^{\ell}$.

$B_{\max}^{q}$ and $B_{\max}^{p}$ are also defined as $B_{\max}^{q} = \max\limits_{\ell \in q} B_{\max}^{\ell}$ and $B_{\max}^{p} = \max\limits_{\ell \in p} B_{\max}^{\ell}$.

$b^{\ell}_{\ell'}$, $b^{q}_{\ell'}$ and $b^{p}_{\ell'}$ are respectively the additional backup bandwidths to reserve beside $B_{\ell'}$ on physical link ℓ' for protecting the new working path p against single failures on link ℓ, sub-path q and entire p. Observe that, $b^{q}_{\ell'} = \max_{\ell \in q} b^{\ell}_{\ell'}$ and $b^{p}_{\ell'} = \max_{\ell \in p} b^{\ell}_{\ell'}$.

$b^{\ell}_{q'}$, $b^{q}_{q'}$ and $b^{p}_{q'}$ are the overall additional backup bandwidths to reserve along sub-path q' for protecting the new working path p against single-failures on link ℓ, sub-path q and entire p. Hence, $b^{\ell}_{q'} = \sum_{\ell' \in q'} b^{\ell}_{\ell'}$, $b^{q}_{q'} = \sum_{\ell' \in q'} b^{q}_{\ell'}$ and $b^{p}_{q'} = \sum_{\ell' \in q'} b^{p}_{\ell'}$.

π, π' are the representations of p and p' in G. They are composed of virtual and inter-domain links. They are called the directive working and backup paths.

$\mathcal{P}_{\pi}$ (resp. $\mathcal{P}_{\pi'}$) is the set of physical paths obtained by substituting all virtual links of π (resp. π') by their instances. Clearly, $p \in \mathcal{P}_{\pi}$, $p' \in \mathcal{P}_{\pi'}$.

α_e is the total bandwidth that p will consume along its sub-path represented by virtual/inter-domain link $e \in E$. Thus, $\alpha_e = \sum_{\ell \in q} a_{\ell}$ where q is the sub-path.

$\beta^{e}_{e'}$ (resp. $\beta^{\pi}_{e'}$) is the overall additional backup bandwidth to reserve along the sub-path represented by link $e' \in E$ in order to protect p against single-failures on its sub-path represented by $e \in E$ (resp. on entire p). Thus, $\beta^{e}_{e'} = b^{q}_{q'}$ and $\beta^{\pi}_{e'} = b^{p}_{q'}$ where q, q' are the sub-paths in $\mathcal{N}$ represented by e, e'.

γ^{res}_{e} is the maximum bandwidth that can be routed over an instance of $e \in E$. $\gamma^{\text{res}}_{e} = \max_{q \in \mathcal{P}_e} \min_{\ell \in q} c^{\text{res}}_{\ell}$. It is called the residual capacity on e.

$\|e\|$ is the length of the shortest instance of $e \in E$. $\|e\| = \min_{q \in \mathcal{P}_e} \|q\|$, where $\|q\|$ is length of q in number of hops.

The parameters a, α and b, β with different indexes are also called *working* and *backup costs*.

Dynamic SPP routing aims to identify, for a request, a working path p and a backup

path p' that are disjoint and minimize the total consumed bandwidth:

$$\min \sum_{\ell \in p} a_\ell + \sum_{\ell' \in p'} b^p_{\ell'}. \tag{4.1}$$

According to the definition of α_e and $\beta^\pi_{e'}$, (4.1) is equivalent to

$$\min \sum_{e \in \pi} \alpha_e + \sum_{e' \in \pi'} \beta^\pi_{e'}. \tag{4.2}$$

We propose then the following two-step routing strategy

- **Inter-domain routing step**: This step is performed on the inter-domain network. The source border node computes π and π' in G while minimizing their bandwidth consumption

$$\min \sum_{e \in \pi} \alpha_e + \sum_{e' \in \pi'} \beta^\pi_{e'}. \tag{4.3}$$

- **Intra-domain routing step**: At this step, the virtual links of π and π' are replaced by physical paths to build the complete working and backup paths. Virtual link e is mapped with (replaced by) one of its instances in $\mathcal{P}_e$. A joint mapping of all virtual links would help to maintain the optimal bandwidth cost obtained at the inter-domain step but involves simultaneously many domains and thus requires global information. Therefore, we first map the virtual links of π and then those of π'. The path instance $q \in \mathcal{P}_e$ that is mapped with the working virtual link $e \in \pi$ should minimize α_e:

$$\min_{q \in \mathcal{P}_e} \sum_{\ell \in q} a_\ell. \tag{4.4}$$

  The path instance $q' \in \mathcal{P}_{e'}$ that is mapped with the backup virtual link $e' \in \pi'$

should minimize $\beta^{\pi}_{e'}$, i.e.,

$$\min_{q' \in \mathcal{P}_{e'}} b^{p}_{q'} = \min_{q' \in \mathcal{P}_{e'}} \sum_{\ell' \in q'} b^{p}_{\ell'} \tag{4.5}$$

Note that the mapping of a virtual link of E_i^{VIRTUAL} involves only its instances in $\mathcal{N}_i$, hence could be performed within the single-domain network $\mathcal{N}_i$ by one border node of the virtual link.

It remains to identify the parameters α_e, $\beta^{\pi}_{e'}$, a_ℓ and $b^{p}_{\ell'}$ and solve the minimization problems (4.3), (4.4) and (4.5). The next section shows how the parameters are identified. Section after presents the algorithms for solving the minimization problems.

4.3 Working and Backup Costs

Until the complete working and backup paths are being identified, the costs α_e and $\beta^{\pi}_{e'}$, which are used in inter-domain routing, cannot be computed exactly but only estimated. In order to satisfy the *scalability constraint*, the estimation should not use the complete and detailed information on each network link. The computation of a_ℓ, $b^{p}_{\ell'}$, which are used by the intra-domain routing, is under the same context.

4.3.1 Underestimation of Working and Backup Costs for Inter-domain Routing

The ultimate goal of the estimations is to relax the dependency of the exact values of α_e, $\beta^{\pi}_{e'}$ on global and detailed information about physical links inside domains. These values will be represented as functions of some domain aggregated information which will become link-states of virtual/inter-domain links.
We underestimate the working cost of link $e \in E$ as the minimal overall bandwidth

that p should consume along e:

$$\alpha_e \simeq \min_{q \in \mathcal{P}_e} \sum_{\ell \in q} a_\ell.$$

Thus:

$$\alpha_e \simeq \begin{cases} \|e\| d & \text{if } d \leq \gamma_e^{\text{res}}, e \in E^{\text{VIRTUAL}} \\ d & \text{if } d \leq \gamma_e^{\text{res}}, e \in L^{\text{INTER}} \\ \infty & \text{otherwise.} \end{cases} \tag{4.6}$$

The estimation of $\beta_{e'}^{\pi}$ is more complicated, let us begin with $b_{\ell'}^{\ell}$. Note that, $b_{\ell'}^{\ell}$, the additional bandwidth to reserve, is the difference between the required bandwidth and the sharable backup bandwidth on ℓ'. The sharable backup bandwidth on link ℓ' for protecting link ℓ is $B_{\ell'} - B_{\ell'}^{\ell}$ see [KL00] for the details. As $b_{\ell'}^{\ell}$ must be non-negative, we have:

$$b_{\ell'}^{\ell} = \max\{0, B_{\ell'}^{\ell} + d - B_{\ell'}\}. \tag{4.7}$$

Here, the detailed information $B_{\ell'}^{\ell}$ is still required (as in SCI). To avoid this, $B_{\ell'}^{\ell}$ is overestimated as in [QX01] by $B_{\max}^{\ell}$. Note that $b_{\ell'}^{\ell}$ cannot be greater than the requested bandwidth:

$$b_{\ell'}^{\ell} \simeq \min\{\max\{0, B_{\max}^{\ell} + d - B_{\ell'}\}, d\}. \tag{4.8}$$

From this estimation, it can be proven that the backup cost of a virtual or inter-domain link to protect a working path is not smaller than the cost of protecting any virtual/inter-domain link of the path (see Appendix A):

$$\beta_{e'}^{\pi} \simeq \max_{e \in \pi} \beta_{e'}^{e}. \tag{4.9}$$

Now what we need to compute is $\beta^e_{e'}$. We underestimate $\beta^e_{e'}$ by the minimum backup bandwidth that p' should reserve along e':

$$\beta^e_{e'} \simeq \min_{q \in \mathcal{P}_e, q' \in \mathcal{P}_{e'}} b^q_{q'}. \tag{4.10}$$

The computational effort for $\beta^e_{e'}$ when e is virtual link may be increasing while its value might have no impact on the max of (4.9) if it is not the greatest element of the max. Therefore, we ignore it by defining $\beta^e_{e'} = 0$, for all $e' \in E, e \in E^{\text{VIRTUAL}}$.
All that remains is to estimate the two following cases of $\beta^e_{e'}$: $\beta^\ell_{e'}$, $\ell \in L^{\text{INTER}}, e' \in E^{\text{VIRTUAL}}$ and $\beta^\ell_{\ell'}$, $\ell, \ell' \in L^{\text{INTER}}$.
In the first case, according to (4.10): $\beta^\ell_{e'} \simeq \min_{q' \in \mathcal{P}_{e'}} b^\ell_{q'}$. Suppose that $e' \in \mathcal{N}_i$ and let $\overline{B}$ be the maximum backup bandwidth reserved on a link of the domain $\mathcal{N}_i$:

$$\overline{B} = \max_{\ell' \in L_i} B_{\ell'}. \tag{4.11}$$

Then, combining with the definition of $b^\ell_{q'}$ and (4.8) we have:

$$b^\ell_{q'} \geq \|q'\| \min\{\max\{0, B^\ell_{\max} + d - \overline{B}\}, d\}.$$

$$\beta^\ell_{e'} \geq \min_{q' \in \mathcal{P}_{e'}} \|q'\| \min\{\max\{0, B^\ell_{\max} + d - \overline{B}\}, d\}.$$

Thus, $\beta^\ell_{e'}$ can be underestimated by:

$$\beta^\ell_{e'} \simeq \|e'\| \min\{\max\{0, B^\ell_{\max} + d - \overline{B}\}, d\}. \tag{4.12}$$

Combining with the capacity constraint, $\beta^{\ell}_{e'}$ for $\ell \in L^{\text{INTER}}$, $e' \in E^{\text{VIRTUAL}}$ is defined:

$$\beta^{\ell}_{e'} \simeq \begin{cases} 0 & \text{if } B^{\ell}_{\max} + d \leq \overline{B} \\ \|e'\|(B^{\ell}_{\max} + d - \overline{B}) & \text{if } B^{\ell}_{\max} + d > \overline{B} > B^{\ell}_{\max}, \gamma^{\text{res}}_{e'} \geq B^{\ell}_{\max} + d - \overline{B} \\ \|e'\|d & \text{if } B^{\ell}_{\max} \geq \overline{B}, \gamma^{\text{res}}_{e'} \geq d \\ \infty & \text{otherwise.} \end{cases} \tag{4.13}$$

In the second case, $\beta^{\ell}_{\ell'} = b^{\ell}_{\ell'}$ for $\ell, \ell' \in L^{\text{INTER}}$, and is defined by (4.8):

$$\beta^{\ell}_{\ell'} \simeq \begin{cases} 0 & \text{if } B^{\ell}_{\max} + d \leq B_{\ell'}, \ell \neq \ell' \\ B^{\ell}_{\max} + d - B_{\ell'} & \text{if } B^{\ell}_{\max} + d > B_{\ell'} > B^{\ell}_{\max}, c^{\text{res}}_{\ell'} \geq B^{\ell}_{\max} + d - B_{\ell'}, \ell \neq \ell' \\ d & \text{if } B^{\ell}_{\max} \geq B_{\ell'}, c^{\text{res}}_{\ell'} \geq d, \ell \neq \ell' \\ \infty & \text{otherwise.} \end{cases} \tag{4.14}$$

Note that instead of $\overline{B}$, other estimations can be used. For example, a less coarse estimation can be obtained by using $\overline{B}_{e'} = \max_{q' \in \mathcal{P}_{e'}} \max_{\ell' \in q'} B_{\ell'}$. It is also possible to consider $\hat{B}_{e'}$, the greatest among all medians of $\{B_{\ell'}, \ell' \in q'\}$, $\forall q' \in \mathcal{P}_{e'}$. At that time, $\beta^{e}_{e'}$ would be estimated by $\frac{1}{2}\|e'\| \min\{\max\{0, B^{\ell}_{\max} + d - \hat{B}_{e'}\}, d\}$. In both cases, the computation effort increases while the scalability decreases.

In summary, the working and backup costs of a virtual or inter-domain link in G are estimated by using only per virtual/inter-domain link values (instead of per link values) such as $\|e\|$, γ^{res}_{e} (or c^{res}_{ℓ}), $\overline{B}$ (or B_{ℓ}) and $B^{\ell}_{\max}$. They are defined as link-state attributes of virtual or inter-domain links.

4.3.2 Computation of Working and Backup Costs for Intra-domain Routing

The working cost a_{ℓ} of the link ℓ is defined as:

$$a_{\ell} = \begin{cases} d & \text{if } d \leq c^{\text{res}}_{\ell} \\ \infty & \text{otherwise.} \end{cases} \tag{4.15}$$

From (4.8) and the definition of $B^p_{\max}$, it is easy to deduce that: $b^p_{\ell'} = \min\{\max\{0, B^p_{\max} + d - B_{\ell'}\}, d\}$, i.e.:

$$b^p_{\ell'} = \begin{cases} 0 & \text{if } B^p_{\max} + d - B_{\ell'} \leq 0 \\ B^p_{\max} + d - B_{\ell'} & \text{if } B^p_{\max} + d > B_{\ell'} > B^p_{\max}, c^{\text{res}}_{\ell'} \geq B^p_{\max} + d - B_{\ell'} \\ d & \text{if } B^p_{\max} \geq B_{\ell'}, c^{\text{res}}_{\ell'} \geq d \\ \infty & \text{otherwise.} \end{cases} \tag{4.16}$$

Hence, the intra-domain routing requires b_ℓ, c^{res}_ℓ of every link ℓ in the domain, and $B^\ell_{\max}$ of every link ℓ of p for computing $B^p_{\max}$.

4.4 Routing Approaches

We propose two approaches to solve the minimization problems (4.3), (4.4) and (4.5). The intra-domain step is identical but the inter-domain step is different in the two approaches. The approaches are named according to their inter-domain routing.

4.4.1 Working Path First (WPF)

The working path is routed first. All Shortest Path problems are in term of cost.

- **Inter-domain routing step**: Instead of minimizing (4.3), we separately minimize each term of the sum. First, the directive working path is set to the shortest path in G between the source and the destination when the working cost α_e is assigned to each link of G. Subsequently, the backup cost $\beta^\pi_{e'}$ is assigned to each link of G. The backup directive path is then set to the shortest path in G between the source and the destination. Note that even when π and π' share a virtual link, their complete paths could still be link disjoint. However, π and π' must be inter-domain link disjoint. This constraint is taken into account in the definition of $\beta^\ell_{\ell'}$.

- **Intra-domain routing step**: First, virtual links of π are mapped one by one within their domains. For mapping the virtual link $e \in E_i^{\text{VIRTUAL}}$ between v and $v' \in \pi$, we search for the shortest path between v and v' in the domain $\mathcal{N}_i$ when physical links of $\mathcal{N}_i$ are weighted by a_ℓ. Once the complete working path p is then obtained, the virtual links of π' are mapped similarly but with the backup cost $b^p_{\ell'}$. Again, disjointness is taken into account through the definition of $b^p_{\ell'} = \infty$ for each physical link ℓ' in the working path.

The Shortest Path problems are solved using Dijkstra's algorithm see e.g. [AMO93]. The request is rejected if one step fails to find paths.
Note that the intra-domain routing of the working path is independent of the inter-domain routing of the backup path. An alternative procedure would be to route completely the working path first, then route the directive backup path and finally map the backup virtual links. In that case, $\beta^e_{e'} \simeq \|e'\| \min\{\max\{0, B^q_{\max} + d - \overline{B}\}, d\}$ for $e \in E^{\text{VIRTUAL}}$ may be obtained in a similar way that we did for $\beta^\ell_{e'}$. This routing is called Complete Working Path First (CWPF) and will not be further developed because its experimental results are similar to those of WPF.

4.4.2 Joint Computing of Directive Paths (JDP)

In this approach, the directive working and backup paths are jointly computed by a mathematical programming. Here we consider each link of E as two directed arcs. However we still keep the notations e and E but the former will represent an arc while the latter denotes the set of arcs. Given $v_i \in V^{\text{BORDER}}$, $\Gamma^+(v_i)$ (resp. $\Gamma^-(v_i)$) denotes the set of outgoing (resp. incoming) arcs at node v_i. We introduce the following notation: x_e=1 if the directive working path π from v_s to v_d goes through arc e, 0 otherwise, y_e=1 if the directive backup path π' from v_s to v_d goes through arc e, 0 otherwise. JDP takes the following procedure:

- **Inter-domain routing step**: We solve an ILP problem (P) defined in the *inter-domain network* G to find π and π' for each lightpath request.

- **Intra-domain routing step**: Similar to the intra-domain routing of WPF.

The lightpath request is rejected if a solution is not found at one of two steps. The ILP formulation (P) for the inter-domain routing step is similar to the one proposed in [KL00], [KL03]:

$$\min \sum_{e \in E} \alpha_e x_e + \sum_{e' \in E} z_{e'} + \nu \sum_{e \in E} x_e + \mu \sum_{e' \in E} y_{e'}$$

subject to:

$$\sum_{e \in \Gamma^+(v_i)} x_e - \sum_{e \in \Gamma^-(v_i)} x_e = \begin{cases} 1 & v_i = v_s \\ 0 & v_i \neq v_s, v_d \\ -1 & v_i = v_d \end{cases} \tag{4.17}$$

$$\sum_{e' \in \Gamma^+(v_i)} y_{e'} - \sum_{e' \in \Gamma^-(v_i)} y_{e'} = \begin{cases} 1 & v_i = v_s \\ 0 & v_i \neq v_s, v_d \\ -1 & v_i = v_d \end{cases} \tag{4.18}$$

$$z_{e'} \geq \beta^e_{e'}(x_e + y_{e'} - 1) \quad e, e' \in E \tag{4.19}$$

$$z_e \geq 0 \quad e \in E \tag{4.20}$$

$$x_e, y_{e'} \in \{0, 1\} \quad e, e' \in E \tag{4.21}$$

The first two terms of the objective function are respectively the cost of the working and backup paths. The cost of the complete paths may be far from that of the directive paths when the number of virtual links increases. Therefore, the last two terms are added to favor short directive paths among those with the same total path cost and thus to limit the number of virtual links. When costs α and β are integers, and ν, μ are sufficiently small so that $\nu \sum_{e \in E} x_e + \mu \sum_{e' \in E} y_{e'} < 1$, it can be easily seen that the solution of (P) is the directive working and backup path pair with the smallest total weighted lengths among those minimizing the total consumed bandwidth. In Section 4.6 we will study the impact of the working and backup path lengths on the cost and the blocking rate.

The two sets of constraints (4.17) and (4.18) are respectively flow conservation constraints for the working path and the backup path. Each set represents a path

from the source border node v_s to the destination border node v_d in G. The parameter z_e is in fact the backup cost β_e^π and is modeled through constraint (4.19).

The links with insufficient residual capacities are automatically excluded from the working and backup paths because their α_e and $\beta_{e'}^e$ are infinity. Once again, the disjoint constraint is taken care by the definition of $\beta_{\ell'}^\ell$ as in WPF.

Besides, as a solution of (P) is defined in the directed graph, $\beta_{e'}^e$ is also updated according to the opposite direction of the arcs e and e'.

4.5 Routing Signaling and Routing Information Update

4.5.1 Routing signaling

The directive working and backup paths are both computed by the source border node. Once finished, the source node asks the border nodes along the working path to map the working virtual links with physical paths. The working segments q and their corresponding $B_{\max}^q$ that are found are returned to the source node. Finally the source identifies $B_{\max}^p$ as the maximum of all $B_{\max}^q$ and sends it to the border nodes along the directive backup path. These nodes use $B_{\max}^p$ to perform the intra-domain routing for mapping their backup virtual links with physical paths.

4.5.2 Routing Information Distribution

Once the routing is completed, the paths are setup and the link states of all physical as well as virtual/inter-domain links are updated. It is worth noting that these link states are stored in a distributed way at different border nodes. A border node also keeps the link-state $\{c_\ell^{\text{res}}, B_\ell, B_{\max}^\ell\}$ of each internal link ℓ of its domain and the link-state $\{\|e\|, \gamma_e^{\text{res}}, \overline{B}\}$ for all adjacent virtual links e. In addition, each internal or border node keeps the set $\mathbb{B}_{\ell'} = \{B_{\ell'}^\ell : \ell \in L\}$ for each link ℓ' and $\mathbb{B}^\ell = \{B_{\ell'}^\ell : \ell' \in L\}$ for each link ℓ adjacent to it. The former set is necessary to compute the exact backup bandwidth to reserve by using (4.7) if the backup path goes through ℓ'. The latter one allows the computation of $B_{\max}^\ell$ if the working path

goes through ℓ.

4.5.3 Routing Information Update through Path Setup Process

The working path will be set up first, then the backup path. For setting up the working path, a signaling message propagates along the working path from the source to the destination carrying the complete working and backup paths. Each node along the working path subsequently makes a cross-connection and updates the set $\mathbb{B}^\ell$ and the link-state $\{c_\ell^{\text{res}}, B_\ell, B_{\max}^\ell\}$ where ℓ is an adjacent working link. The new link states is collected with the signaling message until the domain's egress border node. Here these link states are forwarded to other domain border nodes to synchronize them. The number of update messages is $O(|V_i^{\text{BORDER}}|)$ where $\mathcal{N}_i$ is the current domain. The process continues until the destination is reached.

For reserving the backup path, a similar process is performed from the destination back to the source. However, no cross-connection is made. The backup bandwidth is just reserved by updating $\{c_{\ell'}^{\text{res}}, B_{\ell'}, \mathbb{B}_{\ell'}\}$ on each backup link ℓ'. The number of update messages is also $O(|V_i^{\text{BORDER}}|)$.

Finally, every border nodes locally update the link states $\{\|e\|, \gamma_e^{\text{res}}, \overline{B}\}$ of their virtual/inter-domain links and exchange these link states to each other. The number of exchange messages is $O(|V^{\text{BORDER}}|^2)$.

It is important to emphasize that with the exception of the flow of signaling messages, the routing information update is only performed through communication between border nodes. The overall number of update messages required after a lightpath request is $O(|V^{\text{BORDER}}|^2 + \sum_{j=1}^{K} |V_i^{\text{BORDER}}|^2) = O(|V^{\text{BORDER}}|^2)$, where K is the number of domains crossed either by the working or backup path. Indeed $|V^{\text{BORDER}}| = \sum_{j=1}^{M} |V_i^{\text{BORDER}}|$, thus $|V^{\text{BORDER}}|^2 > \sum_{i=1}^{K} |V_i^{\text{BORDER}}|^2$. The size of each message is always O(1).

Clearly, $O(|V^{\text{BORDER}}|^2)$ is smaller than the number of update messages in the single-domain SPP approaches which is $O(|V||V^{\text{BORDER}}|)$, since an all-to-border node up-

date is required. This proves that our approach is more scalable than the single-domain approaches [KL00], [TC03], [KL03], [XQX02]. The scalability of our approach can be improved if link-state updates are performed only once after several requests. In the next section we will analyze the impact on routing quality.

4.6 Computational results

In this section, we will evaluate the relevance of the information aggregation scenario and the efficiency of WPF compared to JDP. For the relevance of the information aggregation scenario, we compare the results of the proposed two-step routing on a multi-domain network with those of the complete information scenario SCI [KL03] on the same network.

The computational results are conducted on a five-domain network. The five domains are real optical networks: EONet [OSYZ95], RedIRIS [RED05], GARR [GAR05], Renater3 [REN05], SURFnet [SUR05] with real link capacities for the last four networks. Some inter-domain links have been added with OC-192 capacities (see Figure 4.2) for connecting different domains. Requests are randomly generated between border nodes and the requested bandwidth is uniformly distributed among OC-{1, 3, 6, 9, 12}.

JDP (ν,μ) will be used to denote the configurations of the JDP with fixed parameters ν,μ. Configurations with shorter directive working paths are expected to give also shorter complete working paths leading to more possibility of sharing backup bandwidth.

We tested WPF when $\overline{B}$, $\overline{B}_{e'}$ and $\hat{B}_{e'}$ are used. In all cases, the results are very close. We will only present the results of WPF with the $\overline{B}$ estimation .

The commercial software CPLEX and the academic version of OPNET Modeler are respectively used to implement JDP and WPF on a 1.9-GHz Pentium 4. The computational time for routing a request is less than 16 ms for WPF and less than 1 minute for JDP (ν,μ).

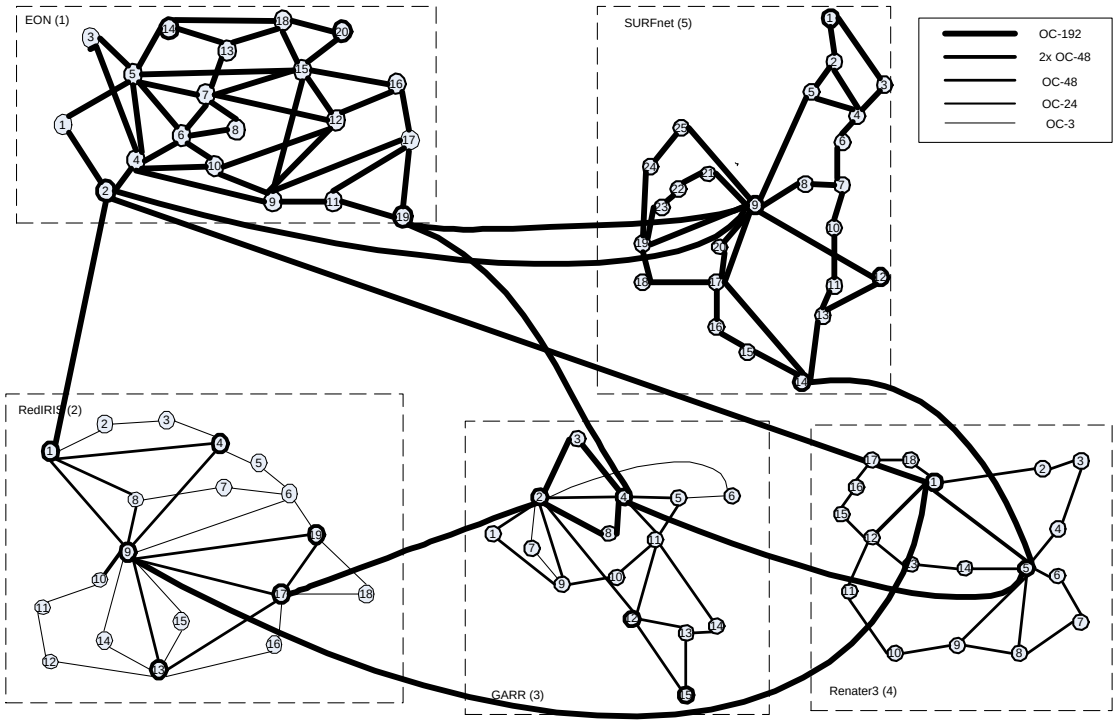

Figure 4.2: Experimental network

4.6.1 Analysis of bandwidth costs

In order to determine how WPF and JDP are far from the optimal solution in terms of bandwidth savings, we compared the total working and backup path costs found by WPF and JDP with SCI. Recall that SCI does not satisfy the scalability constraint. Let $cost^r_{\text{WPF}}$ (resp. $cost^r_{\text{JDP}}$) be the total bandwidth cost of the complete working and backup paths in the case of WPF (resp. JDP), and $cost_{\text{SCI}}$ the total cost of SCI. The relative gap between $cost^r_{\text{WPF}}$ and $cost^r_{\text{SCI}}$ is defined by:

$$gap_{\text{WPF/SCI}} = \frac{cost^r_{\text{WPF}} - cost_{\text{SCI}}}{cost_{\text{SCI}}}$$

and similarly for $gap_{\text{JDP/SCI}}$. Figure 4.3a depicts the distribution of $gap_{\text{WPF/SCI}}$ and $gap_{\text{JDP/SCI}}$. In this figure, the column at abscissa 0.5, for example, represents the percentage of cases that the gap is in the range $]0.25, 0.5]$. Note that the gap is only computed for the requests that are successfully routed by SCI and either WPF or JDP. Figure 4.3a shows that the cost of SCI is generally smaller than that of WPF

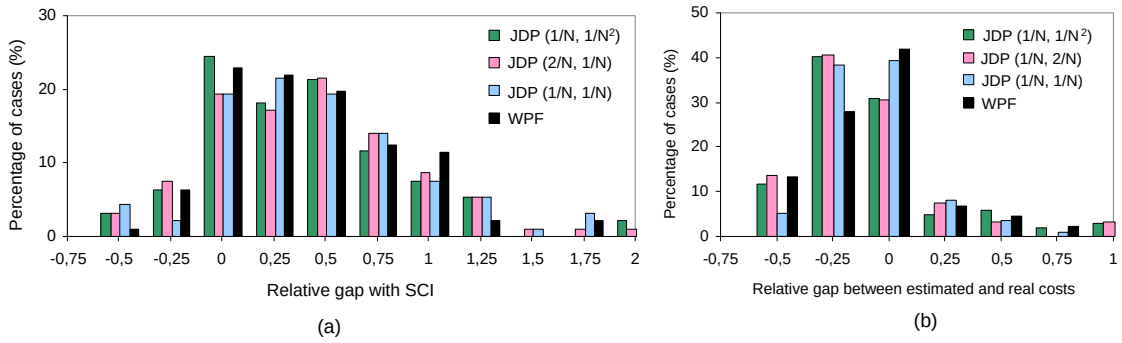

Figure 4.3: Distribution of the relative gap with SCI (a) and the relative gap between the estimated and the real costs for WPF and JDP (b).

and JDP since the gap is positive most of the time. This is a natural observation since the routing in SCI is performed within a complete information scenario. Another observation is that the percentages of cases where the gap is within $]-0.5, 0.5]$ for JDP $(\frac{1}{N}, \frac{1}{N})$, JDP $(\frac{1}{N}, \frac{1}{2N})$, JDP $(\frac{1}{N}, \frac{1}{N^2})$ and WPF are respectively 62, 65, 70

and 70, where $N = |E|$. Thus, most of the time the real cost of the solution found by WPF and by JDP is not so far from the solution found by SCI.

The directive routing given by the inter-domain routing step is accurate if the total estimated cost of the working and backup paths is closed to the total real cost obtained once the routing has been completed. Therefore, to evaluate the accuracy of the inter-domain routing step of each scheme, the relative gap between the estimated and real costs is introduced. For WPF it is defined by:

$$\frac{cost^{e}_{\text{WPF}} - cost^{r}_{\text{WPF}}}{cost^{r}_{\text{WPF}}}$$

and similarly for JDP. Figure 4.3b illustrates the distribution of the relative gap for each routing scheme. We can observe that the gap is within $]-0.5, 0.5]$ for 89%, 82%, 81% and 81% of cases respectively for JDP $(\frac{1}{N},\frac{1}{N})$, JDP $(\frac{1}{N},\frac{1}{2N})$, JDP $(\frac{1}{N},\frac{1}{N^2})$ and WPF. This means the estimations of WPF and JDP are mostly close to their real costs. Moreover, the advantage of shorter working paths is illustrated since JDP $(\frac{1}{N},\frac{1}{N})$ gives better gaps than JDP $(\frac{1}{N},\frac{1}{2N})$, which in turn gives slightly better gaps than JDP $(\frac{1}{N},\frac{1}{N^2})$.

We compare JDP and WPF in frequency of finding smaller estimated and real costs. The comparisons are made with the three configurations of JDP. It should be noted that in this experiment α and β are integer and $\nu \sum_{e\in E} x_e + \mu \sum_{e\in E} y_e < 1$ for the three configurations of JDP. Therefore, the total bandwidth costs are minimized in these cases. In addition, when $(\nu, \mu) = (\frac{1}{N}, \frac{1}{N})$ the total length of the directive working and backup paths is minimized. When $(\nu, \mu) = (\frac{1}{N}, \frac{1}{2N})$ the directive working path π tends to be short. When $(\nu, \mu) = (\frac{1}{N}, \frac{1}{N^2})$ the shortest directive working path π and the shortest directive backup path π' among all candidates associated to π are obtained.

Figures 4.4a shows the percentage of cases for which JDP $(\frac{1}{N}, \frac{1}{N})$ or WPF finds better (smaller) total estimated costs when the number of sent requests increases. Figures 4.4b depicts the percentage of cases for which JDP $(\frac{1}{N}, \frac{1}{N})$ or WPF finds better total real costs when the number of sent requests increases. Figure 4.4c and

4.4d show the same results for JDP ($\frac{1}{N}$, $\frac{1}{2N}$) while Figure 4.4e and 4.4f illustrate the results for JDP ($\frac{1}{N}$, $\frac{1}{N^2}$). Note that WPF is overall slightly better than JDP ($\frac{1}{N}$, $\frac{1}{N}$) in estimated and real costs but JDP is more improved when the length of working path is more minimized. The best result is given with JDP ($\frac{1}{N}$, $\frac{1}{N^2}$). This confirms the expectation that when there are less virtual links the real cost is reduced. Furthermore, when the working path is short, there is more chance to share backup bandwidth with the future lightpath requests because there is less chance of violating the *sharing constraint* due to link-joint working paths. The overall resource utilization will be improved. In fact, WPF follows this strategy since it always looks for the shortest working path first. This explains why WPF obtains a relatively good performance even if it does not jointly compute the working and backup paths.

4.6.2 Blocking Probability Analysis

When the request holding time is infinite, the scheme with better resource allocation rejects less bandwidth and begins to reject later than the others. That is why we chose the bandwidth blocking probability as an index for evaluating the resource allocation capability. This probability is defined as the ratio between the amount of accepted bandwidth and the amount of requested bandwidth. Figure 4.5a shows the bandwidth blocking probability at the inter-domain step. We can see that the blocking of JDP is better than for WPF. The blocking of JDP ($\frac{1}{N}$,$\frac{1}{2N}$) (similar to that of JDP ($\frac{1}{N}$,$\frac{1}{N^2}$)) is better than the blocking of JDP ($\frac{1}{N}$,$\frac{1}{N}$). This can be explained by the fact that a shorter working path length increases the probability of finding a disjoint backup path. Although JDP is more advantageous at the inter-domain step, it finally has slightly more overall blocking due to intra-domain blocking (see Figure 4.5b). It seems that the inter-domain solutions found by WPF are somewhat more realist than those of JDP since the intra-domain blocking probability is smaller. Note that in both approaches, intra-domain blocking may result from the impossibility of finding an instance of a backup virtual link that is disjointed with the fixed working path when they cross the same domain. A joint

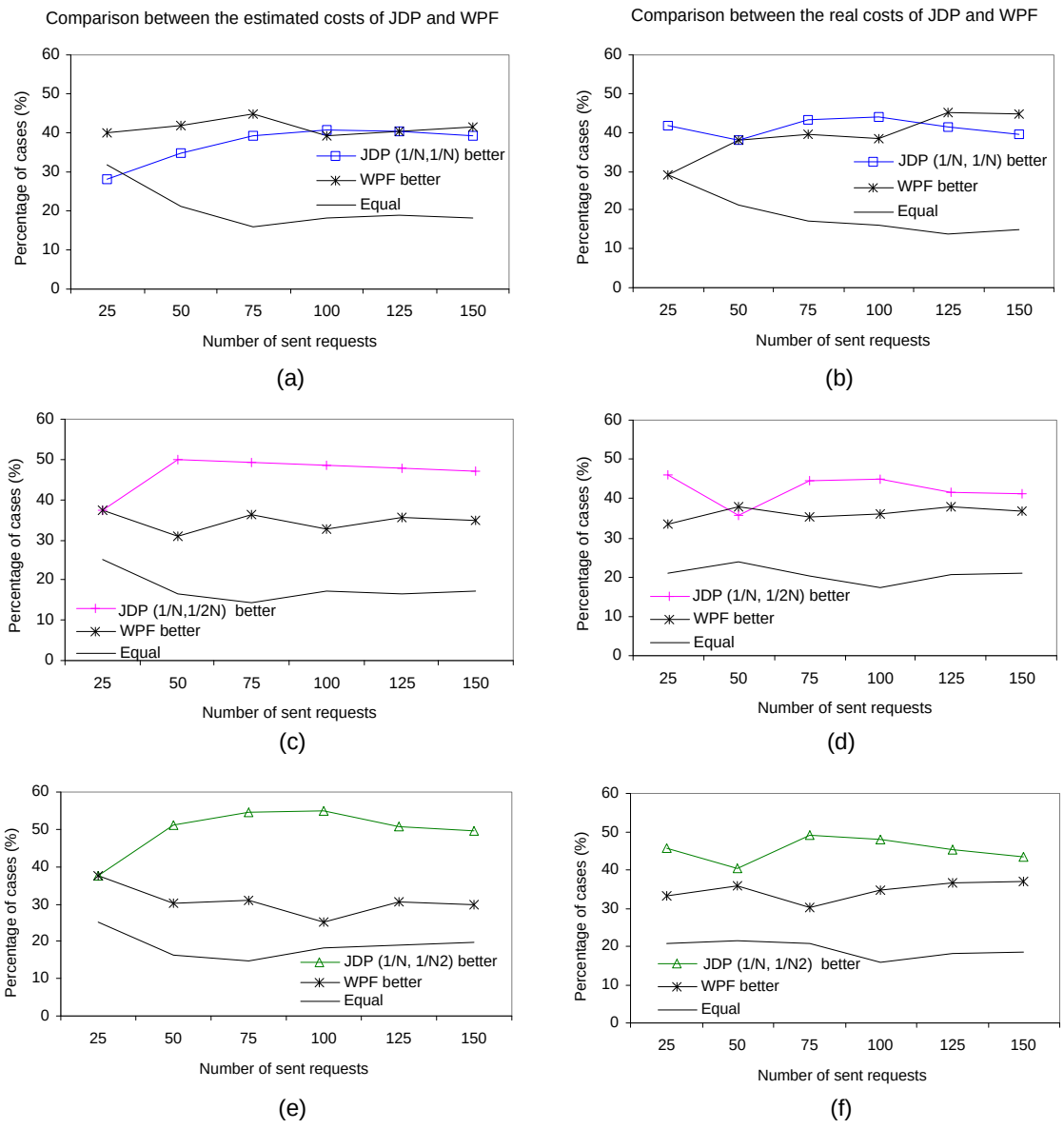

Figure 4.4: Advantages of JDP and WPF in estimated cost (a), (c), (e) and in real cost (b), (d), (f) when the number of sent requests increases.

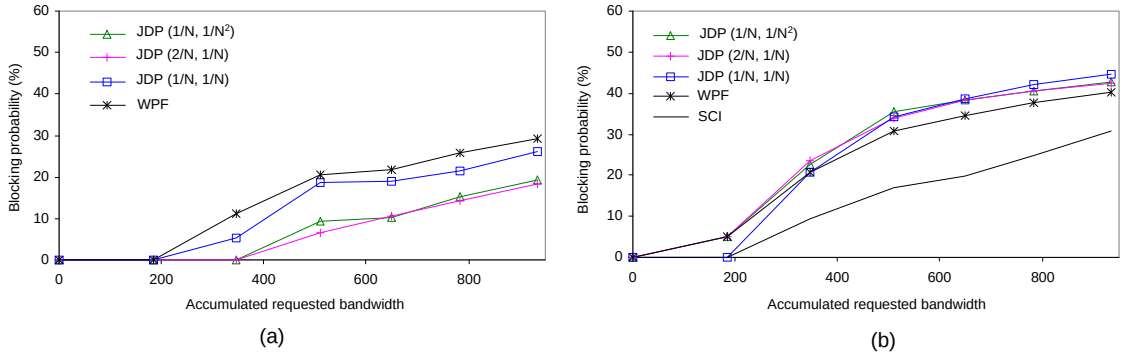

Figure 4.5: Bandwidth blocking probability at the inter-domain step (a) and Overall bandwidth blocking probability (b)

mapping of working and backup virtual links could reduce the blocking. Finally, note that SCI does not block drastically less than WPF and JDP. WPF never blocks 15% more than SCI and the difference tends to be reduced when the network is increasingly loaded.

4.6.3 Impact of update frequency on estimated cost and blocking probability

In the experiments so far, network link states are updated to border nodes immediately once a lightpath has been routed. Although in our solution the number of update messages is significantly reduced, this number can be further reduced by a delayed update. In other words, the updates are performed periodically at a short interval. However, the delayed update leaves the link-state information out of date leading to inaccurate routing. To analyze the impact of short update intervals on the cost and blocking probability, we conducted experiments with WPF. The experiment on JDP is unnecessary since WPF and JDP use the same information scenario and update method. We generated 500 requests according to Poisson process with rates of $\lambda_1 = 0.25$ (requests/second) and $\lambda_2 = 0.125$ (requests/second). The holding time is exponentially distributed with the mean $h = 160$ (seconds). The inter-domain blocking probability (Figure 4.6a) as well as the overall blocking

probability (Figure 4.6b) vary slightly when the update interval increases. The estimated cost, which is not shown here, is almost unchanged over different update intervals. We can conclude that short update intervals do not substantially decrease the routing quality though they make it more scalable.

On the other hand, the number of messages per update increases when the updates are more delayed (Figure 4.7). However, this number increases at a slower rate than the update interval, leading to a decrease in the total number of update messages when the update interval increases. Furthermore, the number of messages per update is almost constant in the range from 128 to 256.

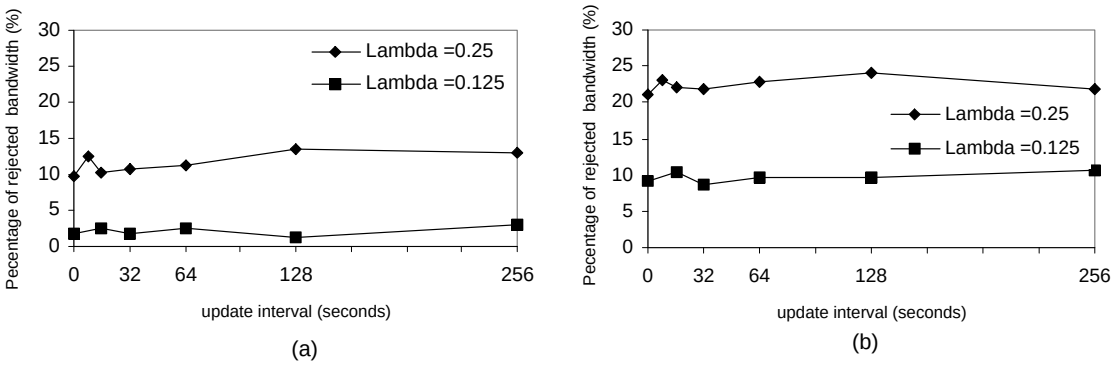

Figure 4.6: Bandwidth blocking probability of WPF at the inter-domain step (a) and Overall bandwidth blocking probability (b) under different update intervals.

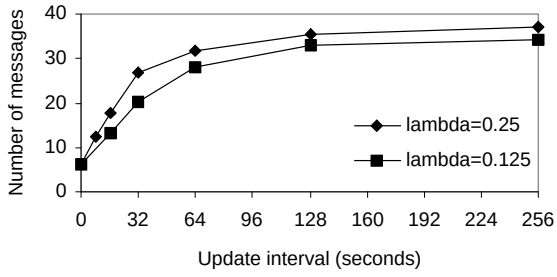

Figure 4.7: Number of update messages received by each border node under different update intervals.

4.7 Conclusion

Existing SPP solutions require global and detailed network information. As such information is not centrally available in multi-domain networks, these solutions are no longer applicable. In this paper, we propose an information aggregation scenario by underestimation and a two-step routing strategy for SPP in multi-domain networks. The main idea is to transform the original multi-domain problem to multiple single-domain problems using a topology aggregation combined with the proposed information scenario. Each single-domain problem is solved by using adapted versions of known single-domain SPP algorithms. The computational results show that our solution is not far from the ideal solution obtained using a complete information scenario. In other words, the proposed scheme is efficient and adequately respects the *scalability constraint* in the same time. Furthermore, we show that a short update interval does not significantly reduce the routing quality but makes the routing more scalable.

The proposed mathematical programming model with the coefficient $(\frac{1}{N}, \frac{1}{N^2})$ jointly computes the directive working and backup paths that minimize total resource costs. In addition, it finds the shortest directive working path among those minimizing the costs and the shortest directive backup path among those with the same directive working path length. The experiment results show that such a scheme leads to a smaller overall resource cost, followed by more efficient resource utilization thanks to a greater possibility of sharing backup bandwidth.

In order to reduce the blocking at the intra-domain step (and thus the overall blocking), especially when single-domain networks are slightly meshed, future works will concern the joint routing of working and backup paths when they cross the same domain.

Acknowledgements

We would like to thank Dr. Brigitte Jaumard from Concordia University (Canada) for providing us with access to ORC laboratory and for precious com-

ments.

Appendix A

Proposition: $\beta^{\pi}_{e'} \simeq \max_{e\in\pi} \beta^{e}_{e'}$

Proof:

By combination the definitions of $b^{q}_{q'}$, $B^{q}_{\max}$, $b^{q}_{\ell'}$ and the approximation of b^{ℓ}_{ℓ} by (4.8), we have:

$$b^{q}_{q'} \simeq \sum_{\ell'\in q'} \min\{\max\{0, B^{q}_{\max} + d - B_{\ell'}\}, d\}.$$

Similarly, $b^{p}_{q'} \simeq \sum_{\ell'\in q'} \min\{\max\{0, B^{p}_{\max} + d - B_{\ell'}\}, d\}$.

We use $q \subset p$ to denote that q is subset of p. It is clear that $B^{p}_{\max} = \max_{q\subset p} B^{q}_{\max}$, then $b^{p}_{q'} \simeq \max_{q\subset p} b^{q}_{q'}$.

Combining with the definition of $\beta^{\pi}_{e'}$ and $\beta^{e}_{e'}$ we conclude that $\beta^{\pi}_{e'} \simeq \max_{e\in\pi} \beta^{e}_{e'}$.

CHAPITRE 5

BACKUP PATH RE-OPTIMIZATIONS FOR SHARED PATH PROTECTION IN MULTI-DOMAIN NETWORKS

B. Jaumard and D. L. Truong

Abstract: Within the context of dynamic routing models for shared path protection in multi-domain networks, we propose a backup path re-optimization phase with possible rerouting of the existing backup paths in order to increase the bandwidth sharing among them while minimizing the network backup cost. The re-optimization phase is activated periodically or when routing a new connection fails because of insufficient capacity. Three re-optimization models are discussed : i) Global rerouting where the re-optimization is performed once for the entire network ; ii) Local rerouting where the re-optimization is serially performed on one domain at a time or on selected domains, and iii) Local rerouting with least effort, i.e., where the smallest possible number of backup path reroutings is performed in order to be able to handle new connection requests. The first model offers the best resource savings while the two others are more scalable in multi-domain networks. Comparative performance of the three models are conducted and numerical results are presented.

Status: Cet article a été présenté et publié à la conférence *IEEE/Globecom 2006*, San Francisco, USA, 27 novembre - 1 decembre 2006.

5.1 Introduction

Shared Protection [Ram99] has been widely studied in the literature. It allows bandwidth sharing amongst backup paths leading to some bandwidth savings while continuing to guarantee 100% failure recovery. Within the single-failure context, 100% failure recovery condition is expressed with the condition that the working paths of the backup paths that share bandwidth must be disjoint. Routing for shared protection aims to identify the working and backup paths that minimize the total bandwidth consumption. We consider the problem for the networks with bandwidth guaranteed connections such as MPLS-TE/ATM or optical networks. The later should be equipped with Multi-service Provisioning Platforms (MSPP) [Muk06] with bandwidth grooming and wavelength conversion capacity at every node. The wavelength assignment problem and wavelength continuity constraint are thus relaxed. Existing solutions follow two paradigms: static routing (off-line) and dynamic routing (on-line). In static routing, the network traffic, i.e. requests for connections, are assumed to be stable and are given as input to the routing model. The working and backup capacities are then optimized for every network link, see, e.g., in [GS98, HC02, XM99, JO01]. Conversely, dynamic routing is proposed for dynamically changed traffic and requests for connections are routed one at a time without taking into account any information on the future requests, see, e.g., [KL03, TC03, XQX02]. As time goes, the total allocated bandwidth will be larger (less optimized) than as if a routing policy with a global view on the forthcoming connections had been applied.

It is known and has been already studied in [SR01, Lab04] that, if we use dynamic routing but reorganize the existing paths in the network, working bandwidth could be freed and increased bandwidth sharing for the backup bandwidth can be obtained leading to a greater resource saving. The reorganization includes finding alternate paths for the existing working and backup paths and then rerouting some working and backup paths. Moving the traffic of a connection on a new working path implies service interruption, and therefore a disorder for the user, that is to

be avoided as much as possible. However, backup paths are generally inactive until a failure occurs. They can be replaced by new ones without any impact on service availability. Therefore, a reorganization scheme in which only backup paths are rerouted offer a good mean to answer to the drawback of the possible bandwidth waste involved in dynamic routing due to the uncertainty about estimating and anticipating the future connection requests.

Few publications exist on rerouting algorithms in the context of dynamic routing. One exception is [Lab04], but no detailed algorithm is provided. We propose here solutions for rerouting backup paths where the objective is to seize the backup capacity. The solutions differ from the backup path reroute solutions, see, e.g., [AHSG05], which aim to improve the service availability at dual failures.

A multi-domain network (see Fig.5.1) is composed of multiple single-domain networks interconnected amongst them by inter-domain links going from border nodes of some domains to border nodes of another. Multi-domain networks are characterized by the *scalability constraint*, defined in [TT06], that no global information is available centrally and limited routing information is exchanged in small scope [BSO02]. In a previous paper [TT06], we have proposed two dynamic routing models, called WPF (Working Path First) and JDP (Joint Directive Path), for Shared Path Protection for multi-domain networks. In these models, a request is routed one at a time with the objective of minimizing its total requested working and backup bandwidth. Although the models satisfy the scalability issue, it suffers from some drawbacks regarding unnecessary bandwidth waste for the backup paths. In this paper, we propose a rerouting model to enhance WPF and JDP. An additional phase will be trigger after WPF and JDP which reroutes existing backup paths in order to re-optimize the network backup capacity. It is called alternatively re-optimization or rerouting phase; it requires extra computation effort and network information exchange while tearing down old paths and setting up new paths. For reducing this effort and information exchange, the rerouting phase should not be activated regularly but once after a given period of time.

In the next Section, we present the backup path rerouting problem. We propose

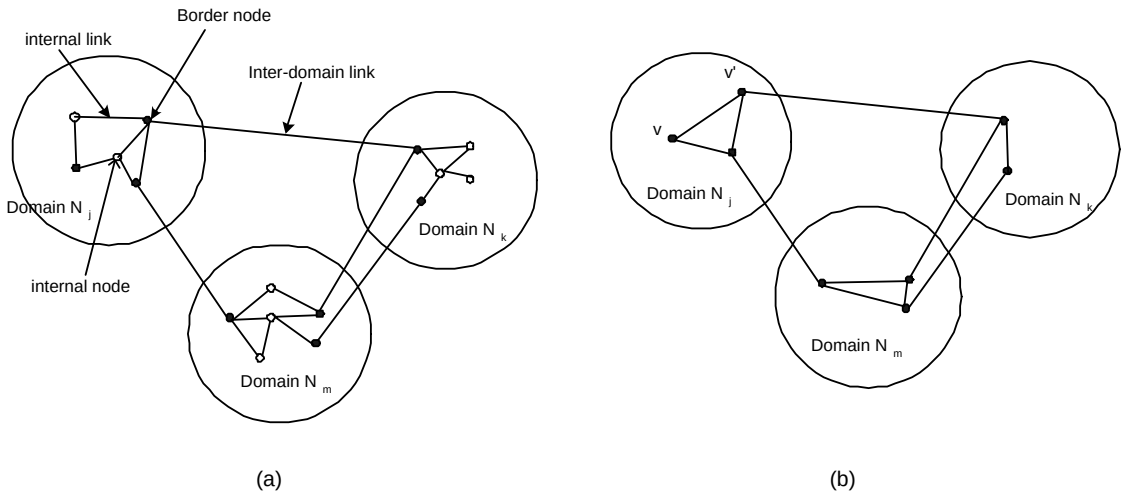

Figure 5.1: Illustration of a Multi-domain network

in Section 5.3.2 the *Global reroute* model, in which the end-to-end backup paths in the network are rerouted at once. Due to the global information requirements, the model is only suitable for single domain network. We next propose a *Local reroute* model to be used for multi-domain network in Section 5.3.3. There, each domain subsequently reroutes the segments of backup paths within it. In Section 5.3.4, the *Least local reroute* model where, in each domain, only a minimal number of backup segments will be rerouted. The integration of WPF and JDP with these rerouting models will be compared with original WPF and JDP without rerouting. Numerical results are described in Section 5.4. The trade off of reroute phase in terms of computational effort and information exchanges is also discussed. Conclusions are drawn in the last section.

5.2 The backup path rerouting problem

Let us consider a multi-domain network with a set K connection requests that are already routed in the network, i.e., a working and a backup path (denoted by p_k and p'_k) has been already defined for them. Request k asks for a connection from source s_k to destination d_k for bandwidth b_k. The backup path rerouting problem is stated as follows. Let $\mathcal{R}^B \subset K$ is the index set of the requests whose backup

paths might be rerouted. While all working paths should remain unchanged, we look for the set of alternative paths of the current backup paths whose indexes are in $\mathcal{R}^B$, that minimizes the overall bandwidth required for the backup paths. If changed, the backup path of the request $k \in \mathcal{R}^B$ must remain disjoint from the working path p_k so that it will not fail when p_k fails upon a single failure. The fewer backup paths are rerouted, the more scalable and practical the solution is, but may be the less bandwidth saving will be obtained. When all backup paths are allowed to be rerouted $\mathcal{R}^B = K$, the best bandwidth saving will be attained.

5.3 Mathematical models

5.3.1 Notations

Let us represent the multi-domain network by a directed graph $G = (E, V)$ where V is the set of nodes and E is the set of fiber links. The reversed fiber link of the link $e \in E$ is denoted by $\bar{e} \in E$. Each network link joins two nodes and is assumed to be bi-directional with two fibers, each carrying the traffic in one direction. The two fibers are assumed to be fold together in the same conduct so they share the same risk upon a single failure. Each fiber is represented by an arc and a network link is represented by a pair $(e, \bar{e})$ of fiber links. We denote by c_e the bandwidth capacity that is available on the fiber link e.

Arc $e \in E$ is associated with binary parameters δ^W_{ek} and δ^B_{ek} such that:

$$\delta^W_{ek} = \begin{cases} 1 & \text{if } e \in p_k \\ 0 & \text{otherwise} \end{cases} \qquad e \in E, k \in K, \tag{5.1}$$

$$\delta^B_{ek} = \begin{cases} 1 & \text{if } e \in p'_k \\ 0 & \text{otherwise} \end{cases} \qquad e \in E, k \in K \setminus \mathcal{R}^B. \tag{5.2}$$

For a given node $v \in V$, we denote by $\Gamma^+(v)$ its set of outgoing edges, and by $\Gamma^-(v)$ its set of incoming edges.

5.3.2 Global reroute

5.3.2.1 Variables

We introduce two sets of variables. The first set, $B_e, e \in E$, defines for each B_e, the bandwidth required for backup paths going through a given arc $e \in E$. We next define variables y_e^k that are decision variables such that:

$$y_e^k = \begin{cases} 1 & \text{if } e \text{ belongs to the backup path of } k \\ & \text{after the rerouting phase} \\ 0 & \text{otherwise.} \end{cases}$$

5.3.2.2 Objective function

In the *Global reroute* model, the objective is to minimize the bandwidth required for all backup paths. The bandwidth required for working paths remain unchanged as no alternative path is sought for them. The objective can then be written:

$$\min \sum_{e \in E} B_e. \tag{5.3}$$

5.3.2.3 Constraints

$$\sum_{e \in \Gamma^+(v)} y_e^k - \sum_{e \in \Gamma^-(v)} y_e^k = \begin{cases} 1 & \text{if } v = s_k \\ 0 & \text{if } v \neq s_k, d_k \\ -1 & \text{if } v = d_k \end{cases}$$

$$v \in V, k \in \mathcal{R}^B \tag{5.4}$$

$$\sum_{k\in\mathcal{R}^B} b_k(\delta^W_{ek}+\delta^W_{\overline{e}k})y^k_{e'} \le B_{e'} - \sum_{k\in K\setminus\mathcal{R}^B} b_k(\delta^W_{ek}+\delta^W_{\overline{e}k})\delta^B_{e'k}$$

$$e, e' \in E \quad (5.5)$$

$$\delta^W_{ek}+\delta^W_{\overline{e}k}+y^k_e+y^k_{\bar{e}} \le 1 \qquad e\in E, k\in\mathcal{R}^B \quad (5.6)$$

$$\sum_{k\in K} b_k\delta^W_{ek} + B_e \le c_e \qquad e\in E. \quad (5.7)$$

Variable domains:

$$y^k_e \in \{0,1\} \qquad e\in E, k\in\mathcal{R}^B \quad (5.8)$$

$$B_e \ge 0 \qquad e\in E. \quad (5.9)$$

Constraints (5.4) are the flow conservation ones for the rerouted backup paths. Constraint (5.5) ensures that the backup bandwidth $B_{e'}$ on e' will never be smaller than the bandwidth needed to protect every working path against a single failure on the fiber pipe containing the pair $(e, \bar{e})$ of fiber links. This latter backup bandwidth is indeed the bandwidth of the working paths over e or $\bar{e}$ that are protected by the backup paths going through e'. Constraint (5.6) assures that p_k and p'_k are always link disjoint. Constraint (5.7) guarantees that the bandwidth used by both working and backup paths over a link will not exceed the link capacity. If there is a loop in p'_k the loop will be removed a posteriori.

This model provides optimal rerouting but is not scalable for multi-domain networks due to global information requirements in model building. For gathering the data of constraints (5.5), a central node needs to keep the routes of all the working paths in the network. It also needs the complete knowledge of the network topology and bandwidth allocation on the fiber links.

5.3.3 Local reroute

In order to overcome the drawback of the *Global reroute* model, we next propose the *Local reroute* one. Instead of rerouting the end-to-end backup paths as in the *Global reroute* model, each domain reroutes locally their inner backup segments in order to minimize its backup capacity. For each segment, the ingress and egress border nodes remain unchanged. The alternate backup paths still go through the same border nodes and inter-domain links. The model that computes the alternate backup segments for domain $D = \{E^{\mathrm{D}}, V^{\mathrm{D}}\}$ is called RRLocal(D). Let $s_k^{\mathrm{D}}, t_k^{\mathrm{D}}$ be respectively the ingress and the egress border nodes of p'_k in the domain D. The model is however similar to the *Global reroute* model in respect to parameter initializations and variable domains.

5.3.3.1 Objective function

We minimize the backup bandwidth consumed by domain D:

$$\text{Minimize} \sum_{e \in E^{\mathrm{D}}} B_e \tag{5.10}$$

5.3.3.2 Constraints

$$\sum_{e \in \Gamma^+(v)} y_e^k - \sum_{e \in \Gamma^-(v)} y_e^k = \begin{cases} 1 & \text{if } v = s_k^{\mathrm{D}} \\ 0 & \text{if } v \neq s_k^{\mathrm{D}}, d_k^{\mathrm{D}} \\ -1 & \text{if } v = d_k^{\mathrm{D}} \end{cases}$$

$$v \in V^{\mathrm{D}}, k \in \mathcal{R}^B \tag{5.11}$$

$$\sum_{k\in\mathcal{R}^B} b_k(\delta^W_{ek}+\delta^W_{\overline{e}k})y^k_{e'} \le B_{e'} - \sum_{k\in K\setminus\mathcal{R}^B} b_k(\delta^W_{ek}+\delta^W_{\overline{e}k})\delta^B_{e'k}$$

$$e\in E, e'\in E^{\mathrm{D}} \quad (5.12)$$

$$\delta^W_{ek}+\delta^W_{\overline{e}k}+y^k_e+y^k_{\overline{e}} \le 1 \qquad e\in E^{\mathrm{D}}, k\in\mathcal{R}^B \quad (5.13)$$

$$\sum_{k\in K} b_i\delta^W_{ek} + B_e \le c_e \qquad e\in E^{\mathrm{D}}. \quad (5.14)$$

Constraints (5.11)-(5.14) are similar to constraints (5.4)-(5.7) of the *Global reroute* model except that they are applied only to the backup segments in domain D. The whole reroute process over multi-domain network follows the pseudo-code:

For all D in G RRLocal(D)

The *Local reroute* model requires smaller scope of information than *Global reroute* model. Except for the constraint (5.12) that requires each border node of D to keep the routes of working paths protected by an arc of D, other constraints are built with the information within the domain D. The solution is much more scalable and the resulting mathematical model, being smaller, is much easier to solve.

5.3.4 Least local reroute model

The *Least local reroute* model (LeastRRLocal) is a further development of the *Local reroute* model with $\mathcal{R}^B = K$. All backup paths are allowed to be rerouted with a rerouting preference level. The rerouting preference level of the backup path p_k is defined by $w_k \in [0,1]$. The smaller w_k is, the less preference is given to the rerouting p'_k. The model looks for a rerouting configuration with minimal backup capacity for the primary objective and the least weighted number of rerouted backup paths for the secondary objective.

5.3.4.1 Variables

A decision variable r_k is associated with each request k indicating if p'_k will be rerouted ($r_k = 1$) or remain unchanged inside the domain D ($r_k = 0$).

5.3.4.2 Objective function

A second term counting the weighted number of rerouted backup segments is added to the objective function with coefficient M_1 sufficiently large as to make the second term smaller than 1. Since the first term is integer, the second term selects the solution with the least weighted number of reroutings when breaking ties is needed.

$$\text{Minimize} \sum_{e \in E^{\mathrm{D}}} B_e + \frac{1}{M_1} \sum_{k \in P} w_k r_k \tag{5.15}$$

5.3.4.3 Constraints

Let M_2 be a large constant.

$$\sum_{e \in \Gamma^+(v)} y_e^k - \sum_{e \in \Gamma^-(v)} y_e^k + M_2(1 - r_k) \geq \begin{cases} 1, & \text{if } v = s_k^{\mathrm{D}} \\ 0, & \text{if } v \neq s_k^{\mathrm{D}}, d_k^{\mathrm{D}} \\ -1, & \text{if } v = d_k^{\mathrm{D}} \end{cases}$$

$$v \in V^{\mathrm{D}}, k \in K. \tag{5.16}$$

$$\sum_{e \in \Gamma^+(v)} y_e^k - \sum_{e \in \Gamma^-(v)} y_e^k - M_2(1 - r_k) \leq \begin{cases} 1, & \text{if } v = s_k^{\mathrm{D}} \\ 0, & \text{if } v \neq s_k^{\mathrm{D}}, d_k^{\mathrm{D}} \\ -1, & \text{if } v = d_k^{\mathrm{D}} \end{cases}$$

$$v \in V^{\mathrm{D}}, k \in K. \tag{5.17}$$

$$y_e^k + M_2 r_k \geq \begin{cases} 1, & \text{if } e \in p'_k \\ 0, & \text{otherwise} \end{cases} \qquad e \in E^{\mathrm{D}}, k \in K \tag{5.18}$$

$$y_e^k - M_2 r_k \leq \begin{cases} 1, & \text{if } e \in p'_k \\ 0, & \text{otherwise} \end{cases} \qquad e \in E^{\mathrm{D}}, k \in K \tag{5.19}$$

$$\delta_{ek}^W + \delta_{\bar{e}k}^W + y_e^k + y_{\bar{e}}^k \leq 1, \qquad e \in E^{\mathrm{D}}, k \in K \tag{5.20}$$

$$\sum_{k \in K} b_k (\delta_{ek}^W + \delta_{\bar{e}k}^W) y_{e'}^k \leq B_{e'}, \qquad e \in E, e' \in E^{\mathrm{D}} \tag{5.21}$$

$$\sum_{k \in K} b_k \delta_{ek}^W + B_e \leq c_e, \qquad e \in E^{\mathrm{D}}. \tag{5.22}$$

Variable domains:

$$r_k \in \{0, 1\}, \qquad k \in K \tag{5.23}$$

$$y_e^k \in \{0, 1\}, \qquad e \in E, k \in K \tag{5.24}$$

$$B_e \geq 0, \qquad e \in E. \tag{5.25}$$

If a path p'_k is rerouted, the flow conservation constraints (5.11) must be enforced, otherwise the parameter initializations (5.2) must hold. Since the set of backup paths to be rerouted is still unidentified, the flow conservation constraints and parameter initializations are built in such a way that only one of them is applied for a given backup path. Inequalities (5.17) and (5.16) are flow conservation constraints for the rerouted backup paths while (5.18) and (5.19) initialize δ_{ek}^B for the unchanged backup paths. Constant M_2 enables only one of the two groups and makes the other one redundant. Indeed, M_2 is sufficiently large if it is greater than the highest incoming and outgoing degrees of a node, thus $M_2 > \max\{\max_{v,k} \sum_{e \in \Gamma^-(v)} y_e^k, \max_{v,k} \sum_{e \in \Gamma^+(v)} y_e^k\}$. The remaining constraints are similar to those of the *Local reroute* model.

When $w_k = 1, k \in K$, the same backup capacity as in the *Local reroute* model

is obtained but only the minimum number of backup segments is rerouted in order to minimize the requested backup bandwidth. Less backup segments must be torn down and reserved and less information need to be exchanged in domains. For further reducing the information exchanges, the rerouting should be activated periodically after a time period, or when we reach a blocking with the routing using WPF or JDP (RRLocal-Block). In the latter case, we reroute only in the blocked domain expecting that some bandwidth could be released, then retry WPF or JDP again.

5.4 Experiment results

The proposed rerouting will be evaluated on their backup bandwidth saving, blocking probability reducing and scalability. The experiment is performed on WPF (because JDP itself provides similar result as WPF) with the following schemes:

- Without reroute: WPF-noRR.
- With *Least local reroute*, when $w_k = 1, k \in K$, after 50 or 100 requests, i.e. WPF-LeastRRLocal-50, WPF-LeastRRLocal-100.
- With *Least local reroute* uniquely in blocked domain upon blocking: WPF-RRLocal-Block.

No experiment was conducted with the *Global reroute* model due to its high computational effort and its lack of scalability for multi-domain networks. Experiment on RRLocal will not be shown neither because when $\mathcal{R}^B = K$, the results are similar to those of LeastRRLocal whereas in the later the minimum number of backup paths are modified. The experiment with $\mathcal{R}^B \neq K$ is left for the future due to the limited space for this paper. The multi-domain network instance is composed of 5 real optical single domain networks: EONnet [OSYZ95], RedIRIS [RED05], GARR [GAR05], Renater3 [REN05], SURFnet [SUR05] with link capacities varying from OC-3 to OC-192 (see Fig.5.2). Some inter-domain links of OC-192 have been added. For each experiment, a sequence of 1000 requests are sent. These requests

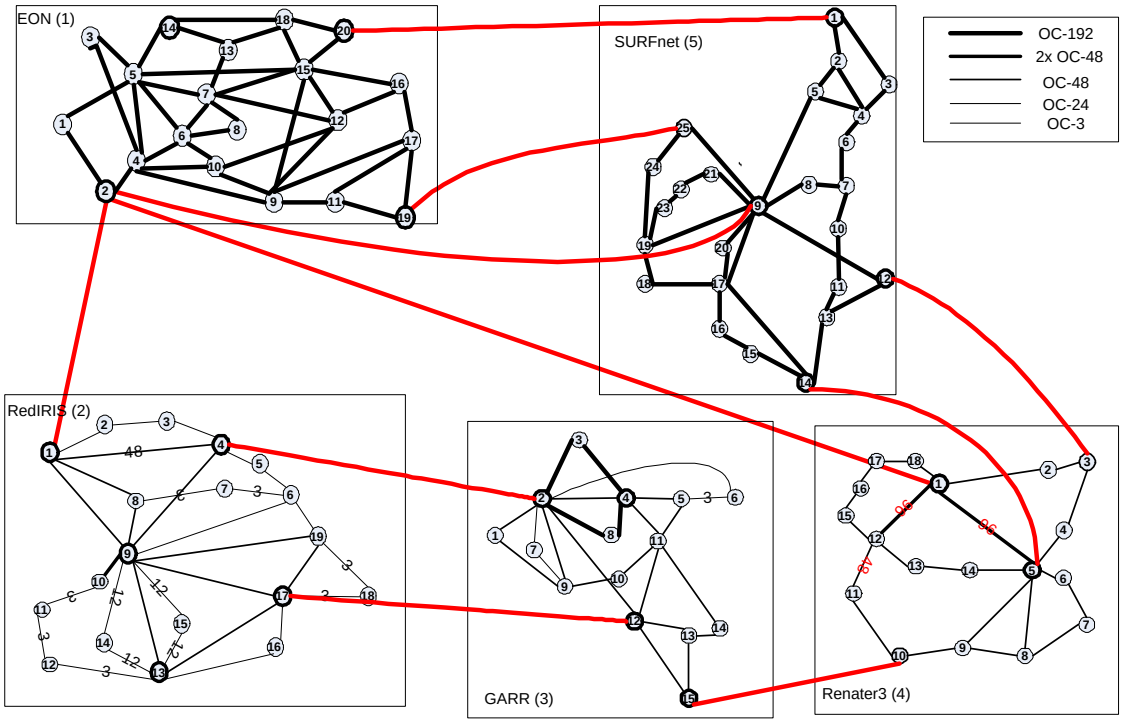

Figure 5.2: Experimental multidomain network.

are between randomly selected border nodes with requested bandwidth uniformly distributed in OC-{1, 3, 6, 9, 12}. Requests arrive according to Poisson process with the rate $\lambda = 0.25$ (requests/s). The request holding time is exponentially distributed with mean $h = 320(s)$. The experiment result will be shown after the 300^{th} request when the network load is stable with an average of 80 simultaneous active connections. This load is sufficient to produce blocking in the network.

CPLEX is used to solve the two rerouting models and Opnet Modeler is used for implementing WPF and simulating the network environment. It takes less than 20 seconds for a rerouting by LeastRRLocal on a Pentium IV-3Ghz. (It takes however about 6 days to solve the *Global reroute* model).

For later convenience, we consider the whole process of rerouting as a single one in RRLocal-Block or LeastRRLocal. It includes multiple simultaneous backup segment reroutings within one or different domains.

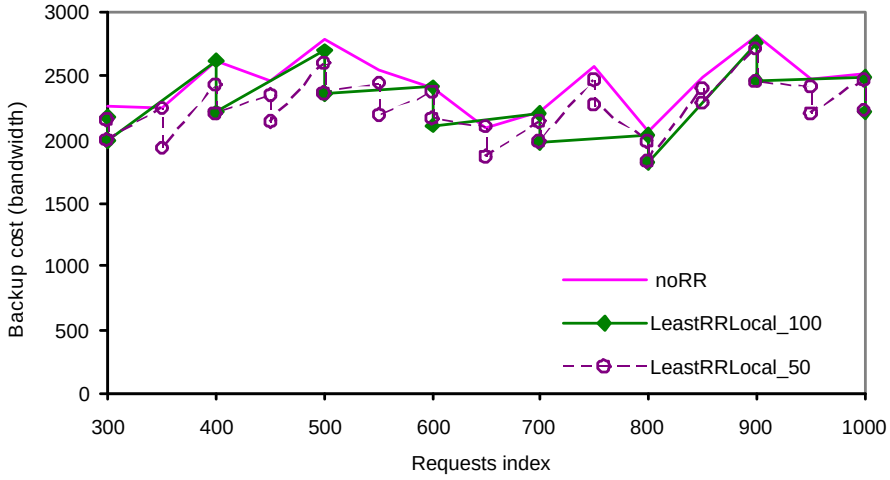

Figure 5.3: Backup costs of WPF in different rerouting schemes.

5.4.1 Backup bandwidth saving

The ability of saving backup bandwidth in RRLocal and LeastRRLocal will be highlighted by comparing the backup capacity (backup cost) obtained in using these schemes with that of WPF-noRR. Here, link capacities are uncapacitated for getting rid of the influence of the blocking cases. Fig.5.3 shows backup costs of WPF-LeastRRLocal-50, WPF-LeastRRLocal-100 and WPF-noRR. The backup cost of the first two schemes reduces at each rerouting illustrating the released bandwidth thanks to backup path reroutings. We define the relative backup cost gain as the fraction between released bandwidth and the network backup capacity before rerouting. Fig.5.4 depicts the gains of each rerouting scheme: for WPF-LeastRRLocal-100 it is an average of 11.5% and for WPF-LeastRRLocal-50 it is an average saving of 9.8%. Less backup bandwidth is released by WPF-LeastRRLocal-50 at each rerouting because the backup paths has been re-organized not so long before. However, WPF-LeastRRLocal-50 frees more frequently backup bandwidth than WPF-LeastRRLocal-100, after each 50 requests against 100 requests; and thus leaves more room for the requests arriving between two reroutings resulting in less blocking as we will see in the next section.

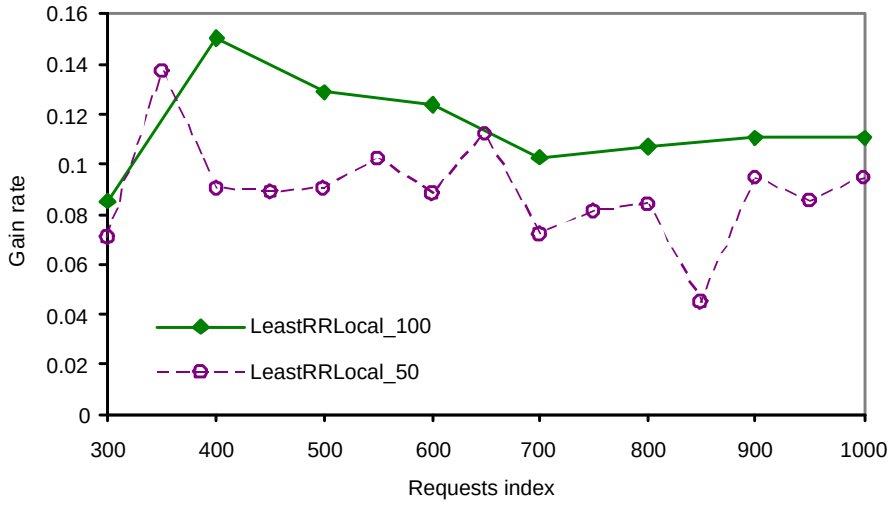

Figure 5.4: Relative backup cost gains of WPF in different rerouting schemes.

5.4.2 Blocking probability

For evaluating the impact of rerouting on blocking probability, capacities are set back on fiber links. Fig.5.5, shows the blocking probabilities of WPF in different rerouting schemes. WPF-RRLocal-Block is the best scheme in terms of blocking probability. It reduces the blocking of WPF-noRR about 3%, note that the original blocking is between 8%-10%. WPF-LeastRRLocal-50 and WPF-LeastRRLocal-100 follow up with more modest results. This is explained by the blocking driven nature of RRLocal-Block. In RRLocal-Block, when a request is blocked, a local rerouting is activated at the domain where the blocking occurs, after that the blocked request is routed again. The rerouting has thus an immediate deblocking impact. It is easy to see in Fig.5.6 the deblocking capacity of WPF-RRLocal-Block through the distance between two blocking probabilities before and after rerouting. Although WPF-LeastRRLocal seizes backup bandwidth regularly, blocking may still occur at a later stage after a rerouting because of non-optimized bandwidth allocation for the subsequent requests, which are not re-organized until the next rerouting. That is why WPF-LeastRRLocal-50 and WPF-LeastRRLocal-100 have a higher blocking probability than RRLocal-Block.

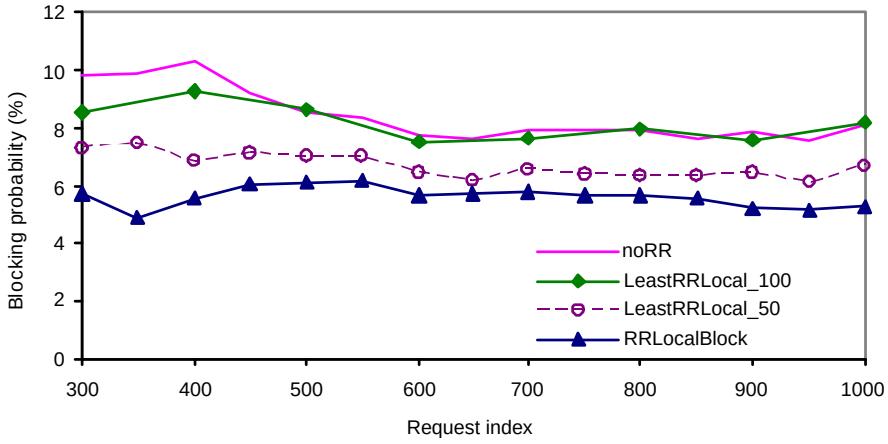

Figure 5.5: Blocking probability of WPF in different rerouting schemes.

Conforming with the expectation in the previous experiment, WPF-LeastRRLocal-50 blocks 0.5% less than WPF-LeastRRLocal-100 because it re-organizes more frequently backup capacity thus leaving more free capacity for the new coming requests.

5.4.3 Scalability evaluation

The scalability of LeastRRLocal and RRLocal-Block over noRR will be first of all evaluated based on the scope of the exchange of the information they require. Let us begin with the computation of rerouted paths. As discussed at the end of Section 5.3.3, for a domain D, RRLocal, therefore LeastRRLocal and RRLocal-Block, requires that border nodes of D keeps the routes of all working paths protected by a link of D. This requirement could be easily satisfied by benefiting from the backup path reservation process of WPF, which forwards the route of a working path along its backup path (see [TT06] for details). Therefore, WPF-LeastRRLocal and WPF-RRLocal-Block do not require any extra information exchange in comparison with WPF-noRR, although a larger information storage is required.

The rerouted path computation is followed by the signaling process which is composed of i) tearing down the old backup segments, ii) reservation of the new backup segments within domains. In LeastRRLocal the signaling is needed in all

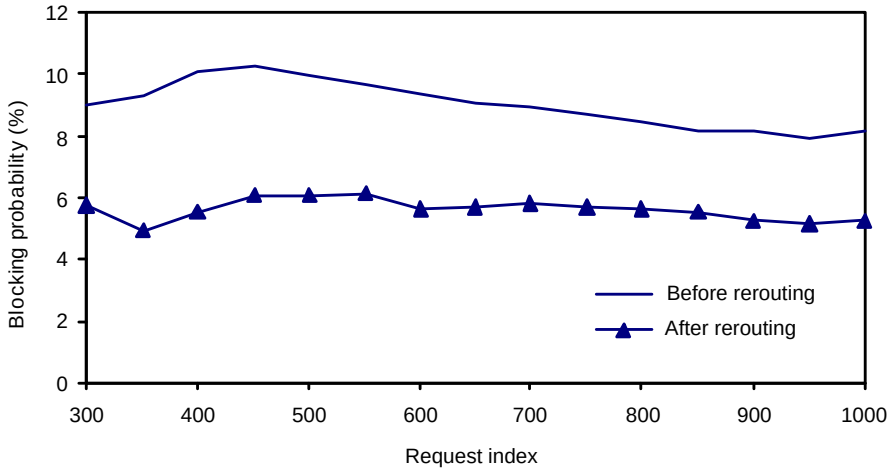

Figure 5.6: Blocking probability of WPF-RRLocal-Block before and after rerouting.

Table 5.1: Information exchange scopes

	LeastRRLocal	RRLocal-Block
Extra information exchange over WPF in path computation	no	no
Signaling scope	All domains	Blocked domain

domains while RRLocal-Block requires it only in one domain because backup paths are rerouted respectively in all domains and one domain. Table 5.1 summarizes the qualitatively comparison on information exchange in the two methods.

Another important factor of the scalability is the number of rerouted backup segments generated in each rerouting method. It is recommended to keep it small as the quantity of information to be exchanged and the number of operations to be performed on network nodes during the signaling process increases proportionally with the number of rerouted backup segments. Table 5.2 presents the average number of rerouted backup segments in each domain, in all domains per rerouting and the total number of rerouted segments after all reroutings in the cases of LeastRRLocal-50, LeastRRLocal-100 and RRLocal-Block. Over the entire network, LeastRRLocal-50 changes nearly as many backup segments per rerouting as LeastRRLocal-100: 43.6 versus 50.4 segments/rerouting. RRLocal-Block reroutes considerably fewer backup segments per rerouting: 8.92 segments, because it only

Table 5.2: Number of rerouted backup segments

Domain (seg./rerouting)	LeastRRLocal-{f}		RRLocal-Block
	f= 50	f=100	
EON	10.75	14.2	-
RedIRIS	13.45	16.5	-
GARR	4.65	5.6	-
Renater3	9.6	8.3	-
SUFnet	5.15	5.8	-
All domains	43.6	50.4	8.92
All domains, reroutings (seg.)	872	504	375

reroutes the backup segments within blocked domains.

Although high rerouting frequency allows a larger reduction of the blocking, it increases the overall number of rerouted backup segments. Globally, LeastRRLocal-50 involves nearly twice the number of backup segments in rerouting than LeastRRLocal-100: 872 versus 504 rerouted segments. On the other hand, RRLocal-Block reroutes only 375 segments. Note that in this experiment, RRLocal-Block re-organizes quite often backup segments, about 8 times per 100 requests, because of blocking due to high network load.

From the above analysis, we can conclude that RRLocal-Block is more scalable than LeastRRLocal-100, which is in it turn more scalable than LeastRRLocal-50.

5.5 Conclusion

This paper presents different backup path rerouting schemes for multi-domain networks. The experiment results demonstrate that these rerouting schemes led to an economy of up to 11.5 % backup capacity and the dropping off of until 3% blocking in comparison to the original blocking of 8%-10%.

A regular (time-driven) rerouting helps to regularly free some capacity and thus reduce the blocking probability. However, it implies extra computational effort and information exchange in rerouted path computation and signaling. The choice of the rerouting frequency is a compromise between the scalability and the blocking

probability. In comparison with LeastRRLocal in different rerouting frequencies, RRLocal-Block (that is blocking-driven) provides smaller blocking probability, requires less information exchanges and less computational effort. We suggest thus RRLocal-Block as an efficient and scalable solution for multi-domain networks.

CHAPITRE 6

USING TOPOLOGY AGGREGATION FOR EFFICIENT SHARED SEGMENT PROTECTION SOLUTIONS IN MULTI-DOMAIN NETWORKS

Dieu-Linh Truong, and Brigitte Jaumard

Abstract: The dynamic routing problem for *Overlapping Segment Shared Protection* (OSSP) in multi-domain networks has not received a lot of interest so far as it is more complex than in single-domain networks. Difficulties lie in the lack of complete and global knowledge about network topologies and bandwidth allocation whereas this knowledge is easily available in single-domain networks. We propose a two-step routing approach for the OSSP based on a topology aggregation scheme and link cost estimation : an inter-domain step and an intra-domain step. We propose two different heuristics, GROS and DYPOS for the inter-domain step, and a "Blocking-go-back" strategy in order to reduce the blocking rate in the intra-domain step. We compare the performance of the two heuristics against an optimal single-domain approach. We show that both heuristics lead to resource efficient solutions that are not far from the optimal ones. Moreover, both heuristics require relatively small computational efforts and are scalable for multi-domain networks.

Keywords: Multi-domain Network, Protection, Routing.

Status: Cet article a été accepté conditionnel à des corrections mineures au *Journal on Selected Areas in Communications/Optical Communications and Networking series* en juin 2007. Ses résultats préliminaires avaient été publiés partiellement dans l'article "Overlapping Segment Shared Protection in Multi-domain Optical Networks", *IEEE/ Asia-Pacific Optical Communication*, Korea, 3-7 Sept. 2006.

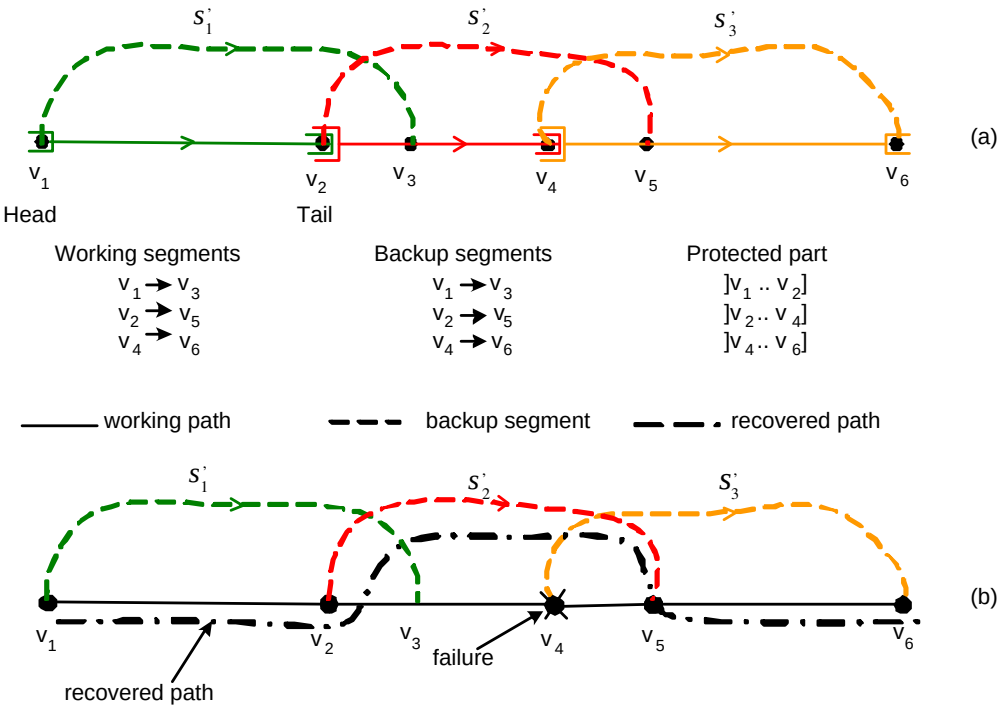

Figure 6.1: Example of Overlapping Segment Protection when v_4 fails. The protected part $]v_2..v_4]$ contains all links and nodes between v_2 exclusively and v_4 inclusively, thus v_4 is recovered by segment s'_2.

6.1 Introduction

In segment protection, an end-to-end working path is divided into segments, each of which are protected by a unique backup segment. Only one backup segment is activated upon a single link or node failure, the other working segments, which are not impaired by the failure, remain used. As a result, segment protection offers a faster recovery than path protection. In the classical segment protection, working segments are non-overlapping. Segment end nodes are then not protected because the failures of those nodes impair both working and backup segments. Overlapping Segment Protection, firstly proposed in [RKM02] and [HM02], overcomes this weakness thanks to the overlapping between working segments (see Fig. 6.1) while still inheriting the fast recovery property of segment protection.

For achieving backup bandwidth efficiency, shared protection has been proposed

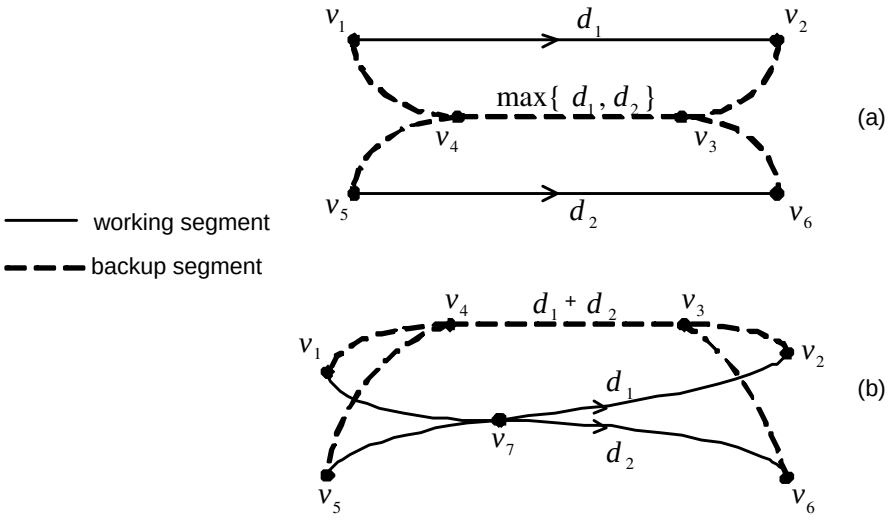

Figure 6.2: Examples of backup bandwidth sharable (a) and non-sharable (b) cases.

for link, path and segment protections [Ram99]. In segment protection, in order to guarantee 100% recovery under a single link or node failure, two backup segments can share some bandwidth if and only if their working segments are link and node-disjoint. We call this *segment sharing condition*. Fig. 6.2 gives an illustration. In case (a), the working segment from v_1 to v_2 with requested bandwidth d_1 and the working segment from v_5 to v_6 with requested bandwidth d_2 are link and node disjoint. Therefore their backup segment can share bandwidth over the common link (v_4, v_3) and the total bandwidth used by the two backup segments on this link is $\max\{d_1, d_2\}$. In case (b), the two working segments share node v_7, therefore their backup segments must reserve separate backup bandwidth. The total backup bandwidth for both backup segments on link (v_4, v_3) is $d_1 + d_2$ which is greater than in case (a).

With the shared protection feature, Overlapping Segment Protection becomes Overlapping Segment Shared Protection (OSSP). This paper aims to investigate the OSSP routing problem in multi-domain networks because of its characteristics: node protection, fast recovery and bandwidth saving.

Shared protection under static traffic has received a lot of interest. Many effi-

cient solutions have been proposed, especially the well-known p-cycle. It was initially introduced in [GS98] and further developed for segment protection in [SG03], [SG04]. However, network traffic changes unpredictably and dynamically today, static traffic is no longer an appropriate assumption unless in the network design or planning contexts. For this reason, we are focusing only on dynamic traffic where a new incoming request needs to be routed without any assuming forecast about the upcoming requests. The objective of the routing is to minimize the bandwidth capacity used by both working and backup segments of the considered request.

A multi-domain network is an interconnection of several single-domain networks [BSO02] (Fig. 6.3a). For the *scalability requirement*, only the aggregated routing information can be exchanged between domains [LRVB04] by an Exterior Gateway Protocol (EGP) such as Border Gateway Protocol (BGP). Consequently, a given node is neither aware of the global multi-domain network topology nor of the detailed bandwidth allocation on each network link. However, the complete routing information is still available within each domain thanks to more frequent routing information exchanges performed by an Interior Gateway Protocol (IGP) such as the link state routing protocols Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS) etc..

Most studies on OSSP remain within the single-domain network context. An optimal solution has been proposed in [HTC04] but it requires a huge computational effort even for small networks. Several heuristics with smaller computational efforts have been proposed such as the work in [RKM02], SLSP-O in [HM03], CDR in [HM02], PROMISE in [XXQ02] or recursive shared segment protection [CGYL07]. The first study ignores the sharing possibility during the routing. The other ones as well as the optimal solution scheme in [HTC04] are restricted to single domain networks as they assume that the global and detailed network information is available at any given internal node.

Some solutions have also been proposed for multi-domain networks with drawbacks. In [OMZ01], the working path is divided into non-overlapping segments at domain border nodes which are then not protected. In [ASL$^+$02], the authors

try to recover those border nodes by using an end-to-end restoration. However in comparison with protection, restoration offers a slower and uncertain recovery. Of course, the recovery quality declines. In [MKAM04], a simple multi-domain network without transit domain is assumed: domains do not directly connect to each other but connect directly to a central backbone domain. Connections from one domain to another one are easily established through some links of the backbone domain. In practice, neighboring domains connect to each other without a backbone domain and a connection between distant domains goes often through one or more transit domains. This makes the routing problem more complex.

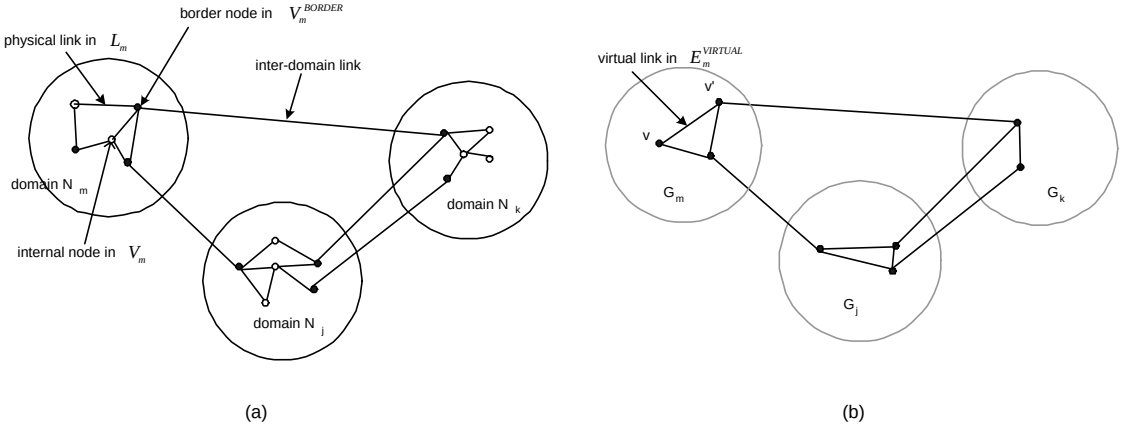

Figure 6.3: A multi-domain network (a) and its *inter-domain network* (b) obtained from Topology Aggregation.

In this study, we have developed a two-step heuristic solution. The multi-domain network is first topologically aggregated to become a compact network called *inter-domain network*, where a rough routing is sketched out. Then detailed routings are performed inside each original domain network. The use of an aggregated topology at the first step eliminates the need for global and detailed information requirements and thus preserves the scalability. The first routing step can be solved by using a greedy or dynamic programming algorithm (to be presented in sections 6.4.2 and 6.4.3) or any single-domain routing solution.

In the proposed solution, the working and backup segment lengths are also restricted. It is known that the failure recovery time consists mainly of the failure notification time and backup segment activation time. The first one is proportional to the working segment length and the second one is proportional to the backup segment length [CGYL07]. Therefore, the restrictions on the working and backup segment lengths will guarantee a fast recovery. Except for SLSP-O, in most published OSSP solutions, this restriction is not considered, leading usually to solutions with single segment patterns. The segment protection solutions degenerate thus to path protection solutions. The cost may decrease but the recovery time increases. In some experiments of PROMISE in [XXQ03b], the length of a backup segment is bounded by a function of its working segment length. However, this bound is not tight as the working segment length is not bounded. On the contrary, we restrict both working and backup segment lengths in our study. The single-domain solution in [CGYL07] and SHALL in [LYL06] restrict the working and backup segment lengths. However, SHALL uses Suurballe and Tarjan′s algorithm [ST84], which does not require overlapping between segments, and thus cannot offer the node protection capability. Similarly, the work in [CGYL07] does not require backup segment overlapping either.

We consider networks with bandwidth guaranteed connections such as SONET/SDH, MPLS-TE, ATM or DWDM networks. In the case of DWDM networks, each network node is assumed to be equipped with Multiservice Provisioning Platform (MSPP, see i.e. [Muk06]) with bandwidth grooming and wavelength conversion abilities. The wavelength continuity constraint and wavelength assignment problem are thus relaxed. Without bandwidth grooming, the proposed solution is still applicable on DWDM network as long as one wavelength is considered as a bandwidth unit.

This paper is organized as follows notations and fundamental concepts are introduced in the next section. Section 6.3 presents link costs which will be used in the routing algorithms proposed in Section 6.4. Section 6.5 outlines the signaling processes that coordinate the routing, the connection setup as well as the informa-

tion update. Section 6.6 shows the computational results. Section 6.7 concludes the paper.

6.2 Fundamental concepts and Notations

The multi-domain network is represented by a graph $\mathcal{N} = (V, L)$ composed of M connected single-domain networks $\mathcal{N}_m = (V_m, L_m)$, $m = 1, .., M$ where V, V_m are sets of nodes and L, L_m are sets of links. Each single-domain network contains border nodes which connect with the border nodes of other domains through inter-domain links (see Fig. 6.3a). The set of border nodes of $\mathcal{N}_m$ is V_m^{BORDER}. The set of inter-domain links of the multi-domain network is $L^{\text{INTER}} \subset L$. Thus:

$$V = \bigcup_{m=1..M} V_m,$$

$$L = \left(\bigcup_{m=1..M} L_m \right) \cup L^{\text{INTER}}.$$

A full mesh topology aggregation (TA) will be applied to each domain network. The TA on domain $\mathcal{N}_m$ results in an aggregated graph $G_m = (V_m^{\text{BORDER}}, E_m^{\text{VIRTUAL}})$ containing only border nodes of $\mathcal{N}_m$ and a set of virtual links connecting all pairs of border nodes $E_m^{\text{VIRTUAL}} = \{(v_1, v_2) : v_1, v_2 \in V_m^{\text{BORDER}}\}$. A virtual link $(v_1, v_2) \in G_m$ represents the set of intra domain paths (called intra-paths) inside $\mathcal{N}_m$ from v_1 to v_2. The multi-domain network is transformed into the compact network $G = (V^{\text{BORDER}}, E)$, called *inter-domain network* (see illustration on Fig. 6.3b), where

$$V^{\text{BORDER}} = \bigcup_{m=1..M} V_m^{\text{BORDER}},$$

$$E = \left(\bigcup_{m=1..M} E_m^{\text{VIRTUAL}} \right) \cup L^{\text{INTER}}.$$

We will denote by e an edge of G, e can then be a virtual link or an inter-domain

link. Let $\mathcal{P}_e$ be the set of intra-paths represented by e if e is a virtual link and $\mathcal{P}_e = \{e\}$ if e is an inter-domain link. Edge e will be associated with some link-states containing aggregated routing information obtained from its intra-paths. Such aggregated information can be exchanged between border nodes without impairing the *scalability requirement*. The *inter-domain network* can be then viewed as a single-domain network.

Let us consider a request with bandwidth d from node v_s to node v_d that just arrives in the network and needs to be routed without bifurcation. We refer to this request by "the new incoming request". We have to find an end-to-end working path p made of $|I|$ segments $\{p_i, i \in I\}$, and a set of backup segments $\{p'_i, i \in I\}$ such that their total bandwidth capacity is minimized. The backup segment p'_i protects the working segment p_i. The working path consumes bandwidth d on each of its links without any sharing. Before describing the routing algorithms, we need to introduce additional notations.

6.2.1 Notations used for the original multi-domain network

Because most notations are related on the new incoming request, for the shake of simplification, the indexes concerning this request such as p and p'_i will be omitted.

c_ℓ^{res} total residual bandwidth capacity on physical link $\ell \in L$.

a_ℓ bandwidth to be used by the working path p of the new incoming request on physical link $\ell \in L$ if p contains ℓ.

$B_{\ell'}$ total backup bandwidth already reserved by backup segments on physical link $\ell' \in L$ before the routing of the new incoming request.

$B_{\ell'}^v$ fraction of backup bandwidth on physical link $\ell' \in L$ that is reserved by the backup segments whose working segments go through node $v \in V$. Of course, $B_{\ell'}^v \leq B_{\ell'}$. This backup bandwidth cannot be shared with the backup segments of the new incoming request that protect v otherwise it would violate the *segment sharing condition*.

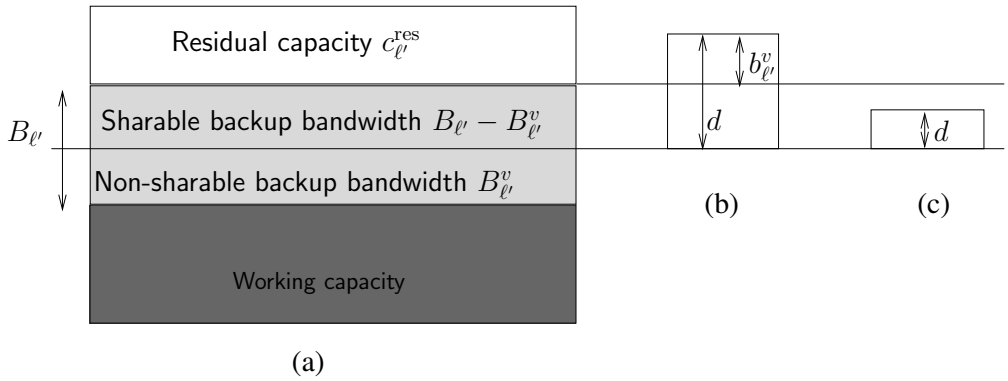

Figure 6.4: Bandwidth structure on a physical link ℓ' (a) two examples of the required additional backup bandwidth on that link (b), (c).

$B_{\max}^v = \max_{\ell' \in E} B_{\ell'}^v$ and $B_{\max}^q = \max_{v \in q} B_{\max}^v$ the maximum backup bandwidths reserved on a network link in order to protect the working segments going through node v and through sub-path q respectively.

$b_{\ell'}^v$ additional backup bandwidth (with regard to the existing backup bandwidth $B_{\ell'}$) that a backup segment of the new incoming request needs to reserve on a physical link ℓ' in order to ensure the survival of all already protected working segments as well as of its own working segment when node v fails.

$b_{\ell'}^q$ additional backup bandwidth (with regard to the existing backup bandwidth $B_{\ell'}$) that a backup segment of the new incoming request needs to reserve on a physical link ℓ' in order to ensure the survival of all already protected working segments as well as its own working segment when a node or a link of the sub-path q fails.

Fig. 6.4a illustrates the bandwidth structure on a physical link.

Clearly, the sharable backup bandwidth on ℓ' for protecting a node v is $(B_{\ell'} - B_{\ell'}^v)$. This bandwidth is profitable for the new incoming request in order to protect v. There are two cases: i) the sharable backup bandwidth is not greater than the requested bandwidth d, we need an additional bandwidth $b_{\ell'}^v = d - (B_{\ell'} - B_{\ell'}^v)$ on ℓ' (Fig. 6.4b); ii) the sharable backup bandwidth is greater than d, no additional backup bandwidth is necessary and $b_{\ell'}^v = 0$ (Fig. 6.4c). In other words:

$$b^v_{\ell'} = \max\{0, B^v_{\ell'} + d - B_{\ell'}\}. \tag{6.1}$$

Readers are encouraged to read [KL00] and [XQX02] for detailed and similar computations in the case of link protection.

Observe that with OSSP, for a given node, the same backup segments must be activated when either this node fails or all its adjacent links fail simultaneously. The solution that protects a node is sufficient to protect every adjacent link of the node. We deduce the following result:

Property 1. *The backup bandwidth required on a link (ℓ') by one backup segment of a connection in order to protect a working segment (q) is equal to the largest backup bandwidth needed on this link in order to protect a node of the working segment:*

$$b^q_{\ell'} = \max_{v \in q} b^v_{\ell'}. \tag{6.2}$$

6.2.2 Notations used for the *inter-domain network*

$\pi, \pi_i, \pi'_i (i \in I)$ representations of the working path p, working and backup segments p_i, p'_i of the new incoming request in the *inter-domain network* G.

$q \mapsto e$ indicates that the intra-path $q \in \mathcal{P}_e$ is the part, represented by $e \in G$, in the working or backup segments of the new incoming request.

α_e total working bandwidth that the working path p of the new incoming request consumes along its sub-path $q \mapsto e \in E$. Thus, $\alpha_e = \sum_{\ell \in q} a_\ell$.

$\beta^e_{e'}$ (resp. $\beta^{\pi_i}_{e'}$) total additional backup bandwidth that a backup segment of the new incoming request needs to reserve along $q' \mapsto e' \in \pi'_i$ to protect $q \mapsto e \in \pi_i$ (resp. to protect p_i). Thus, $\beta^e_{e'} = \sum_{\ell' \in q'} b^q_{\ell'}$ and $\beta^{\pi_i}_{e'} = \sum_{\ell' \in q'} b^{p_i}_{\ell'}$.

$\overline{B}_{e'} = \begin{cases} B_{\ell'} & \text{if } e' = \ell' \in L^{\text{INTER}} \\ \max\limits_{\ell' \in L_m} B_{\ell'} & \text{if } e' \in E_m^{\text{VIRTUAL}} \end{cases}$. If e is a virtual link, $\overline{B}_{e'}$ is the maximum backup bandwidth reserved on a physical link of the domain to whom e belongs. If e is an inter-domain link, it is the existing backup bandwidth on e.

γ_e^{res} maximal bandwidth that can be routed over any intra-path $q \in \mathcal{P}_e$ of $e \in E$. $\gamma_e^{\text{res}} = \max\limits_{q \in \mathcal{P}_e} \min\limits_{\ell \in q} c_\ell^{\text{res}}$.

$\|e\|$ length of the shortest intra-path represented by e. It is also called the estimated length of e.

The parameters a and b with different indexes denote the working and backup costs of physical links. Similarly, α and β denote the working and backup costs of the virtual and inter-domain links.

6.3 Costs of virtual and physical links

In this section, the costs of virtual and physical links are presented. We will see later in Section 6.4 that these costs are essential parameters for the proposed routing algorithms.

6.3.1 Estimations of the costs of virtual links

The exact values of the costs $\alpha_e, \beta_{e'}^e, \beta_{e'}^{\pi_i}$ of a virtual link e or $e' \in E$ depend on the intra-path q on p or $p'_i, i \in I$ that e or e' represents, e.g., $q \mapsto e$ or $q \mapsto e'$. However, q, p and $p'_i, i \in I$ are unknown until the routing completion making the exact computation of these values impossible before the end of the routing. Moreover, these costs are associated with the *inter-domain network* where physical link information is inaccessible. Therefore, we will use approximations to remove the physical link dependent parameters and define these costs as functions of virtual link dependent parameters.

The working cost of $e \in E$ is defined as the smallest total bandwidth that the working path p consumes along e. The choice of taking the smallest total bandwidth but not the average or other estimations is originated from the objective of minimizing bandwidth cost. The intra-path of $\mathcal{P}_e$ with minimum total bandwidth will be the best intra-path that p should go through. Thus:

$$\alpha_e = \begin{cases} \|e\| \times d & \text{if } d \leq \gamma_e^{\text{res}}, e \in E^{\text{VIRTUAL}} \\ d & \text{if } d \leq \gamma_e^{\text{res}}, e \in L^{\text{INTER}} \\ \infty & \text{otherwise.} \end{cases} \tag{6.3}$$

The approximation of the backup cost $\beta_{e'}^{\pi_i}$ is more complex. Let us begin with $b_{\ell'}^v$ defined in (6.1). In order to eliminate the dependency of $b_{\ell'}^v$ on detailed information $B_{\ell'}^v$, $b_{\ell'}^v$ is overestimated by: $\max\{0, B_{\max}^v + d - B_{\ell'}\}$. Remind that $b_{\ell'}^v$ cannot be greater than the requested bandwidth. We get the following overestimation:

$$b_{\ell'}^v = \min\{\max\{0, B_{\max}^v + d - B_{\ell'}\}, d\}. \tag{6.4}$$

From this, it can be proved that the backup cost of a virtual or inter-domain link for protecting a working segment is not smaller than the cost for protecting a virtual/inter-domain link of the segment. Thus:

$$\beta_{e'}^{\pi_i} = \max_{e \in \pi_i} \beta_{e'}^e. \tag{6.5}$$

The cost $\beta_{e'}^e$ is also approximated in its turn. Since $\beta_{e'}^e = \sum_{\ell' \in q'} b_{\ell'}^q$, it is lower bounded by the minimum backup bandwidth that should be reserved along e':

$$\beta_{e'}^e \geq \min_{q \in \mathcal{P}_e, q' \in \mathcal{P}_{e'}} \sum_{\ell' \in q'} b_{\ell'}^q, \tag{6.6}$$

where

$$b_{\ell'}^q \geq \min\{\max\{0, B_{\max}^q + d - B_{\ell'}\}, d\}$$

as $b^q_{\ell'} = \max_{v \in q} b^v_{\ell'}$ and $B^q_{\max} = \max_{v \in q} B^v_{\max}$.

Thus:

$$\beta^e_{e'} \geq \min_{q \in \mathcal{P}_e, q' \in \mathcal{P}_{e'}} \sum_{\ell' \in q'} \min\{\max\{0, B^q_{\max} + d - B_{\ell'}\}, d\}.$$

Since $\overline{B}_{e'} \geq B_{\ell'}$, for all $\ell' \in q \mapsto e$ then:

$$\beta^e_{e'} \geq \min_{q \in \mathcal{P}_e} \|e'\| \times \min\{\max\{0, B^q_{\max} + d - \overline{B}_{e'}\}, d\}. \tag{6.7}$$

Let v_1, v_2 be the border end nodes of e and $B^e_{\max} = \max\{B^{v_1}_{\max}, B^{v_2}_{\max}\}$. Clearly $B^e_{\max} \leq B^q_{\max}$. Thus we have:

$$\beta^e_{e'} \geq \|e'\| \times \min\{\max\{0, B^e_{\max} + d - \overline{B}_{e'}\}, d\}. \tag{6.8}$$

Let us underestimate $\beta^e_{e'}$ by the right-hand side of (6.8) which is in fact the lower bound of the backup bandwidth that should be reserved along e' for p'. Taking into account the link capacity, we define:

$$\beta^e_{e'} = \begin{cases} 0 & \text{if } B^e_{\max} + d \leq \overline{B}_{e'} \\ \|e'\| \times (B^e_{\max} + d - \overline{B}_{e'}) & \text{if } B^e_{\max} + d > \overline{B}_{e'} > B^e_{\max} \\ & \text{and } \gamma^{\text{res}}_{e'} \geq B^e_{\max} + d - \overline{B}_{e'} \\ \|e'\| \times d & \text{if } B^e_{\max} \geq \overline{B}_{e'} \text{ and } \gamma^{\text{res}}_{e'} \geq d \\ \infty & \text{otherwise.} \end{cases} \tag{6.9}$$

In summary, the working and backup costs of a virtual or inter-domain link are represented by functions of the virtual link dependent parameters: $\|e\|$, γ^{res}_e, $\overline{B}_{e'}$, $B^e_{\max}$. These parameters define the link-states of e. Border nodes exchange among themselves these link-states in order to get a common view of the compact *inter-domain network*.

6.3.2 Costs of physical links

The working cost a_ℓ of physical link ℓ is exactly defined by:

$$a_\ell = \begin{cases} d & \text{if } d \leq c_\ell^{\text{res}} \\ \infty & \text{otherwise.} \end{cases} \tag{6.10}$$

From (6.4) and the definitions of $b_{\ell'}^q$ and $B_{\max}^q$, it is easy to deduce that: $b_{\ell'}^{p_i} = \min\{\max\{0, B_{\max}^{p_i} + d - B_{\ell'}\}, d\}$, i.e.:

$$b_{\ell'}^{p_i} = \begin{cases} 0 & \text{if } B_{\max}^{p_i} + d - B_{\ell'} \leq 0 \\ B_{\max}^{p_i} + d - B_{\ell'} & \text{if } B_{\max}^{p_i} + d > B_{\ell'} > B_{\max}^{p_i}, \\ & \quad c_{\ell'}^{\text{res}} \geq B_{\max}^{p_i} + d - B_{\ell'} \\ d & \text{if } B_{\max}^{p_i} \geq B_{\ell'}, c_{\ell'}^{\text{res}} \geq d \\ \infty & \text{otherwise.} \end{cases} \tag{6.11}$$

6.4 Routing solutions

6.4.1 Outline of the solution

In this study, the objective of the routing is to minimize the total bandwidth consumed by p and $p'_i, i \in I$ of the new incoming request. It can be expressed as follows:

$$\min \sum_{\ell \in p} a_\ell + \sum_{p'_i, i \in I} \sum_{\ell' \in p'_i} b_{\ell'}^{p_i}. \tag{6.12}$$

In the *inter-domain network*, it is equivalent to:

$$\min \sum_{e \in \pi} \alpha_e + \sum_{\pi'_i, i \in I} \sum_{e' \in \pi'_i} \beta_{e'}^{\pi_i}. \tag{6.13}$$

In multi-domain networks, paths tend to be long. In order to guarantee a fast recovery, we require that each working and backup segments are not longer than l^{W} and l^{B} thresholds respectively. This requirement is afterward referred as segment

length constraints.

We propose a two-step routing as follows:

- *Inter-domain step*: We first optimize (6.13) in the *inter-domain network* where virtual and inter-domain links are assigned costs α_e and $\beta_e^{\pi_i}$. The constraints on working and backup segment lengths are also taken into account. This gives us segments π_i and $\pi_i', i \in I$ as paths of virtual/inter-domain links. If no solution is found, the routing fails. Otherwise, the *intra-domain step* will follow.

  In fact, (6.13) is an OSSP single-domain routing problem. All OSSP single-domain routing algorithms cited in this paper can be used to solve (6.13) as long as they are applied on the *inter-domain network* with the proposed virtual link costs and the segment length constraints are integrated. Two solution schemes, GROS and DYPOS, are proposed in the next two paragraphs 6.4.2, 6.4.3.

- *Intra-domain step*: The segment pairs $(\pi_i, \pi_i'), i \in I$ are considered one after the other. For each pair, the virtual links of the working segment are mapped first to the intra-paths with the least working costs:

$$\min_{q \in \mathcal{P}_e} \sum_{\ell \in q} a_\ell (= \alpha_e). \tag{6.14}$$

  The selected intra-path for the virtual link e is indeed the Shortest Path (SP) in terms of physical working costs a_ℓ between the end nodes of the virtual link. Once the complete working segment p_i is obtained, the virtual links of π_i' will be mapped similarly into the SP but in terms of $b_{\ell'}^{p_i}$:

$$\min_{q' \in \mathcal{P}_{e'}} \sum_{\ell' \in q'} b_{\ell'}^{p_i} (= \beta_{e'}^{\pi_i}). \tag{6.15}$$

  Note that the nodes along p_i are excluded in this mapping in order to guarantee the disjointness between the working and backup segments of a given

pair.

Each mapping relates to only one domain and can be solved using Dijsktra's SP algorithm within the domain while respecting the *scalability requirement*.

6.4.2 GROS: A greedy solution

The first routing solution for the *inter-domain step* is a greedy heuristic denoted by GROS (GReedy Overlapping Short segment shared protection). For each new incoming request, the GROS heuristic works as follows.

1. Working path π is the shortest path in the *inter-domain network* between the source and the destination in terms of working costs α_e.

2. The working path is greedily divided into segments. The first segment π_1 originates from the source node of the working path. The tail node of each segment is chosen so that the segment is as long as possible with a total estimated length that does not exceed l^{W}. However, if no such tail node is found, the node that is closest to the head node will be designated as tail node. From the tail node, we go back toward the head node with the smallest number of hops until reaching a new node with nodal degree larger than 2. This last node will be the head of the next segment. The process continues until the destination node is reached.

3. For each previously identified working segment, a backup segment is computed as the shortest path in terms of backup costs $\beta_{e'}^{\pi_i}$ between the segment end nodes. The total estimated length of the segment must not be larger than l^{B}. The shortest path with additive constraint algorithm A*Prune (or A*Dijkstra) [LR01] is used for computing each backup segment.

We can remark that, in the GROS heuristic, the working segment length may sometimes exceed the threshold l^{W}. In other words the working segment length constraint is soft.

If the algorithm does not find a solution at a given step, the routing fails.

GROS differs from CDR in [HTC04]. In CDR, a set of segment end nodes are predefined for each pair of source and destination before the working path is identified. From these segment end nodes, the working and backup segments are computed. In GROS, we determine only the segment end nodes once the working path is routed in the *inter-domain network*.

6.4.3 DYPOS: A Dynamic Programming solution

The second routing solution for the *inter-domain step* is called DYPOS (DYnamic Programming Overlapping Short segment shared protection). It is inspired from PROMISE Dynamic programing solution (PRO-D) [XXQ03b] for single-domain networks. The difference is in the integration of the working and backup segment length constraints.

Let us first briefly recall PRO-D. In PRO-D, the working path is the shortest path between the source and the destination. The backup segment is computed as follows. Assume that the nodes along the working path are numbering from 0 to T. Let $i \rightarrow j$ denotes the working segment from node i to node j. Let D_m be the "best known" solution to protect a part of the working path from node m to node T exclusively. D_m divides possibly that part into multiple overlapping segments and protects each of them by one segment. The current D_m is compared with each alternative solution built from $D_i, i \in [m+1..T-1]$ and the least cost backup segment that protects the part $m \rightarrow i$ and overlaps with the part $i \rightarrow T$. The backup segment is denoted by $p'_{m \rightarrow i}$. The best solution will be newly assigned to D_m. The algorithm starts by building the segment for the last hop ($m = T-1$) using the shortest backup path. The protected part is growing up until the entire working path is protected ($m = 0$) (Fig. 6.5).

In DYPOS, for computing each D_m, we consider only the alternative solutions associated with D_i such that the estimated length of the part $m \rightarrow i$ does not exceeds l^{W}. In addition, while computing $p'_{m \rightarrow i}$, we use again the A*Prune algorithm in order to find a backup segment with an estimated length smaller than or equal to l^{B}.

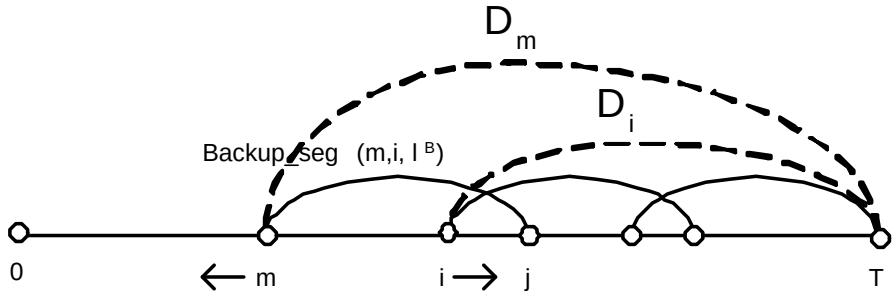

Figure 6.5: Working mechamism of the Dynamic programming algorithm

The pseudo-code in Alg.1 describes formally DYPOS. Function $\texttt{CSP}^{\text{B}}(m, T, l^{\text{B}})$ implements the A*Prune algorithm. It identifies the shortest path from m to T (using the backup cost $\beta_{e'}^{\pi_i}$) that must not be longer than l^{B} (in terms of estimated length). We denote by $||m \rightarrow i||$ the total estimated length of $m \rightarrow i$. `Backup_seg` (m, i, l^{B}) computes $p'_{m \rightarrow i}$. The backup segment $p'_{m \rightarrow i}$ must end at a node $j > i$ in order to create overlapping between its working segment and the working part $i \rightarrow T$. `Backup_seg` (m, i, l^{B}) identifies N least cost segment candidates from m to j with $j = [i+1..i+N]$ using $\texttt{CSP}^{\text{B}}(m, j, l^{\text{B}})$ and returns the least cost one.

DYPOS differs from GROS as the segment length constraints are hard constraints. If DYPOS finds no solution, it reports a failed routing.

6.4.4 Blocking-go-back option

A request may be successfully routed at the *inter-domain step* but blocked at the *intra-domain step* because of insufficient bandwidth for mapping a virtual link or impossibility of mapping a virtual link of a backup segment while maintaining the disjointness with its working segment. Let us call *blocking virtual link* the virtual link where the blocking occurs. In order to avoid such blocking cases, a second routing iteration is added to GROS and DYPOS. The second routing iteration is identical to the first one except that in the *inter-domain step*, the blocking virtual link is removed before the working path or backup segment computation, depending

Algorithm 1 DYPOS

for $m = T - 1$ down to 0 **do**
  if $||m \rightarrow T|| \leq l^{\mathrm{W}}$ **then**
    $D_m \leftarrow \mathtt{CSP}^{\mathrm{B}}(m, T, l^{\mathrm{B}})$
  else
    $D_m = \infty$
  end if
  for $i = m + 1$ to $T - 1$ **do**
    if $||m \rightarrow i|| \leq l^{\mathrm{W}} - 1$ **then**
      $p'_{m \rightarrow i} = \mathtt{Backup\_seg}\ (m, i, l^{\mathrm{B}})$
      $D_m \leftarrow \min(D_m, \mathtt{Combine}\ (D_i, p'_{m \rightarrow i}))$
    end if
  end for
end for
return D_0

Algorithm 2 Backup seg (m, i, l^{B})

$bs = \infty$
for $j = i + 1$ to $\min(i + N, T)$ **do**
  if $||m \rightarrow j|| \leq l^{\mathrm{W}}$ **then**
    $bs \leftarrow \min(bs, \mathtt{CSP}^{\mathrm{B}}(m, j, l^{\mathrm{B}}))$
  end if
end for
return bs

on whether the blocking virtual link was on the working path or backup segments. This removal prevents from repeating the previous blocking. Then the intra-domain step, as described in 6.4.1, is applied again. A failed routing is reported if a new blocking is produced.

Although the Blocking-go-back step allows the reduction of blocking probability, it takes longer to route a request when the routing fails at the first routing iteration. The routing time without Blocking-go-back step is: $T_r = T_r^{\text{INTER}} + T_r^{\text{INTRA}}$ where T_r^{INTER} (resp. T_r^{INTRA}) is the execution time of the *inter-domain step* (resp. *intra-domain step*). When the routing is blocked, the Blocking-go-back step is triggered. Certainly, the Blocking-go-back step introduces extra routing time arising from the second routing iteration, the total routing time is $T_r = 2 \times T_r^{\text{INTER}} + 2 \times T_r^{\text{INTRA}}$. In other words, the routing time with Blocking-go-back is twice the routing time without Blocking-go-back. We verified this through experimental results on two multi-domain networks LARGE-5 and LARGE-8 which will be described later. In LARGE-5, the average routing time when Blocking-go-back step is involved is 89.06 milliseconds while it takes 37.16 milliseconds without Blocking-go-back step. In LARGE-8, the average routing time with Blocking-go-back is 329.71 milliseconds and without Blocking-go-back is 134.09 milliseconds. Note that this routing time sacrifice is compensated in lower blocking probability.

6.5 Signaling and routing information update

Contrary to the OSSP routings existing in the literature, the OSSP routing in multi-domain networks is performed in a distributed way in different domains and requires signaling processes for coordinating the segment computation, segment setup and also routing information update. We will not discuss here the details of how the signaling protocols should be implemented as well as which message formats should be used. We describe only the interaction among network nodes.

6.5.1 Signaling for working and backup segment computation

The *inter-domain step* is performed centrally at the border source node without impairing the *scalability requirement* since the *inter-domain network* is considered as a single-domain network. In this step, the border source node computes the working and backup costs $\alpha_e, \beta_{e'}^{\pi_i}$ for each link of the *inter-domain network* by using link-states $\|e\|$, γ_e^{res}, $\overline{B}_{e'}$, $B_{\max}^e$ which are available at each border node thanks to the routing information update process that will be described later. Then GROS or DYPOS can be used for performing the *inter-domain step*. Once the computation is finished, the border source node asks the other border nodes along its working and backup segments to map subsequently the adjacent virtual links into intra-paths.

At the reception of the mapping request, the border node triggers the *intra-domain step* within its domain. It first computes the costs $a_\ell, b_\ell^{p_i}$ using the detailed information available in the domain and then solves mapping problems (6.14) and (6.15). The border node returns the mapped intra-path to the border source node.

From the mapped intra-paths, the border source node builds the complete working and backup segments.

6.5.2 Signaling for working and backup segment setup

A message carrying the information of the complete working path and backup segments is propagated along the working path from the source node to the destination node. At each node on the working path, switch is made in order to establish the end-to-end working path. At each segment head node an additional message is created carrying the information on the corresponding backup segment. The message is propagated along the route of the backup segment until the segment tail node. At each node, it asks to reserve an additional amount of bandwidth $b_{\ell'}^v$ on its outgoing link that belongs to the backup segment. Note that no switch is made there. The process terminates when the destination node is reached.

6.5.3 Routing information update

After each routing, link-states of virtual links change. They should be updated for serving the *inter-domain step* of the next routing. Link-states $\|e\|$, γ_e^{res}, $\overline{B}_{e'}$, $B_{\max}^e$ are computed locally in the domain containing e by a border node of e. This node writes all these link-states in one message and sends it to other border nodes of the multi-domain network. A BGP like protocol could be used for link-state message diffusion.

Of course, for computing the link-states of e, the border nodes of e needs also the detailed routing information of its domain. A domain scope routing information exchange between domain nodes is also needed.

Routing information update is the most expensive process regarding the flow of messages. A number of $O({V^{\text{BORDER}}}^2)$ messages are exchanged among border nodes and of $O(V_m^2)$ within each domain leading to an overall number of $O({V^{\text{BORDER}}}^2) + \sum_{m=1}^{M} O(V_m^2)$ messages. Nevertheless, this number is still smaller than $O(V^2) = O((V^{\text{BORDER}} + \sum_{i=1}^{M} V_m)^2)$, the number of messages required by a single-domain solution.

For reducing furthermore the charge of update message flows, the update could be triggered less regularly in a time driven way. However, the routing will be less accurate since some routing information will be out of date.

6.6 Experimental results

We use different network and traffic instances for evaluating the efficiency of GROS and DYPOS. The following metrics: backup overhead and overall blocking probability are used for their evaluation.

6.6.1 Metrics

The working network cost is defined as the total working bandwidth used by all network links. The network cost is defined as the total working and backup

bandwidth used by all network links.

The *Backup overhead* is defined as the ratio between the network cost and the smallest working network cost less 1. This amounts to the backup bandwidth redundancy of a protection scheme. The smallest working network cost can be obtained when all working paths are the shortest paths.

The *Overall blocking probability* is defined as the percentage of the total rejected bandwidth out of the total bandwidth requested by all connections.

6.6.2 Comparison with optimal single-domain solution

We evaluate the efficiency of GROS and DYPOS by comparing their results on a multi-domain network with the results of the single-domain optimal solution [HTC04], denoted by Opt, on the equivalent flattened network. Due to the extremely high computational effort required by Opt, the comparison is made only on a small 5-domain network of 28 nodes with 70 requests. The Transit-Stub model of GT-ITM [ZCB96], a well known multi-domain network generator, is used for generating this network instance that we denote by SMALL-5 and present in Fig. 6.6. To route a request, GROS and DYPOS take few milliseconds whereas Opt takes few minutes. That means GROS and DYPOS are thousands of times faster than Opt. In a larger network, we cannot obtain any result from Opt. Due to the small scale of the network, the constraint on backup segment length is ignored by setting l^{B} very large for GROS and DYPOS. In Opt, neither working and backup segment lengths are restricted. We made also comparison with the results obtained from dedicated protection denoted by NoShare.

Fig. 6.7 shows that the proposed two-step solution with either GROS or DYPOS provides a backup overhead close to the backup overhead of Opt and far better than the backup overhead of NoShare. In other words, GROS and DYPOS yield a very good bandwidth saving rate. Do not forget that the constraint on l^{W} is present in GROS and DYPOS, while it is absent in Opt, therefore giving a slightly advantage to Opt. Recall also that, while GROS and DYPOS are scalable for multi-domain networks, Opt is clearly not. In this experiment and also in others

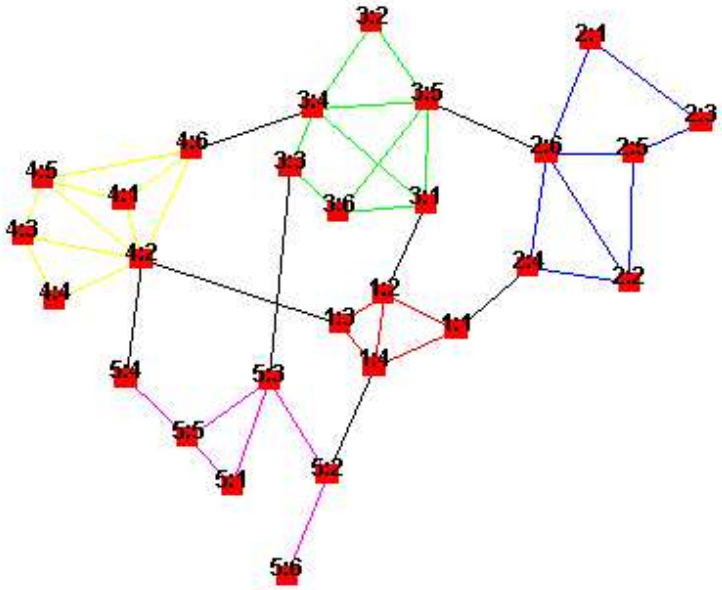

Figure 6.6: SMALL-5 network.

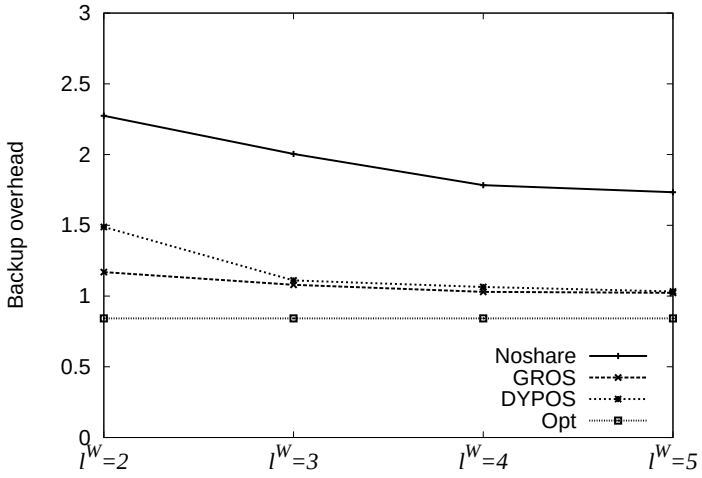

Figure 6.7: Backup overhead in SMALL-5.

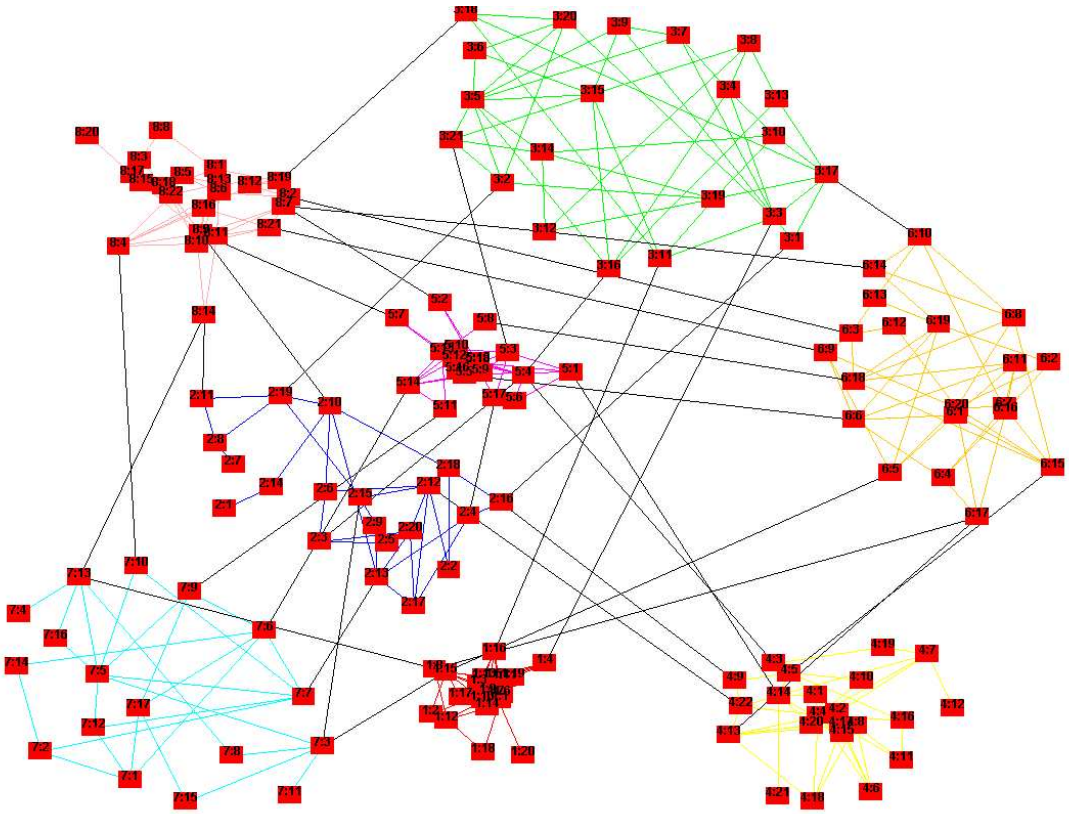

Figure 6.8: LARGE-8 network.

afterward, DYPOS yields sometimes larger backup overhead than GROS due to the working segment length constraint that is hard in DYPOS and soft in GROS. That forces DYPOS to consider the solutions with larger cost than those of GROS if the later violate the constraint on l^{W}. This phenomena reduces when l^{W} increases.

6.6.3 Backup overhead

From now on, the experiments are made on large multi-domain networks with heuristics only. The Transit-Stub model of GT-ITM, is again used for generating a larger multi-domain network with 8 domains, 36 inter-domain links and 60 border nodes. The network is denoted by LARGE-8 and is shown in Fig. 6.8. Each domain has in average 4 neighboring domains. According to [MP01], this number

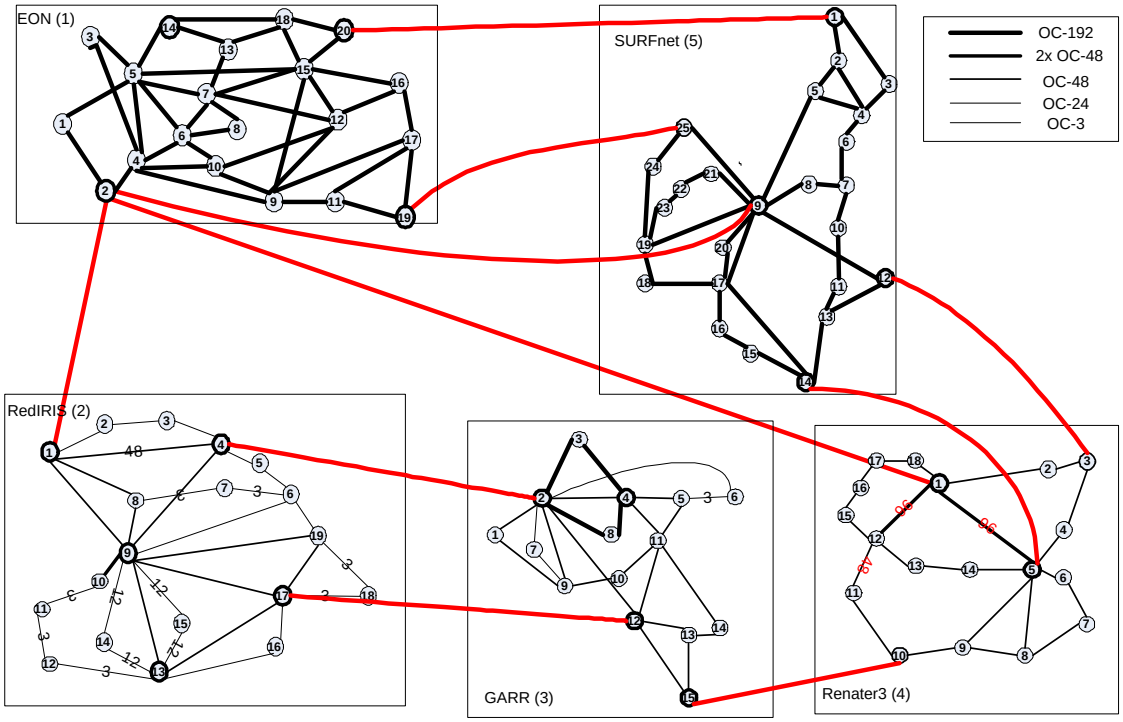

Figure 6.9: LARGE-5 network.

reflects faithfully the Internet interconnections. The numbers of nodes and links of each domain are: $(20, 53)$, $(20, 29)$, $(21, 48)$, $(22, 41)$, $(18, 36)$, $(20, 44)$, $(17, 27)$, $(22, 47)$, see [LAR07] for the details of the topology.

We also consider another multi-domain network that we used for experiments in previous papers [TT06], [JT06], [TJ06]. The network is built from 5 real optical networks: EON [OSYZ95], [RED05], [GAR05], [REN05], [SUR05]. Inter-domain links are added with capacity OC-192. The network is denoted by LARGE-5 and is shown in Fig. 6.9.

An incremental traffic is generated by submitting subsequently 1000 connection requests to the network with all requests remaining active. The incremental traffic allows keeping active more requests in the network and thus allows evaluating more accurately the bandwidth allocation characteristics of each solution scheme. Network links are uncapacitated in order to avoid the impact of blocking which is

different from one protection scheme to the other. Backup overhead is computed after 1000 requests.

Fig. 6.10 depicts the obtained backup overhead in LARGE-8 when using GROS, DYPOS in comparison with those of NoShare and WPF. This last scheme is a multi-domain Shared Path Protection proposed in [TT06], which will be renamed PATH in order to distinguish from the other protection schemes, i.e. of segment type. The working segment length thresholds are $l^{\mathrm{W}} = 3$ and $l^{\mathrm{W}} = 5$ and backup segment length threshold varies. Similar backup overheads are found in GROS and DYPOS with both $l^{\mathrm{W}} = 3$ and $l^{\mathrm{W}} = 5$. We observe that GROS and DYPOS require only around 0.55 and 0.8 times the working capacity for their backup, meanwhile NoShare requires 1.5 and up to 2.2 times the same amount with $l^{\mathrm{W}} = 5$ and $l^{\mathrm{W}} = 3$ respectively. In LARGE-5 (Fig. 6.11), we find a smaller but still significant difference between the backup overheads of NoShare and of the other schemes. This shows the advantage of shared protection over dedicated protection as well as the efficiency of GROS and DYPOS in favoring backup bandwidth sharing.

In general, we should also accept that OSSP (GROS and DYPOS in particular) can use more backup resources than Shared Path Protection (PATH in particular) because the backup segments need some extra links in order to joint the working path at segment end nodes. For some configurations, for example when $l^{\mathrm{W}} = 5$ in LARGE-8 or when $l^{\mathrm{W}} = 5, l^{\mathrm{B}} = 6$ or $l^{\mathrm{W}} = 3, l^{\mathrm{B}} = 4$ in LARGE-5, OSSP can be close to Shared Path Protection. In spite of the disadvantages with backup overhead, OSSP is still attractive due to its fast recovery characteristic which will be presented later in other experiments.

6.6.4 Blocking probability

The blocking probability is examined under dynamic traffic. In dynamic traffic, connections arrive and tear down after a holding time. Requests arrive according to Poisson process with rate $r = 1$ and holding time exponentially distributed with mean $h = 320$. There are, on average, 320 active connections in the network.

Fig. 6.12 depicts the overall blocking probability of GROS, of GROS with the

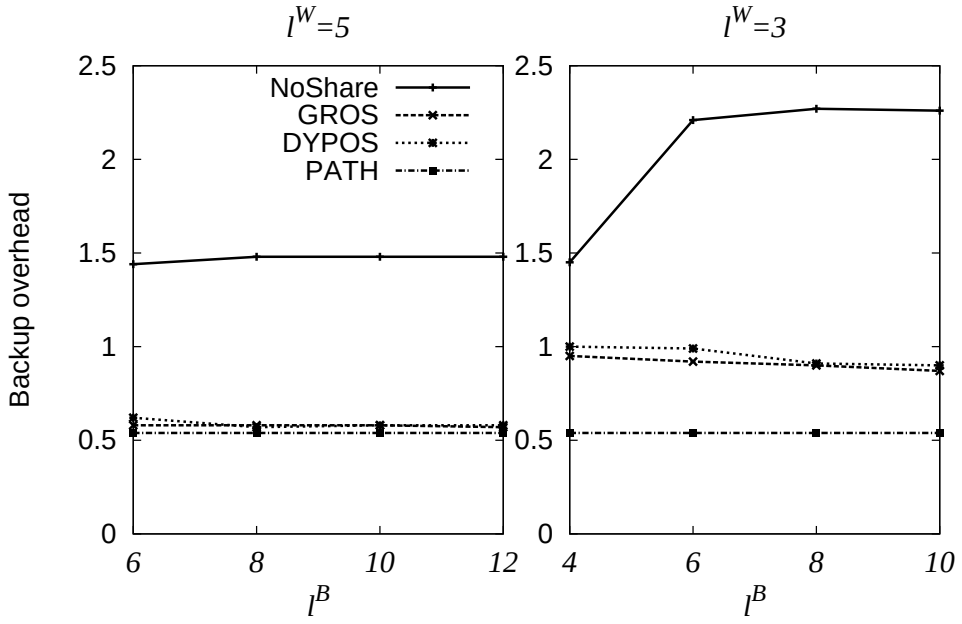

Figure 6.10: Backup overhead in LARGE-8.

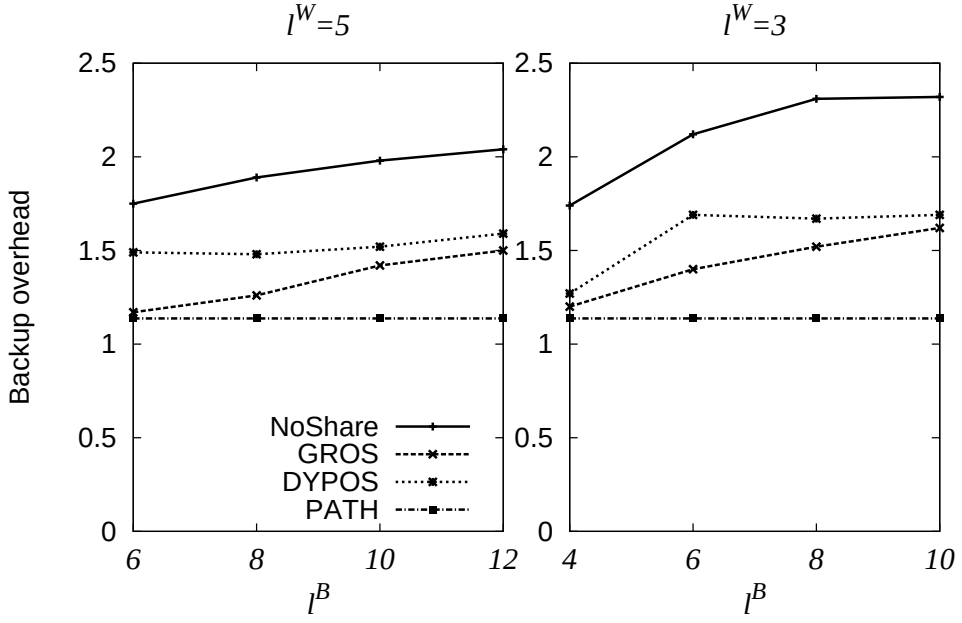

Figure 6.11: Backup overhead in LARGE-5.

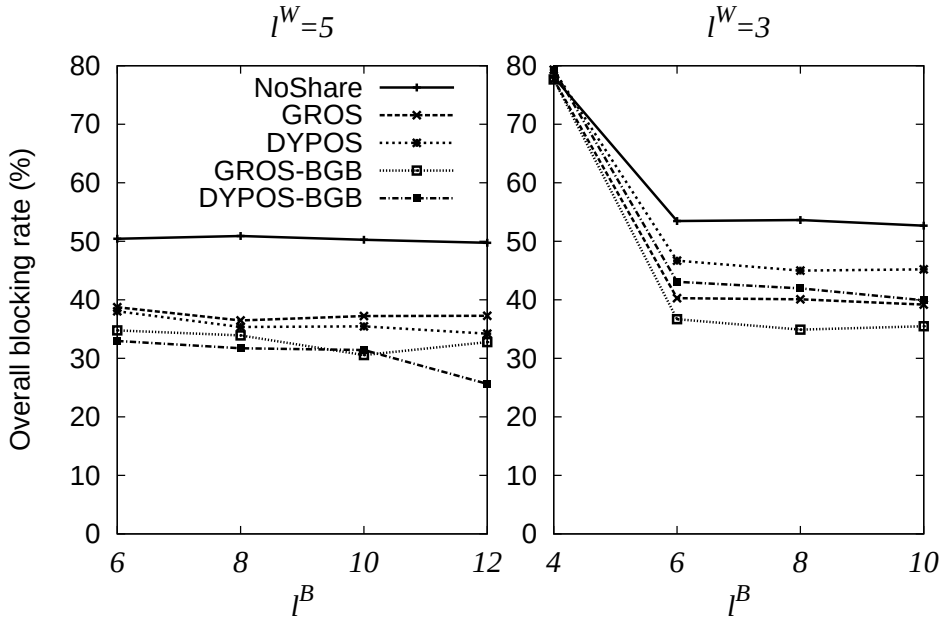

Figure 6.12: Overall blocking probabilities in LARGE-8.

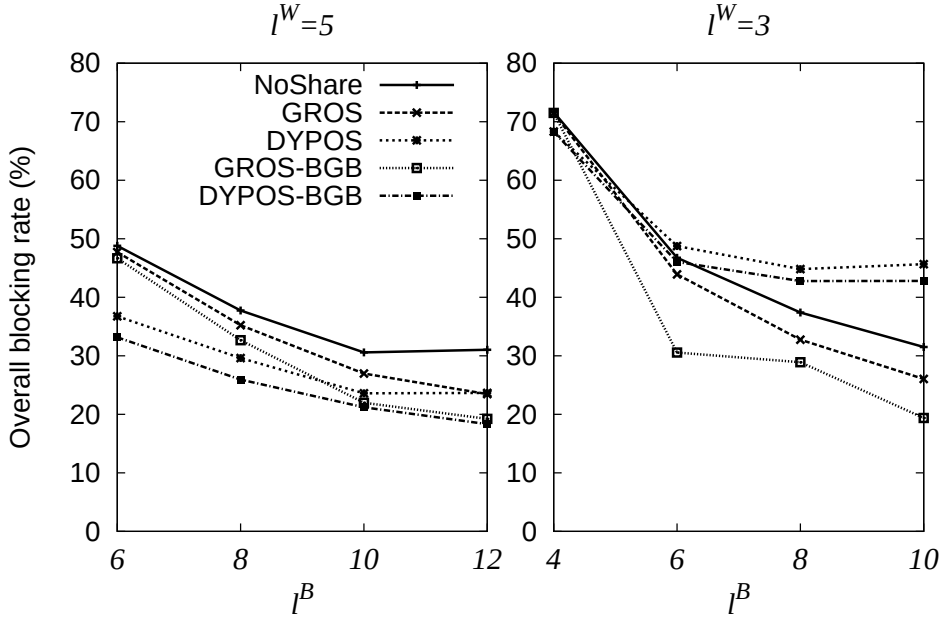

Figure 6.13: Overall blocking probabilities in LARGE-5.

Blocking-go-back option (denoted by GROS-BGB), of DYPOS and of DYPOS with the Blocking-go-back option (denoted by DYPOS-BGB) in LARGE-8. The four schemes keep NoShare at a distance. In LARGE-5 (see Fig. 6.13), the four schemes are still better than NoShare when $l^{\mathrm{W}} = 5$. However, when $l^{\mathrm{W}} = 3$, DYPOS and DYPOS-BGB become worse than GROS and sometimes even than NoShare. Two reasons can be given. First, the constraint on working segment length is hard in DYPOS and soft in the others. Second, LARGE-5 is less connected than LARGE-8 leading to less opportunity for dividing working paths into segments of 3 hops or less. This shows also the pertinence of properly defining segment lengths in low connected networks.

The blocking probabilities drop off for all schemes in both network topologies when the Blocking-go-back option is adopted. Fig. 6.14 and 6.15 show more clearly the advantage of the Blocking-go-back step. The GROS Inter and DYPOS Inter curves depict the percentages of the requests that are successfully routed in the *inter-domain step* of the second routing. Similarly, the GROS Intra and DYPOS Intra curves depict the percentages of the requests that are successfully routed after the *intra-domain step* of the second routing. A large number of requests that fails in the first routing is successfully routed in the *inter-domain step* of the second routing and about 30%-50% of them are successfully routed in the *intra-domain step* except when thresholds become two small, e. g. $l^{\mathrm{W}} = 3, l^{\mathrm{B}} = 4$. We can conclude that the Blocking-go-back step is useful for increasing the grade of service.

6.6.5 Impact of segment length

Fig. 6.16 and 6.17 show the average working segment lengths of GROS and DYPOS in comparison with other schemes in LARGE-5 and LARGE-8. Fig. 6.18 and 6.19 show the average backup segment lengths of these schemes. For GROS and DYPOS, as the segment length constraints are enforced in the inter-domain step with estimated virtual link lengths, the actual segment lengths in the original multi-domain network may exceed the thresholds l^{W} and l^{B}. However, the experiment results still show that the average working and backup segment lengths are always

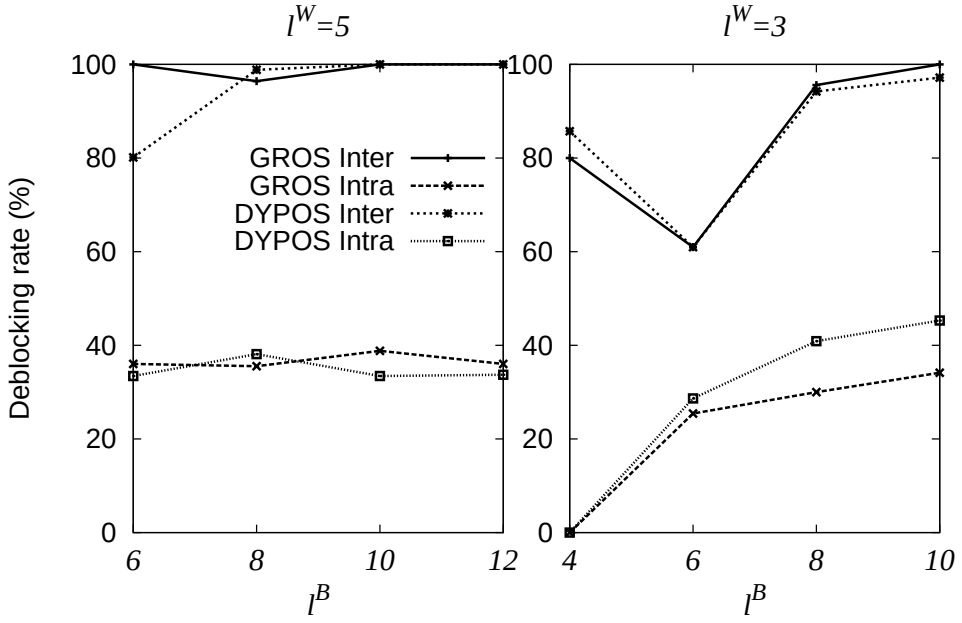

Figure 6.14: De-blocking capacity of the Blocking-go-back step in LARGE-8.

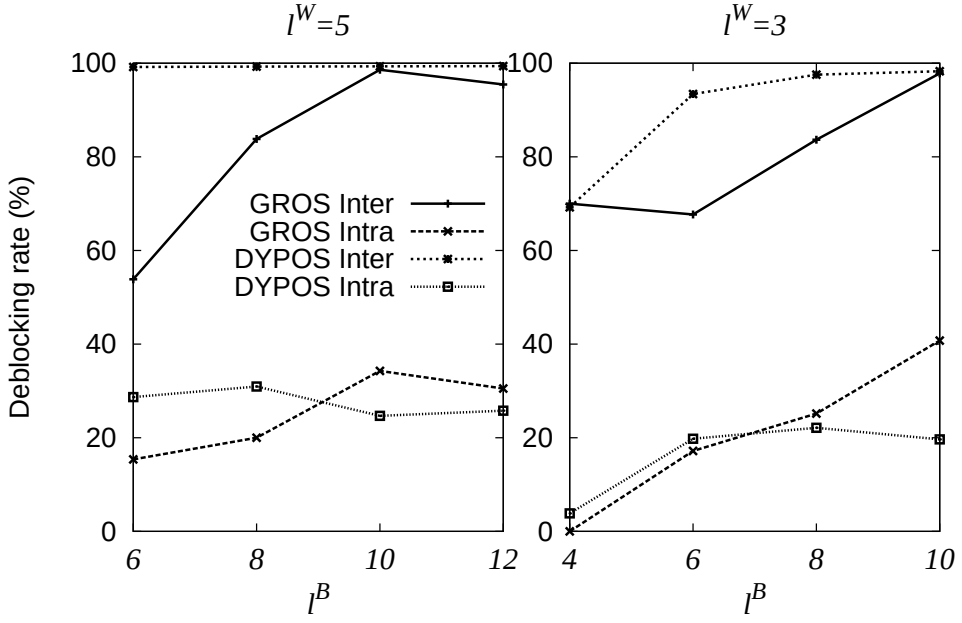

Figure 6.15: De-blocking capacity of the Blocking-go-back step in LARGE-5.

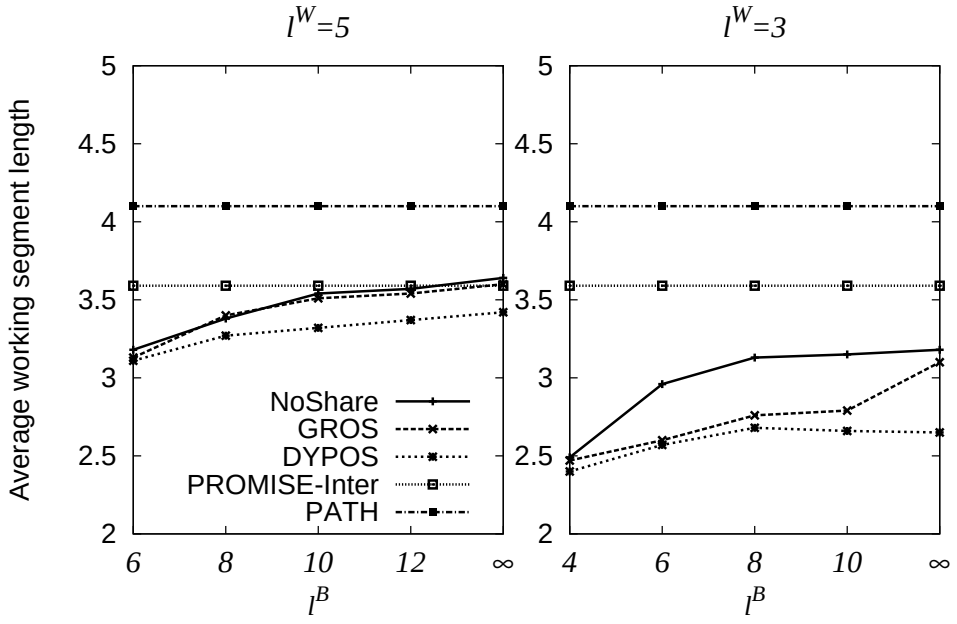

Figure 6.16: Average working segment lengths in LARGE-5

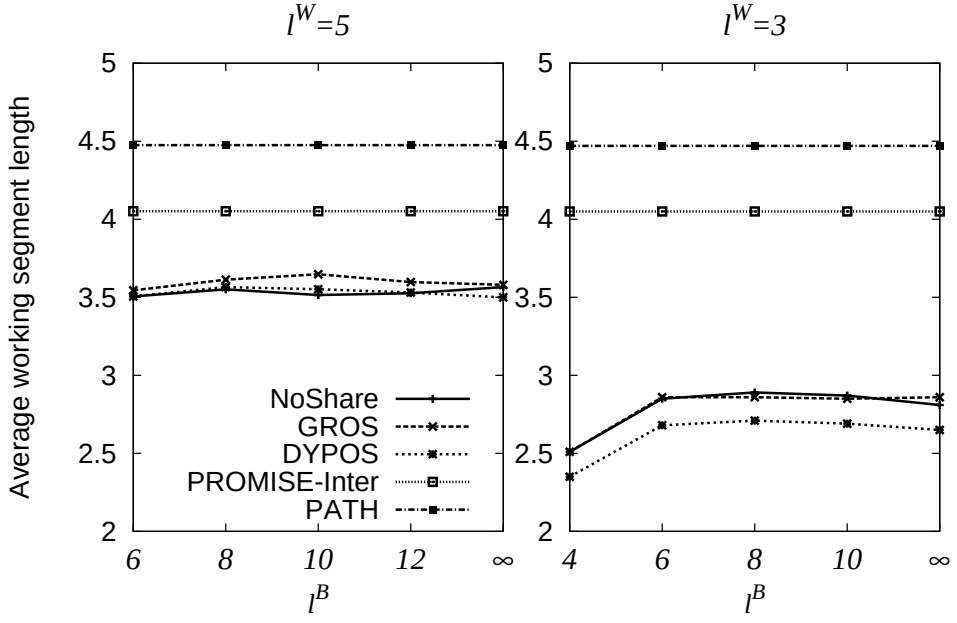

Figure 6.17: Average working segment lengths in LARGE-8

under the given limits l^{W} and l^{B}. The results at $l^{\mathrm{B}} = \infty$ corresponds to the elimination of the backup segment length constraint. Logically, working segments are longer when the threshold l^{W} increases from 3 to 5. Similarly, for a given working segment length threshold l^{W}, the average backup segment length becomes generally larger when l^{B} increases. Note that in both LARGE-5 and LARGE-8, there are 10%- 50% of cases that contain from 2 to 5 segments.

In Fig. 6.19, we observe that the backup segment lengths of NoShare are nearly constant with $l^{\mathrm{B}} \geq 6$ and that they are smaller than those of GROS and DYPOS. This can be explained by the fact that the backup segments of GROS and DYPOS take sometimes longer routes in order to benefit from the sharable backup bandwidth on some network links. On the contrary, NoShare does not allow backup bandwidth sharing, it has thus no interest to take these long routes, it simply takes the shortest routes for the smallest costs.

We also run DYPOS without both working and backup segment length constraints. This is similar to using PROMISE on the inter-domain routing. The corresponding results are denoted by PROMISE-Inter. The results show that GROS and DYPOS always give smaller average working segment lengths (Fig. 6.16, 6.17) and mostly give smaller average backup segment lengths (Fig. 6.18 and 6.19) than PROMISE-Inter. This illustrates the role of segment length constraints implemented in GROS and DYPOS. This demonstrates also that GROS and DYPOS recover from failures faster than PROMISE-Inter.

In comparison with the multi-domain Shared Path Protection scheme PATH, the average working segment lengths of GROS and DYPOS are clearly smaller than the average working path lengths of PATH (Fig. 6.16 and 6.17). Backup segments of GROS and DYPOS are also generally shorter than backup paths of PATH (Fig. 6.18 and 6.19). As a result, the recovery times of GROS and DYPOS are shorter than that of PATH.

Although short segment length promises fast recovery, it sometimes impairs backup overhead. When the segment length thresholds are too small, there are few choices for working path division and backup segment building. This leads to

the selection of a solution with high backup cost which satisfies the segment length constraints. As a result the overall backup overhead increases. Indeed, in LARGE-8 as shown in Fig. 6.10, backup overhead increases from around 0.55 when $l^{\mathrm{W}} = 5$ to around 0.8 when l^{W} reduces to 3. An increment is also found with LARGE-5 in Fig. 6.11, but it is smaller.

Again, too small segment length thresholds make the blocking probability worse. There might be no solution satisfying the required working and backup segment lengths. This is illustrated in Fig. 6.12 and 6.13. The blocking probability increases slightly from $l^{\mathrm{W}} = 5$ to $l^{\mathrm{W}} = 3$ in the case of LARGE-8 and even more in the case of LARGE-5. In both topologies, at $l^{\mathrm{W}} = 3$, the blocking probabilities raise up drastically when the backup segment length threshold reduces to $l^{\mathrm{B}} = 4$. Smaller impact is observed with $l^{\mathrm{W}} = 5$ in LARGE-5 and nearly invisible in LARGE-8 because the thresholds are nevertheless large enough to provide a reasonable number of segment choices.

6.7 Conclusion

In this paper, we have presented a two-step routing solution for OSSP in multi-domain networks. The solution is scalable for multi-domain networks thanks to the use of Topology Aggregation. A greedy and a dynamic programming algorithms, GROS and DYPOS, with and without Blocking-go-back option are also proposed for the *inter-domain step*. The comparison with optimal single-domain solution shows the efficiency of GROS and DYPOS. Other experiments illustrate that GROS and DYPOS promote well the backup bandwidth sharing. They also show the advantage of the Blocking-go-back phase in reducing the blocking probability.

As the working and backup segments are restricted in length, the proposed solutions guarantee fast recovery. The experiment results show that these solutions can offer a faster recovery than Shared Path Protection as well as the other OSSP solution without segment length restriction.

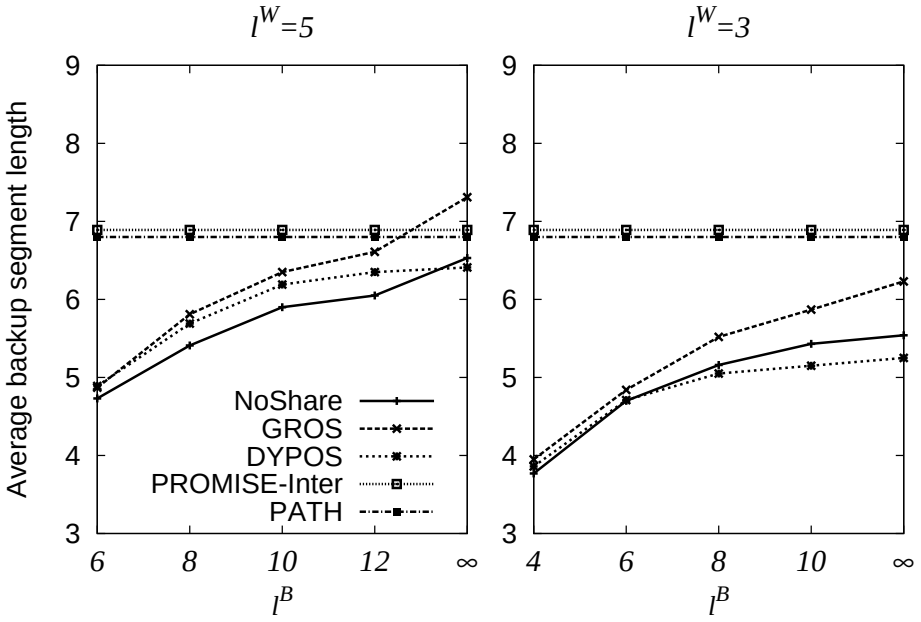

Figure 6.18: Average backup segment lengths in LARGE-5

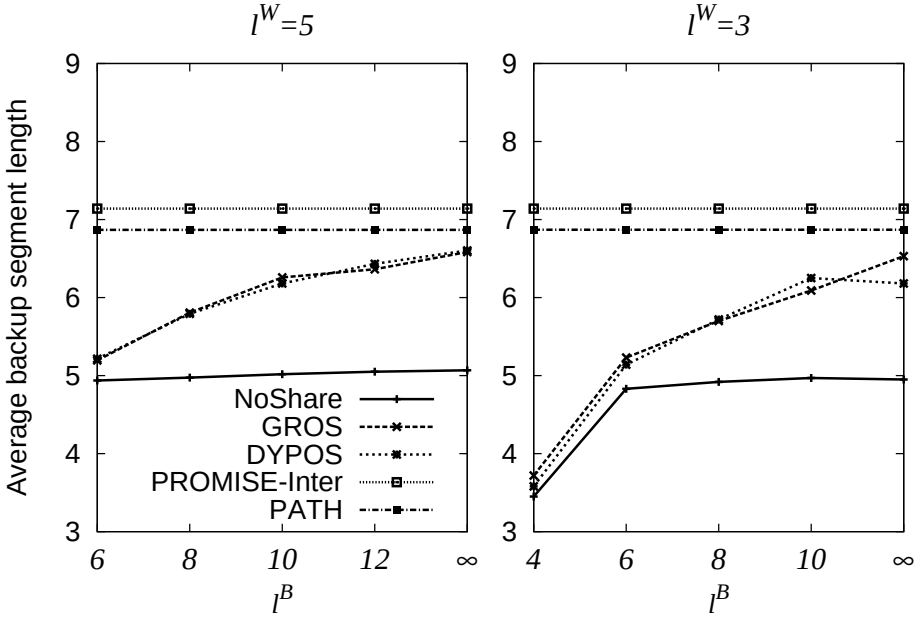

Figure 6.19: Average backup segment lengths in LARGE-8

Obviously, the smaller the segment lengths are, the shorter the recovery is. However, the experiment results show that segment length thresholds should be set carefully as too small thresholds may entail a significant increase of the blocking probability as well as of the backup overhead.

Acknowledgment

The work of the second author has been supported by the NSERC (Natural Sciences and Engineering Research Council of Canada) grant GP0036426 and the Concordia Research Chair on Optimization of Communication Networks.

CHAPITRE 7

A NOVEL APPROACH FOR OVERLAPPING SEGMENT SHARED PROTECTION IN MULTI-DOMAIN NETWORKS

Dieu-Linh Truong, and Brigitte Jaumard

Abstract: Routing for Overlapping Segment Shared Protection (OSSP) in multi-domain networks is more difficult than that in single domain networks because of the scalability requirements. We propose a novel approach for OSSP routing in multi-domain networks where the underlying idea is the prior identification of potential intra-domain paths (or intra-paths for short) for carrying working and backup traffic between domain border nodes. The intra-path identification comes from the solution of a complex model favoring bandwidth saving. The resulting intra-paths are next viewed as virtual edges and used to reduce the multi-domain network to a simpler aggregated network. Routing is performed in this aggregated network at the intra-path level without going down to the physical links. The novel approach performs an accurate and highly scalable routing thanks to the prior identification of the potential intra-paths and the introduction of the maximal share risk group feature. The experiments show that the quality of the proposed approach is close to the optimal one in single-domain networks and outperforms the previously proposed scalable solutions in multi-domain networks.

Keywords: Multi-domain networks, Protection, Dynamic Routing.

7.1 Introduction

Many studies have been published for path and segment protections in single-domain as well as in multi-domain networks with bandwidth guaranteed connec-

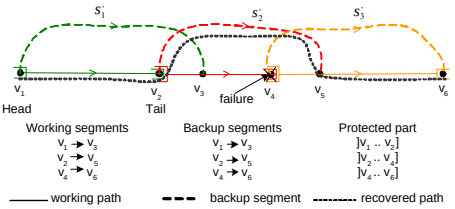

Figure 7.1: Example of Overlapping Segment Protection when v_4 fails. The protected part $]v_2..v_4]$ contains all links and nodes between v_2 exclusively and v_4 inclusively, thus v_4 is recovered using segment s'_2.

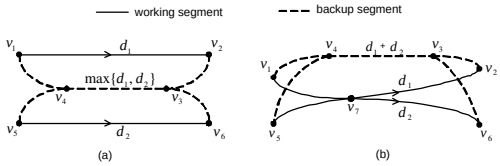

Figure 7.2: Examples of cases where two backup segments can share backup bandwidth (a) and cannot (b).

tions. Overlapping Segment Shared Protection (OSSP) is, however, new and has not received a lot of interest, mostly in the context of multi-domain networks. This paper addresses the routing problem for OSSP in multi-domain networks. Before describing this problem, we briefly recall the OSSP concept.

7.1.1 OSSP concept

In classical segment protection, an end-to-end working path is divided into concatenated segments and each one is protected by a unique backup segment. Upon a single link or node failure, only the failed working segment is replaced by its backup segment. As a result, segment protection offers a faster recovery than path protection. However, segment end nodes are not protected as the failures at those nodes impair both the working and backup segments. Overlapping Segment Protection, first proposed by [HM02] and [RKM02], overcomes this weakness thanks to the overlapping between working segments (see Fig. 7.1) while still inheriting the fast recovery of segment protection.

Shared protection has been proposed for link, path and segment protections [Ram99] in order to save backup bandwidth. In segment protection, in order to guarantee 100% recovery of any single link or node failure, two backup segments can share bandwidth if and only if their working segments are link and node-disjoint.

This condition is called *segment sharing condition*, see Fig. 7.2 for an illustration. In case (a), the working segment from v_1 to v_2, with requested bandwidth d_1, and the working segment from v_5 to v_6, with requested bandwidth d_2, are link and node disjoint. Their backup segments can share bandwidth over the common link (v_4, v_3) and the needed bandwidth on this link is $\max\{d_1, d_2\}$ in order to ensure protection for both working paths. In case (b), the two working segments share node v_7, their backup segments cannot share backup bandwidth. The needed backup bandwidth on link (v_4, v_3) is $d_1 + d_2$, which is greater than in case (a).

With the backup bandwidth sharing feature, Overlapping Segment Protection becomes Overlapping Segment Shared Protection (OSSP). Shared protection under static traffic has received a lot of interest. Several efficient solutions have been proposed, especially the well-known p-cycle initially introduced in [GS98] and further developed for segment protection in [SG03], [SG04]. However, network traffic todays changes dynamically, static traffic is no longer an appropriate assumption except for planning. For this reason, we focus only on dynamic traffic.

For a given new incoming request, the dynamic routing problem for OSSP consists of establishing a working path and associated backup segments for it, while minimizing the bandwidth they use. This routing should be done without any forecast on upcoming requests. The amount of backup bandwidth to reserve for a backup segment depends on the working segment to be protected as well as on the existing working and backup segments. This dependency makes the problem complex. Indeed, it is even more combinatorial than the Shared Path Protection problem which is already NP-hard [LMDL92]. An optimal solution proposed in [HTC04] requires a huge computational effort even for small networks. Several heuristics with smaller computational effort such as the work in [RKM02], SLSP-O in [HM03], CDR in [HM02], PROMISE in [XXQ02] or recursive shared segment protection in [CGYL07] have been proposed. The first study ignores the sharing possibility during the routing. The other studies as well as the one in [HTC04] require detailed information on bandwidth allocation on each network link for their complex bandwidth cost computations. Such requirements can be satisfied only in

single domain networks. Therefore, we refer to all these solutions by single-domain solutions.

7.1.2 State of the art of OSSP in multi-domain networks

OSSP for multi-domain networks is much more complex than that for single domain networks due to the network characteristics and size. A multi-domain network is made of the interconnection of several single-domain networks [BSO02], see Fig. 7.3a. In order to satisfy the *scalability requirements*, only the aggregated routing information can be exchanged amongst domains [LRVB04] by an Exterior Gateway Protocol (EGP) such as BGP. Consequently, a given node is neither aware of the global multi-domain network topology nor of the detailed bandwidth allocation on each network link. However, the complete routing information is still available within each domain thanks to more frequent routing information updates performed by an Interior Gateway Protocol (IGP) such as OSPF.

The *scalability requirement* makes the OSSP routing in multi-domain networks more difficult as working and backup segments go through different domains while the detailed knowledge on bandwidth allocation in physical links of each domain is unavailable at a central node.

Few solutions have been proposed for multi-domain networks, and they all suffer from some drawbacks. In [ASL+02] and [OMZ01], border nodes are not protected because working paths are divided into non-overlapping segments whose endpoints are border nodes. In [MKAM04] the authors consider an unusual multi-domain network without any transit domain. In practice, a connection between distant domains often goes through one or more transit domains leading to a more complex routing problem than the one studied in [MKAM04].

This paper aims at solving the OSSP routing problem in multi-domain networks with the objective of minimizing the total working and backup bandwidth capacity required by a new incoming request under a dynamic traffic pattern. We address this problem for networks with bandwidth guaranteed connections such as SONET/SDH, MPLS-TE, ATM and DWDM networks with wavelength conversion

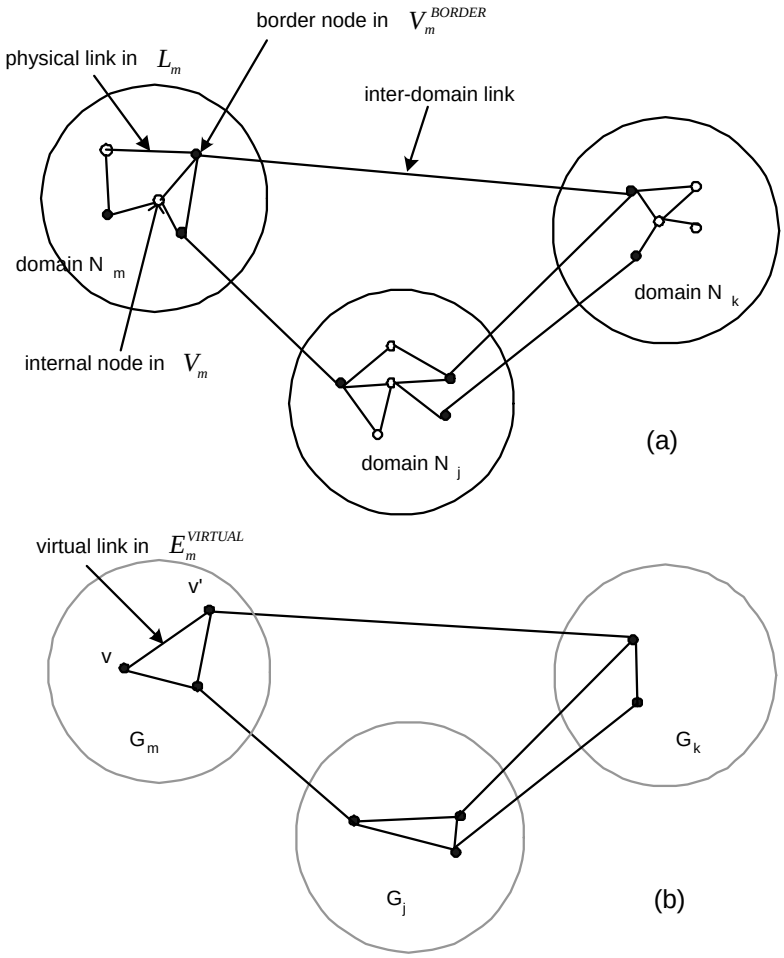

Figure 7.3: A multi-domain network (a) and its *inter-domain network* (b) obtained from Topology Aggregation.

capability.

In [TJ07b] we have proposed some solutions for OSSP in multi-domain networks. In these solutions, each domain network is topologically aggregated. A domain $N_m = (V_m, L_m)$, where V_m and L_m are the sets of nodes and links, becomes an aggregated graph $G_m = (V_m^{\text{BORDER}}, E_m^{\text{VIRTUAL}})$ composed of a border node set V_m^{BORDER} and a virtual link set E_m^{VIRTUAL} (see Fig. 7.3b). A virtual link represents the possibility of going from one border node to another one through intra-paths. The multi-domain network becomes a so called inter-domain network. A rough routing is determined first in the inter-domain network based on approximations of the working and backup costs of the virtual links. The resulting working and backup segments are the paths of virtual and inter-domain links. These virtual links are then mapped to intra-paths. Several algorithms have been proposed for the routing and the mapping, resulting in two solutions GROS and DYPOS. We referred to both of them by "Route-and-Map" approach, denoted by RaM.

Although RaM offers good routing results and scalability, we propose here to eliminate the approximation in RaM by reversing the order of the Routing and the Mapping phases. The resulting approach is called "Map-and-Route" or MaR for short. Each virtual link is mapped to several intra-paths whose working and backup costs can be computed exactly. The routing is performed on a network where links are associated with the selected intra-paths and the link costs are thus exact.

This paper is organized as follows. The next section provides the general ideas of the new approach. Section 7.3 describes the Routing solution and Section 7.4 discusses its scalability. Section 7.5 states the Mapping sub-problem in each domain and its exact and heuristic solutions are presented in Sections 7.6 and 7.7 respectively. Periodic Mapping refreshing is discussed in Section 7.8. The experimental results are shown and discussed in Section 7.9. Conclusions follow in Section 7.10. A simplified model as well as a path-based formulation for the Mapping sub-problem are also presented in the Appendix.

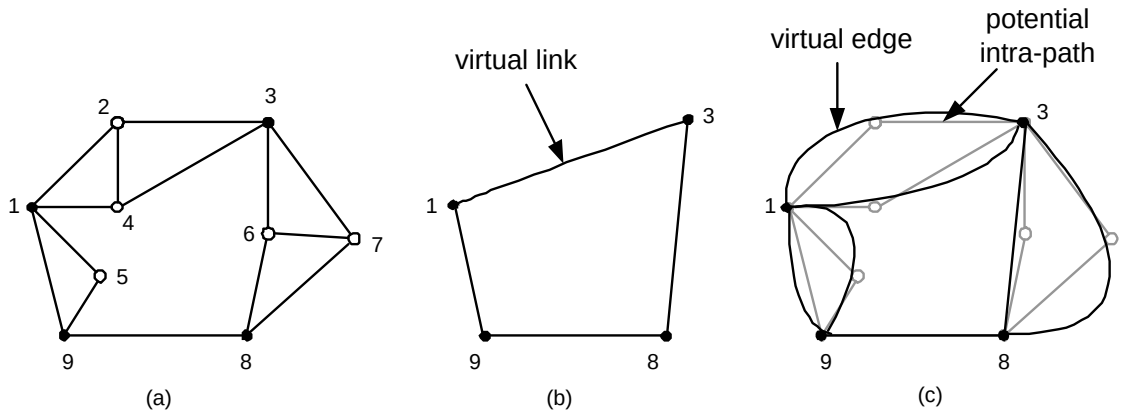

Figure 7.4: (a) An original domain at the intra-domain level; (b) the aggregated domain at the inter-domain level; (c) the mapped domain at the mapped level with a maximum of 2 intra-paths/virtual link for both working and backup traffic.

7.2 A Map and Route approach

In MaR, each domain network N_m is first aggregated as in RaM. Next, for each virtual link e of N_m, a set of potential intra-paths $\mathcal{P}_e^{\mathrm{W}}$ (resp. $\mathcal{P}_e^{\mathrm{B}}$ with possibly $\mathcal{P}_e^{\mathrm{B}} = \mathcal{P}_e^{\mathrm{W}}$), will be selected for carrying the working (resp. backup) traffic between the border nodes of e (see Fig. 7.4 where border nodes correspond to filled black points while internal nodes correspond to white points, as in all subsequent figures). The intra-paths carrying working (resp. backup) traffic are referred as working (resp. backup) intra-paths. An intra-path can be viewed as a lightpath between an ingress and an egress node of a domain. The larger $\mathcal{P}_e^{\mathrm{W}}$ and $\mathcal{P}_e^{\mathrm{B}}$ are, the larger are the choices for intra-paths at the expense of a reduced scalability.

Definition 1. *A direct intra-path is an intra-path that does not go through any intermediate border node.*

All potential intra-paths to be selected for a virtual link must be direct intra-paths. This requirement does not reduce the choice of intra-paths as a non direct intra-path can be represented by multiple direct intra-paths. A connection may take several direct intra-paths to go through a domain.

In order to work at the level of intra-paths, each selected potential intra-path

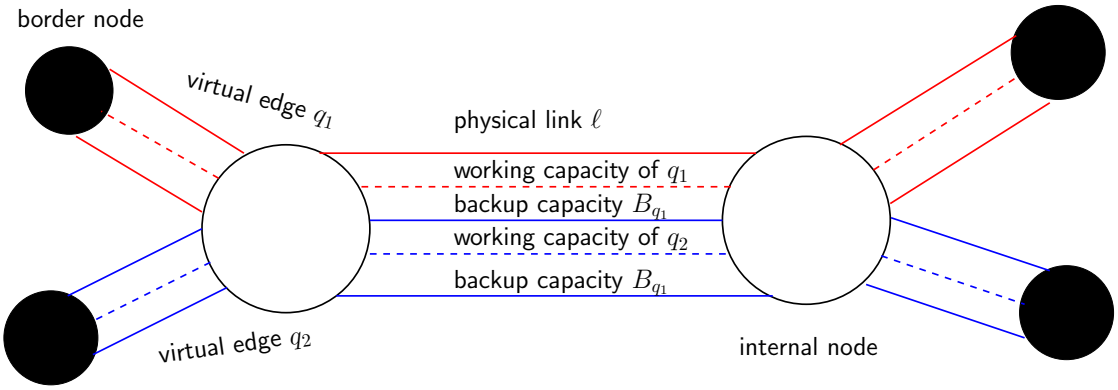

Figure 7.5: Example of two virtual edges that share physical link. Their B_{q_1}, B_{q_2} differs from the total backup capacity of link ℓ. The free capacities are not shown.

will be abstracted and handled as a single link called "virtual edge" (see Fig. 7.4c). Since a virtual edge is in one-to-one correspondence with an intra-path, we use the two terms alternatively depending on whether we are dealing with the abstracted (mapped) or detailed (intra-domain) level. Each virtual edge has its own working and backup capacities (see Fig. 7.5 for an illustration). The working capacity of a virtual edge is defined as the bandwidth occupied by the working paths routed along the entire intra-path associated with the virtual edge. Similarly, the backup capacity of a virtual edge is defined as the bandwidth occupied by the backup segments routed along the entire intra-path associated with the virtual edge. Although the virtual edges between different nodes may use a common physical link, they manage separately their working and backup capacities on the common physical link. Therefore, the working (resp. backup) capacity of a physical link is the sum of, the working (resp. backup) capacities of the virtual links using it.

We also introduce the following new sharing rule in which an intra-path is again handled as a single entity (see an illustration in Fig. 7.6).

> *Sharing rule*: Two backup segments are allowed to share bandwidth only if they go through an **identical intra-path**.

In other words, two backup segments either share bandwidth along their entire common intra-path or share no bandwidth. Backup segments over two intra-paths

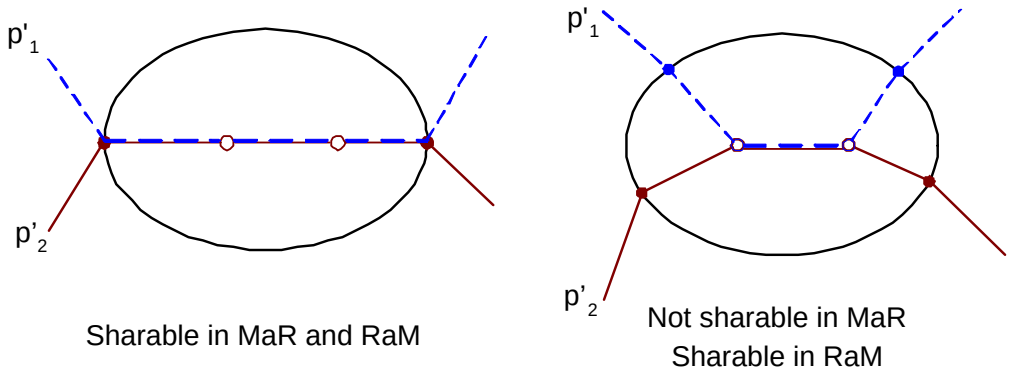

Figure 7.6: The cases where two backup segments p'_1, p'_2 can share and cannot share backup bandwidth under MaR and RaM.

that differ by at least one link are not allowed to share bandwidth.

Therefore, a virtual edge is really considered as a single link. *The working capacity, backup capacity as well as the backup bandwidth sharing possibility of a virtual edge is identical for every physical link along its corresponding intra-path.* Of course in MaR, some bandwidth sharing possibility is omitted on some particular physical links. However, the advantage is that we do not have to go down to the level of the physical links in order to identify some sharable bandwidth for protecting an intra-path, which would impair the scalability. Note that in RaM, in order to avoid going down to the physical link level, approximations are made on sharable bandwidth computations.

The OSSP routing problem is divided into 2 sub-problems:

- Mapping sub-problem: Potential intra-path sets $\mathcal{P}_e^{\mathrm{W}}, \mathcal{P}_e^{\mathrm{B}}$ are selected for each virtual link e of each domain. Each intra-path is then abstracted as a single "virtual edge". The multi-domain network resulting from this abstraction is called "mapped network". The Mapping is performed once for a long term use and should be refreshed for updating the potential intra-paths only when the current ones are saturated.

- Routing sub-problem: The working path and backup segments are computed in the mapped network. Unlike RaM, there is no need to go down to the

intra-domain level for identifying the intra-paths within each domain as they are in one-to-one correspondence with virtual edges.

Both sub-problems will be discussed in detail in the next sections.

7.3 Routing sub-problem

The objective of the Routing sub-problem is, for a new incoming request, to find a working path and a set of backup segments so that their total bandwidth cost is minimized. Let d be the requested bandwidth. Before detailing the analytical expression of the total bandwidth cost of the incoming request, we introduce some further notations and define the cost functions of virtual edges.

Let q (resp. q') be an intra-path/virtual edge that is considered for a working segment (resp. backup segment) for the new incoming request. Let us assume that each bandwidth unit on a physical link has a unit cost and the bandwidth cost of a segment is the sum of the bandwidth costs of its links.

c_ℓ^{res} residual capacity of physical link ℓ.

$B_{q'}$ backup bandwidth reserved by backup segments going through the entire virtual edge q'. Be aware that $B_{q'}$ may differ from the total backup bandwidth reserved on a physical link of q'. (see an example in Fig. 7.5).

$B_{q'}^v$ sum of requested bandwidth for the connections of which a backup segment goes through q' and the corresponding working segment goes through node v. Those backup segments cannot share bandwidth amongst them because they will be activated simultaneously when v fails. Their backup bandwidth is not profitable for a backup segment of the new incoming request if this segment goes through q' and protects a working segment going through v.

$B_{q'}^q$ sum of requested bandwidth for the connections of which a backup segment goes through q' and the corresponding working segment goes through q.

$\delta^q_{q'}$ disjointness between two intra-paths q, q'. It is set to 1 if q and q' are node disjoint and 0 otherwise.

α_q total working bandwidth cost of the new incoming request on virtual edge q.

$\beta^q_{q'}$ backup bandwidth cost of the new incoming request on virtual edge q' for protecting virtual edge q against any single link or node failure. Note that $\beta^q_{q'}$ is the additional bandwidth that the new incoming request needs on q' excluding the fraction of sharable backup bandwidth it can benefit from $B_{q'}$.

π working path of the new incoming request in the mapped network. It is a path made of virtual edges.

π_i working segment of the new incoming request indexed by i in the mapped network. It is a path made of virtual edges.

π'_i backup segment of the working segment π_i in the mapped network. It is a path made of virtual edges.

I set of segment indices of the new incoming request.

$\beta^{\pi_i}_{q'}$ backup bandwidth cost of the new incoming request on virtual edge q' for protecting working segment π_i against a single failure on any node or link.

$\|q\|$ length of intra-path q in terms of the number of hops.

γ_q residual capacity of virtual edge q.

Definition 2. *The residual capacity of a virtual edge is the maximum bandwidth that we can route along it.*

Theorem 1. *The residual capacity of virtual edge q is the minimum of the residual capacity on each of its physical links:*

$$\gamma_q = \min_{\ell \in q} c^{\text{res}}_\ell. \tag{7.1}$$

Proof. Indeed, the smallest residual capacity on q is $\min_{\ell \in q} c_\ell^{\text{res}}$. We can always route an amount of bandwidth $\min_{\ell \in q} c_\ell^{\text{res}}$ over q because every link along q has at least this amount of available capacity. On the other hand, there is at least one link of q whose residual capacity is $\min_{\ell \in q} c_\ell^{\text{res}}$, thus we cannot route more than this bandwidth over q. As a result, $\gamma_q = \min_{\ell \in q} c_\ell^{\text{res}}$ □

We are now going to identify the working and backup costs of virtual edges. Unlike RaM, all costs will be computed exactly in MaR. Since working segments do not share any bandwidth, each working segment uses exactly bandwidth d on each of its physical links. The working cost of the new incoming request on virtual edge q amounts to:

$$\alpha_q = \begin{cases} \|q\| \times d & \text{if } d \leq \gamma_q \\ \infty & \text{otherwise.} \end{cases} \tag{7.2}$$

Theorem 2. *If the bandwidth reserved on a backup segment is sufficient for protecting its working segment against a failure on a node, it is also sufficient for protecting the same working segment against a failure on a link adjacent to the node.*

Proof. Let us denote the backup segment by p' and the working segment by p. Let v be a node of p. Let e be an adjacent link of v. Let $\mathbb{P}^v_{p'}$ and $\mathbb{P}^e_{p'}$ be respectively the set of working segments going through v and e such that their backup segments go through p'. Hence, $\mathbb{P}^e_{p'} \subseteq \mathbb{P}^v_{p'}$. Assume that the backup bandwidth reserved on p' is sufficient against a failure at node v, it means that this bandwidth is sufficient for activating simultaneously all backup segments in $\mathbb{P}^v_{p'}$. The backup bandwidth is thus sufficient for activating simultaneously all backup segments in $\mathbb{P}^e_{p'}$. In other words, it is sufficient for covering the failure at e. □

This theorem shows that in order to protect a working segment against failures on nodes and links, we only need to protect nodes and then links will be automatically protected.

Theorem 3. *The backup cost of the new incoming request on virtual edge q' for protecting a virtual edge q against any single link or node failure is:*

$$\beta_{q'}^{q} = \begin{cases} \|q'\| \times (\max\limits_{v \in q} B_{q'}^{v} + d - B_{q'}) & \text{if } \delta_{q'}^{q} = 1 \text{ and} \\ & 0 \leq \max\limits_{v \in q} B_{q'}^{v} + d - B_{q'} \leq \gamma_{q'} \\ 0 & \text{if } B_{q'} - \max\limits_{v \in q} B_{q'}^{v} \geq d \\ \infty & \text{otherwise.} \end{cases} \tag{7.3}$$

Proof. When q, q' are not disjoint, i.e., $\delta_{q'}^{q} = 0$, they both fail upon a single failure at a common link or node, therefore $\beta_{q'}^{q} = \infty$. Otherwise, let us consider the backup bandwidth needed by the new incoming request on a physical link of q' in order to protect q against a failure on node $v \in q$. Within the existing backup bandwidth $B_{q'}$ on q', $B_{q'}^{v}$ is non sharable for covering a failure at v. The remaining bandwidth is sharable and amounts to $B_{q'} - B_{q'}^{v}$ for every physical link of q'. Thus, the additional bandwidth that the new incoming request needs on each physical link of q' for protecting v is: $B_{q'}^{v} + d - B_{q'}$. In the single failure context, only one node can fail at a time. Hence, the additional backup bandwidth needed on a physical link of q' for protecting q against any single failure is: $\max\limits_{v \in q}(B_{q'}^{v} + d - B_{q'})$. As the cost of q' is proportional to its length, we deduce the formula (7.3). □

From the backup cost $\beta_{q'}^{q}$, we deduce the backup cost $\beta_{q'}^{\pi_i}$ of virtual edge q' for protecting working segment π_i against any single link or node failure:

$$\beta_{q'}^{\pi_i} = \max_{q \in \pi_i} \beta_{q'}^{q}. \tag{7.4}$$

Therefore, the Routing sub-problem is defined formally as finding a working path π and a set of backup segments $\{\pi'_i, i \in I\}$ so that the total bandwidth cost is minimized, i.e.,:

$$\min \sum_{q \in \pi} \alpha_q + \sum_{\pi'_i, i \in I} \sum_{q' \in \pi'_i} \beta_{q'}^{\pi_i}.$$

It is equivalent to the single domain OSSP routing problem described in [HM02], [HM03], [HTC04] and [XXQ02] where virtual edges are considered as links of a single domain network. Single domain OSSP routing solutions as well as the inter-domain routing solutions GROS and DYPOS proposed in [TJ07b] can be used to solve it by applying them on the mapped network. In the experiments presented in this paper, DYPOS is used for the Routing step.

7.3.1 An exact and scalable solution for computing the backup cost of a virtual edge

The computation of backup cost $\beta^q_{q'}$ as expressed in (7.3) requires the knowledge of $B^v_{q'}$ for each node v and intra-path q'. It is an intra-domain information which changes dynamically after each routing. Therefore, maintaining all $B^v_{q'}$ up-to-date is a non-scalable requirement.

We propose a more scalable method for computing $\beta^q_{q'}$ based on maximal shared risk groups. The idea is as follows. In each domain, there exist some critical nodes which belong to many intra-paths. The protection of these nodes can be sufficient for protecting some other nodes (see Th. 4 below). Therefore, the backup cost for protecting an intra-path can be deduced from the backup cost for protecting some given critical nodes.

Definition 3. *For a given domain, the Share Risk Group (*SRG*) of a node v, denoted by* $\mathrm{SRG}(v)$*, is the set of virtual edges that share the same risk at v in the domain. It corresponds to the set of intra-paths going through v.*

SRGs have the following characteristic:

Theorem 4. *Let* $\mathrm{SRG}(v_i)$ *and* $\mathrm{SRG}(v_j)$ *be two SRGs so that* $\mathrm{SRG}(v_i) \subseteq \mathrm{SRG}(v_j)$*, then*

$$B^{v_i}_{q'} \leq B^{v_j}_{q'}. \tag{7.5}$$

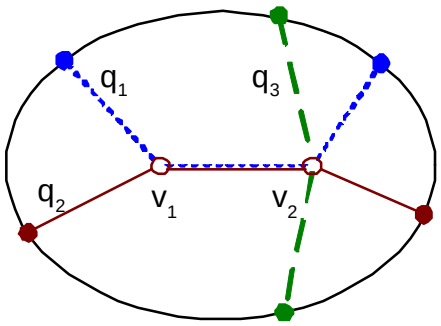

Figure 7.7: $\mathrm{SRG}(v_1) = \{q_1, q_2\} \subset \mathrm{SRG}(v_2) = \{q_1, q_2, q_3\}$ because all intra-paths going through v_1 are going through v_2.

In other words, if all the intra-paths that go through one node (v_i), go also through another node (v_j), the backup bandwidth reserved on the intra-path for protecting the first node (v_i) does not exceed the backup bandwidth reserved on the same intra-path for protecting the second node (v_j), see Fig. 7.7 for an illustration.

Proof. Since $\mathrm{SRG}(v_i) \subseteq \mathrm{SRG}(v_j)$ then:

$$\mathrm{SRG}(v_j) = \mathrm{SRG}(v_i) \cup \Big(\mathrm{SRG}(v_j) \setminus \mathrm{SRG}(v_i) \Big).$$

Thus:

$$\sum_{q \in \mathrm{SRG}(v_j)} B^q_{q'} = \sum_{q \in \mathrm{SRG}(v_i)} B^q_{q'} + \sum_{q \in \mathrm{SRG}(v_j) \setminus \mathrm{SRG}(v_i)} B^q_{q'}.$$

Consequently,

$$\sum_{q \in \mathrm{SRG}(v_j)} B^q_{q'} \geq \sum_{q \in \mathrm{SRG}(v_i)} B^q_{q'}.$$

From the definition of $B^v_{q'}$, we have $B^v_{q'} = \sum_{q \in \mathrm{SRG}(v)} B^q_{q'}$. Thus,

$$B^{v_j}_{q'} \geq B^{v_i}_{q'}.$$

□

Definition 4. *A SRG is maximal if it is not contained in another SRG.*

If two SRGs of two different nodes are identical and maximal, one node will be chosen as the representative node for the maximal SRG. If one of the two nodes is a border node, it will be chosen. When both nodes are internal nodes, we can choose any of them.

From Th. 4 we deduce the following theorem.

Theorem 5. *Let q be a sub-path and v_j be a node on q such that* $\mathrm{SRG}(v_j)$ *is maximal and denoted by* $\mathrm{SRG}^{\mathrm{MAX}}(v_j)$. *We have:*

$$\max_{v \in q} B^{v}_{q'} = \max_{v_j : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_j)} B^{v_j}_{q'}. \tag{7.6}$$

Proof. First note that $v_j : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_j)$ is the formal expression of the fact that v_j is in q and $\mathrm{SRG}(v_j)$ is maximal. The proof is now as follows.

On the one hand, since v_j is in q, set $\{v_j : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_j)\}$ is a subset of set $\{v : v \in q\}$. Therefore,

$$\max_{v \in q} B^{v}_{q'} \geq \max_{v_j : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_j)} B^{v_j}_{q'}. \tag{7.7}$$

On the other hand, for all $v \in q$, there exists a maximal SRG that contains or is equal to $\mathrm{SRG}(v)$. Let $\mathrm{SRG}^{\mathrm{MAX}}(v_k)$ be such a SRG, then $\mathrm{SRG}(v) \subseteq \mathrm{SRG}^{\mathrm{MAX}}(v_k)$. According to Th. 4, $B^{v}_{q'} \leq B^{v_k}_{q'}$. In addition, since $q \in \mathrm{SRG}(v)$ then $q \in \mathrm{SRG}^{\mathrm{MAX}}(v_k)$ and thus v_k is in q. In brief, for each $v \in q$ there exists $v_k : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_k)$ such that and $B^{v}_{q'} \leq B^{v_k}_{q'}$. Hence,

$$\max_{v \in q} B^{v}_{q'} \leq \max_{v_k : q \in \mathrm{SRG}^{\mathrm{MAX}}(v_k)} B^{v_k}_{q'}. \tag{7.8}$$

Inegalities (7.7) and (7.8) lead to (7.6). □

Th.5 provides a way to compute $\max_{v \in q} B^{v}_{q'}$. Let v_1, v_2 be two border nodes of

virtual edge q and N_m be the domain containing q. Then:

$$\max_{v \in q} B^v_{q'} = \max \left\{ B^{v_1}_{q'}, B^{v_2}_{q'}, \max_{\substack{v_j \in V_m \setminus V_m^{\text{BORDER}}: \\ q \in \text{SRG}^{\text{MAX}}(v_j)}} B^{v_j}_{q'} \right\}. \tag{7.9}$$

The third term of which we take the maximum in (7.9) relates only to non-border nodes $v_j \in V_m \setminus V_m^{\text{BORDER}}$ while the first and second terms take care of border nodes v_1, v_2. We do not need to consider the other border nodes since q is a direct intra-path. Note that $B^{v_1}_{q'}$ and $B^{v_2}_{q'}$ can be easily identified by looking at the virtual edges going through v_1 and v_2.

Equation (7.9) holds also when q is an inter-domain link as we then have $V_m \setminus V_m^{\text{BORDER}} = \emptyset$.

In substituting $\max_{v \in q} B^v_{q'}$ in (7.3) by the right hand-side of (7.9), we obtain the following scalable formula for computing the backup cost $\beta^q_{q'}$:

$$\beta^q_{q'} = \begin{cases} \|q'\| \times \left(\max\{B^{v_1}_{q'}, B^{v_2}_{q'}, \max\limits_{\substack{v_j \in V_m \setminus V_m^{\text{BORDER}}: \\ q \in \text{SRG}^{\text{MAX}}(v_j)}} B^{v_j}_{q'}\} + d - B_{q'} \right) \\ \qquad \text{if } \delta^p_{p'} = 1 \text{ and} \\ \qquad\qquad 0 \leq \max\{B^{v_1}_{q'}, B^{v_2}_{q'}, \max\limits_{\substack{v_j \in V_m \setminus V_m^{\text{BORDER}}: \\ q \in \text{SRG}^{\text{MAX}}(v_j)}} B^{v_j}_{q'}\} + d - B_{q'} \leq \gamma_{q'} \\ 0 \qquad \text{if } B_{q'} - \max\{B^{v_1}_{q'}, B^{v_2}_{q'}, \max\limits_{\substack{v_j \in V_m \setminus V_m^{\text{BORDER}}: \\ q \in \text{SRG}^{\text{MAX}}(v_j)}} B^{v_j}_{q'}\} \geq d \\ \infty \qquad \text{otherwise.} \end{cases} \tag{7.10}$$

In conclusion, in order to identify the backup cost of an intra-path for protecting another intra-path, we only need to compute the backup cost for protecting some non-border maximal SRG nodes and those for protecting border nodes of the intra-paths.

7.4 Scalability discussion

Since the Mapping step is performed independently within each domain, it does not encounter any scalability difficulty. Let us discuss the scalability issue in the Routing process.

When a new request comes in, the source border node is responsible for identifying working and backup segments by performing the Routing process. First of all, it needs to compute the working and backup costs associated with each virtual edge. According to (7.2), the required parameters for computing the working cost of each virtual edge q are: $\|q\|, \gamma_q$. According to (7.3) and (7.9), the required parameters for computing the backup cost of virtual edge q (q' in the formula) in order to protect all the other virtual edges q^{W} (q in the formula) are: $\|q\|, B_q^v$ for all $v \in V^{\mathrm{BORDER}}$, B_q, γ_q, B_q^v for all non-border nodes v whose SRG is maximal.

Let E^{VEDGE} be the set of virtual edges in the mapped network. Each border node should store the parameters that we categorize as follows:

Cat.A : $\|q\|, B_q$ for all $q \in E^{\mathrm{VEDGE}}$;

Cat.B : B_q^v for all $v \in V^{\mathrm{BORDER}}$ and for all $q \in E^{\mathrm{VEDGE}}$;

Cat.C : All non-border $\mathrm{SRG}^{\mathrm{MAX}}$ in every domain as well as their associated internal nodes v;

Cat.D : B_q^v for all internal nodes v that are associated with the non-border $\mathrm{SRG}^{\mathrm{MAX}}$ of Cat.C, and for all $q \in E^{\mathrm{VEDGE}}$;

Cat.E : γ_q for all $q \in E^{\mathrm{VEDGE}}$.

These parameters should be kept up-to-date and exchanged between border nodes, by using an extension of BGP in order to adapt with the newly introduced parameters. These exchanges provide the border nodes with the identical information for working and backup cost computing.

In Cat.A, values $\|q\|$, for all $q \in E^{\mathrm{VEDGE}}$ do not need to be updated because they are constant unless the network topology changes. The parameter B_q for all

$q \in E^{\text{VEDGE}}$ is easily managed in the mapped network by increasing or decreasing it by $\beta_q^{\pi_i}/\|q\|$ at each setting up or tearing down of a given request. In Cat.B, B_q^v for all $v \in V^{\text{BORDER}}$, $q \in E^{\text{VEDGE}}$ can also be managed in a similar way except that the increasing and reduction are equal to d. In Cat.C, the non-border maximal SRGs depend uniquely on the Mapping step and are therefore stable. The experimental results in Section 7.9.2 will show that the number of non-border SRG^{MAX} is quite small, therefore the number of B_q^v in Cat.D is also small.

The values of B_q^v in Cat.B and Cat.D for a given virtual edge q can be stored at a border node of q. The values of B_q^v are required only for the computations of backup segments whose working path is previously identified in most routing algorithms used for the Routing step. In a backup segment computation, we can collect B_q^v for all v in the working path by sending a signaling message, using, e.g, RSVP-PATH, along the working path and get those B_q^v back with returning message, using, e.g., RSVP-RESV.

It now remains the values γ_q of Cat.E which need to be kept track for each virtual edge of E^{VEDGE}. While the residual capacity on every physical link that participates in q is not smaller than the maximal requested bandwidth, the residual capacity γ_q of virtual edge q is sufficient for any new request and does not need to be updated. Otherwise, γ_q needs to be recalculated exactly by using (7.1). This operation impairs the most the scalability of our solution. Luckily we have to perform it rarely, only when the network is nearly saturated.

In summary, most of the information required in the routing of MaR is per virtual edge and is managed at the mapped level. The quantity of required internal domain information is small. The scalability is thus preserved.

7.5 Mapping sub-problem

The Mapping sub-problem as briefly described in Section 7.2 consists of identifying a set of potential working intra-paths and a set of potential backup intra-paths for each virtual link. Such a Mapping is performed independently in each domain.

The Mapping sub-problem for domain N_m is stated as follows.
Given:

- n^{W} and n^{B} the maximum numbers of potential working and backup intra-paths needed for each virtual link of E_m^{VIRTUAL};
- n_e the number of *direct* intra-paths associated with e.

Let $n_e^{\mathrm{W}} = \min\{n_e, n^{\mathrm{W}}\}$ and $n_e^{\mathrm{B}} = \min\{n_e, n^{\mathrm{B}}\}$. They correspond to the exact number of potential working and backup intra-paths for each $e \in E_m^{\mathrm{VIRTUAL}}$. We need to identify:

- $\mathcal{P}_e^{\mathrm{W}} = \{q_{e,i}^{\mathrm{W}}, i = 1..n_e^{\mathrm{W}}\}$, the set of potential working intra-paths of e;
- $\mathcal{P}_e^{\mathrm{B}} = \{q_{e,i}^{\mathrm{B}}, i = 1..n_e^{\mathrm{B}}\}$, the set of potential backup intra-paths of e.

As the Routing objective is to minimize the total working and backup cost of each request, in the Mapping, we encourage the intra-paths supporting this objective through the following selection criteria.

Criterion 1. *A selected working intra-path should minimize its working cost while maintaining enough residual bandwidth for allocating future connections. It amounts to balance the network load.*

We can associate each physical link with a weight which is the inverse of the residual capacity of the link. A selected intra-path is then a weighted shortest path. Hence, from the global viewpoint, the set of all potential intra-paths to be selected for a domain should minimize their total weighted length, which leads to:

$$\min \sum_{e \in E_m^{\mathrm{VIRTUAL}}} \sum_{q \in \mathcal{P}_e^{\mathrm{W}}} \sum_{\ell \in q} \frac{1}{c_\ell^{\mathrm{res}}}. \tag{7.11}$$

Criterion 2. *A selected backup intra-path should minimize its backup cost while maintaining enough residual bandwidth for allocating future connections.*

From a global sight, a backup segment uses an homogeneous amount of bandwidth along an intra-path as in a working segment. Hence, this criterion will be similar to Criterion 1, which leads to:

$$\min \sum_{e \in E_m^{\text{VIRTUAL}}} \sum_{q \in \mathcal{P}_e^{\text{B}}} \sum_{\ell \in q} \frac{1}{c_\ell^{\text{res}}}. \tag{7.12}$$

Criterion 3. *The working intra-paths should be selected so as to increase the possibility of finding pairwise disjoint working intra-paths.*

This criterion originates from the fact that backup segments can share bandwidth only if their working segments are disjoint, according to the segment sharing condition. The criterion is interpreted as maximizing the number of pairs of disjoint working intra-paths:

$$\max \sum_{\substack{q_1 \in \mathcal{P}_{e_1}^{\text{W}}, q_2 \in \mathcal{P}_{e_2}^{\text{W}}, \\ e_1, e_2 \in E_m^{\text{VIRTUAL}}}} \delta_{q_1}^{q_2}. \tag{7.13}$$

Criterion 4. *Virtual links should be mapped so that the possibility that a pair of working and backup virtual links is disjoint is maximized.*

Definition 5. *Two virtual links are disjoint iff there exists a potential intra-path of one virtual link that is link and node disjoint with a potential intra-path of the other virtual link.*

The disjointness can be formally stated as follows:

$$\delta_{e'}^{e} = \begin{cases} 1 & \text{if } \exists q \in \mathcal{P}_e^{\text{W}}, q' \in \mathcal{P}_{e'}^{\text{B}} : q \cap q' = \emptyset, \\ 0 & \text{otherwise.} \end{cases} \tag{7.14}$$

Criterion 4 is justified as follows. In the case of lightly inter-connected multi-domain networks, we may need to route a request over two fixed virtual links, one for the working segment and the other one for the backup segment. These two virtual links should have at least one pair of disjoint intra-paths otherwise the considered working and backup segments would have some common links or nodes,

and this would impair the protection. From the global viewpoint, this criterion is interpreted as maximizing the number of pairs of disjoint working and backup virtual links:

$$\max \sum_{e,e' \in E_m^{\text{VIRTUAL}}} \delta_{e'}^{e}. \tag{7.15}$$

When (7.15) provides multiple optimal solutions, we break the ties by maximizing the total number of disjoint working and backup intra-paths, which is similar to Criterion 3:

$$\max \sum_{\substack{q \in \mathcal{P}_e^{\text{W}}, q' \in \mathcal{P}_{e'}^{\text{B}}, \\ e,e' \in E_m^{\text{VIRTUAL}}}} \delta_{q'}^{q}. \tag{7.16}$$

7.5.1 Putting all together

The Mapping sub-problem is a multi criteria optimization problem. For solving it, we put the criteria all together in a single objective by associating different weights for each criterion, we obtain:

$$\begin{aligned} \min\Bigg(\mu_1 \sum_{e \in E_m^{\text{VIRTUAL}}} \sum_{q \in \mathcal{P}_e^{\text{W}}} \sum_{\ell \in q} \frac{1}{c_\ell^{\text{res}}} \quad &+ \mu_2 \sum_{e \in E_m^{\text{VIRTUAL}}} \sum_{q \in \mathcal{P}_e^{\text{B}}} \sum_{\ell \in q} \frac{1}{c_\ell^{\text{res}}} \\ - \mu_3 \sum_{e_1,e_2 \in E_m^{\text{VIRTUAL}}} \sum_{\substack{q_1 \in \mathcal{P}_{e_1}^{\text{W}}, \\ q_2 \in \mathcal{P}_{e_2}^{\text{W}}}} \delta_{q_1}^{q_2} &- \mu_4 \sum_{e,e' \in E_m^{\text{VIRTUAL}}} \delta_{e'}^{e} \\ &- \mu_5 \sum_{e,e' \in E_m^{\text{VIRTUAL}}} \sum_{\substack{q \in \mathcal{P}_e^{\text{W}}, \\ q' \in \mathcal{P}_{e'}^{\text{B}}}} \delta_{q'}^{q} \Bigg). \end{aligned} \tag{7.17}$$

The coefficients should be set carefully in order to define a meaningful objective. In general, μ_1 and μ_2 should be set large enough so that the two first terms, and thus bandwidth saving, are prioritized. The three last terms will help to select the solutions with the most disjoint virtual links and intra-paths. Since working and

backup intra-paths are relatively symmetrical in the Mapping, we can set $\mu_1 = \mu_2$ and $\mu_3 = \mu_5$. The coefficient μ_4 and μ_5 should be chosen so that the fifth term is always smaller than the granularity of the fourth term in order to not act upon the maximization of the fourth term.

7.6 Exact Mapping solution

The Mapping sub-problem is complex as intra-paths need to be found for multiple pairs of border nodes. In addition, intra-paths depend on each other. That is why we use an integer linear program for modeling the optimal Mapping for each domain. Let us consider domain N_m.

Let $x^{e,i}_{(u,v)}$ be the decision variable such that it is equal to 1 if link (u, v) belongs to working intra-path $q^{\mathrm{W}}_{e,i}$, indexed i, of virtual link $e \in E_m^{\mathrm{VIRTUAL}}$ and 0 otherwise, i.e.,:

$$x^{e,i}_{(u,v)} = \begin{cases} 1 & \text{if } (u,v) \in q^{\mathrm{W}}_{e,i} \in \mathcal{P}^{\mathrm{W}}_e \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1..n^{\mathrm{W}}_e, \quad e \in E_m^{\mathrm{VIRTUAL}}. \quad (7.18)$$

Let $y^{e,i}_{(u,v)}$ be the decision variable such that it is equal to 1 if link (u, v) belongs to backup intra-path $q^{\mathrm{B}}_{e,i}$, indexed i, of virtual link $e \in E_m^{\mathrm{VIRTUAL}}$ and 0 otherwise, i.e.:

$$y^{e,i}_{(u,v)} = \begin{cases} 1 & \text{if } (u,v) \in q^{\mathrm{B}}_{e,i}, q^{\mathrm{B}}_{e,i} \in \mathcal{P}^{\mathrm{B}}_e \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1..n^{\mathrm{B}}_e, \quad e \in E_m^{\mathrm{VIRTUAL}}. \quad (7.19)$$

7.6.1 Flow conservation constraint for working intra-paths

Let h^e and t^e be the head and the tail border nodes of $e \in E_m^{\text{VIRTUAL}}$. The flow conservation constraint for the working intra-path $q_{e,i}^{\text{W}} \in \mathcal{P}_e^{\text{W}}$ is:

$$\sum_u x_{(u,v)}^{e,i} - \sum_u x_{(v,u)}^{e,i} = \begin{cases} 1 & \text{if } v = h^e \\ 0 & \text{if } v \neq h^e, t^e \\ -1 & \text{if } v = t^e \end{cases},$$
$$v \in V_m, i = 1..n_e^{\text{W}}, e \in E_m^{\text{VIRTUAL}}. \quad (7.20)$$

In order to guarantee that $q_{e,i}^{\text{W}}$ is a direct intra-path the following constraint is added:

$$x_{(u,v)}^{e,i} = 0 \text{ and } x_{(v,u)}^{e,i} = 0,$$
$$v \in V_m^{\text{BORDER}}, u \in V_m, v \neq h^e, v \neq t^e,$$
$$i = 1..n_e^{\text{W}}, e \in E_m^{\text{VIRTUAL}}. \quad (7.21)$$

Note that dummy loops might be obtained. Although they do not affect neither the feasibility nor the value of the optimal solution, they are not desirable because they lead to identical solutions in the practice. The following constraints can be added to eliminate the loops and are needed as we will introduced the diversity condition in section 7.6.3.

$$\sum_u x_{(u,v)}^{e,i} \begin{cases} \leq 1, & \text{if } v \in V_m, v \neq h^e \\ = 0, & \text{if } v = h^e \end{cases} \quad e \in E_m^{\text{VIRTUAL}}. \quad (7.22)$$

7.6.2 Flow conservation constraint for backup intra-paths

Similar to the flow conservation for working intra-paths, the following constraints apply for each backup intra-path:

$$\sum_u y^{e,i}_{(u,v)} - \sum_u y^{e,i}_{(v,u)} = \begin{cases} 1 & \text{if } v = h^e \\ 0 & \text{if } v \neq h^e, t^e \\ -1 & \text{if } v = t^e \end{cases},$$
$$v \in V_m, i = 1..n^{\mathrm{B}}_e, e \in E^{\mathrm{VIRTUAL}}_m. \quad (7.23)$$

$$y^{e,i}_{(u,v)} = 0 \text{ and } y^{e,i}_{(v,u)} = 0,$$
$$v \in V^{\mathrm{BORDER}}_m, u \in V_m, v \neq h^e, v \neq t^e,$$
$$i = 1..n^{\mathrm{B}}_e, e \in E^{\mathrm{VIRTUAL}}_m. \quad (7.24)$$

$$\sum_u y^{e,i}_{(u,v)} \begin{cases} \leq 1, & \text{if } v \in V_m, v \neq h^e, \\ = 0, & \text{if } v = h^e. \end{cases}, e \in E^{\mathrm{VIRTUAL}}_m. \quad (7.25)$$

7.6.3 Diversity condition

Potential intra-paths for the same virtual link must be mutually different. For this reason we introduce the diversity condition which forces the intra-paths in $\mathcal{P}^{\mathrm{W}}_e$ to be distinct, so do the intra-paths in $\mathcal{P}^{\mathrm{B}}_e$. Let us start with the diversity condition for working intra-paths. For each pair of working intra-paths $q^{\mathrm{W}}_{e,i}, q^{\mathrm{W}}_{e,j} \in \mathcal{P}^{\mathrm{W}}_e$, let $B^{q^{\mathrm{W}}_{e,i} q^{\mathrm{W}}_{e,j}}_{(u,v)}$ indicates whether one of them goes through link (u, v). $B^{q^{\mathrm{W}}_{e,i} q^{\mathrm{W}}_{e,j}}_{(u,v)}$ is equal to

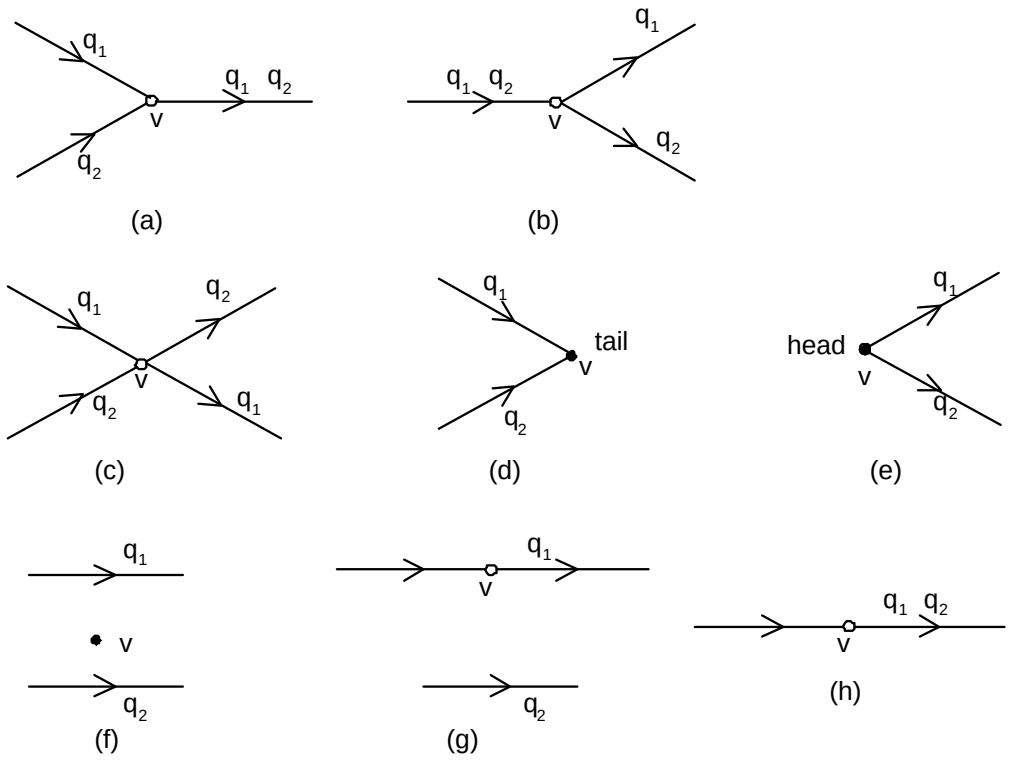

Figure 7.8: Possible cases for node v with respect to two intra-paths associated with the same virtual link. Cases (a), (b), (c), (d), (e): node v is a merging or switching point. Cases (f), (g), (h): node v is not a switching or merging point.

1 if $q^{\mathrm{W}}_{e,i}$ or $q^{\mathrm{W}}_{e,j}$ goes through (u,v) and 0 otherwise, thus:

$$B^{q^{\mathrm{W}}_{e,i}q^{\mathrm{W}}_{e,j}}_{(u,v)} = \begin{cases} 0 & \text{if } (u,v) \notin q^{\mathrm{W}}_{e,i}, q^{\mathrm{W}}_{e,j} \\ 1 & \text{otherwise.} \end{cases},$$

$$i \neq j, e \in E^{\mathrm{VIRTUAL}}_m. \quad (7.26)$$

$B^{q^{\mathrm{W}}_{e,i}q^{\mathrm{W}}_{e,j}}_{(u,v)}$ is computed as follows:

$$\frac{1}{2}\left(x^{e,i}_{(u,v)} + x^{e,j}_{(u,v)}\right) \leq B^{q^{\mathrm{W}}_{e,i}q^{\mathrm{W}}_{e,j}}_{(u,v)} \leq x^{e,i}_{(u,v)} + x^{e,j}_{(u,v)} \quad (7.27)$$

$$B^{q^{\mathrm{W}}_{e,i}q^{\mathrm{W}}_{e,j}}_{(u,v)} \in \{0,1\}.$$

Two intra-paths of the same virtual link intersect each other at the head and the tail nodes of the virtual link. If intra-paths are different, they must have at least one merging and one switching points such as v in cases (a), (b), (c) (d) or (e) of Fig. 7.8. If case (b) occurs at a node, case (a) or (c) or (d) must occur at another node because q_1, q_2 have to join each other at the tail node of their virtual link. Similarly, if case (e) occurs at a node, case (a) or (c) or (d) must occur at some other nodes. Therefore, the diversity condition is satisfied if there exists a node v on q_1, q_2 so that at least one of cases (a), (c) or (d) occurs. In these cases, $\sum_u B_{(u,v)}^{q_1 q_2} = 2$. If v is neither a switching nor a merging point, i.e., cases (f), (g), (h) of Fig. 7.8, then $\sum_u B_{(u,v)}^{q_1 q_2} = 0$ or 1. The diversity condition is thus stated as:

$$\exists v \in V_m, \sum_u B_{(u,v)}^{q_1 q_2} = 2. \tag{7.28}$$

We introduce $r_v^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} \in \{0, 1\}$ as the decision variable which takes the value 1 if $\sum_u B_{(u,v)}^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} = 2$ and 0 otherwise. Hence, $r_v^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} = 1$ if $q_{e,i}^{\mathrm{W}}$ cuts $q_{e,j}^{\mathrm{W}}$ at v. Then, (7.28) can be rewritten as follows:

$$\frac{1}{2} \sum_u B_{(u,v)}^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} \geq r_v^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} \geq \sum_u B_{(u,v)}^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} - 1,$$
$$v \in V_m, i, j = 1..n_e^{\mathrm{W}}, i \neq j, e \in E_m^{\mathrm{VIRTUAL}} \tag{7.29}$$

$$\sum_{v \in V_m} r_v^{q_{e,i}^{\mathrm{W}} q_{e,j}^{\mathrm{W}}} \geq 1, \quad i, j = 1..n_e^{\mathrm{W}}, i \neq j, e \in E_m^{\mathrm{VIRTUAL}}. \tag{7.30}$$

The diversity among backup intra-paths is defined similarly. For two backup

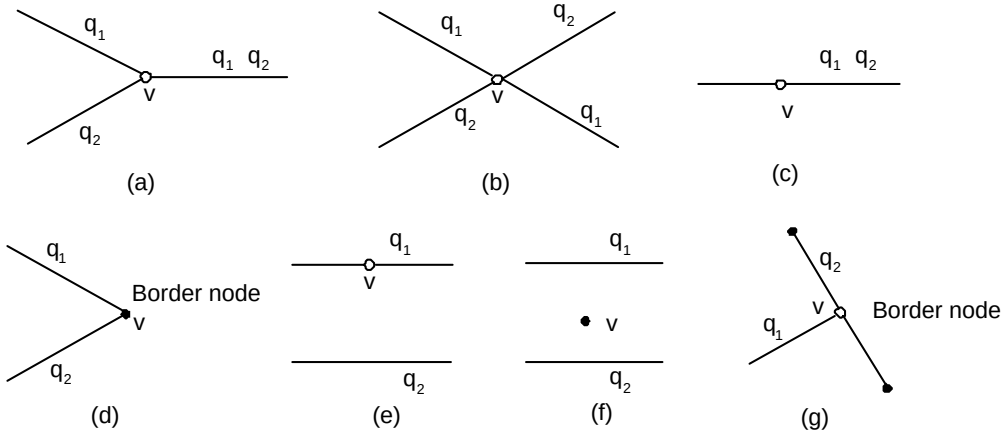

Figure 7.9: Positions of a node v with respect to two intra-paths of two virtual links regardless their directions.

intra-paths $q^{\mathrm{B}}_{e,i}$ and $q^{\mathrm{B}}_{e,j} \in \mathcal{P}^{\mathrm{B}}_{e}$, $B^{q^{\mathrm{B}}_{e,i} q^{\mathrm{B}}_{e,j}}_{(u,v)}$ must satisfy:

$$\frac{1}{2}\left(y^{e,i}_{(u,v)} + y^{e,j}_{(u,v)}\right) \leq B^{q^{\mathrm{B}}_{e,i} q^{\mathrm{B}}_{e,j}}_{(u,v)} \leq y^{e,i}_{(u,v)} + y^{e,j}_{(u,v)} \tag{7.31}$$

$$B^{q^{\mathrm{B}}_{e,i} q^{\mathrm{B}}_{e,j}}_{(u,v)} \in \{0,1\}. \tag{7.32}$$

7.6.4 Disjointness between intra-paths

First, we consider the disjointness between two working intra-paths $q^{\mathrm{W}}_{e_1,i} \in \mathcal{P}^{\mathrm{W}}_{e_1}, q^{\mathrm{W}}_{e_2,j} \in \mathcal{P}^{\mathrm{W}}_{e_2}$. Let:

$$A^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{W}}_{e_2,j}}_{v} = \sum_{u} x^{e_1,i}_{(u,v)} + x^{e_1,i}_{(v,u)} + x^{e_2,j}_{(u,v)} + x^{e_2,j}_{(v,u)},$$

$$v \in V_m, e_1, e_2 \in E^{\text{VIRTUAL}}_m, i = 1..n^{\mathrm{W}}_{e_1}, j = 1..n^{\mathrm{W}}_{e_2}. \tag{7.33}$$

In Fig. 7.9, cases from (a) to (f) show the possible positions of a node v with respect to two intra-paths q_1, q_2 of two virtual links. In particular, case (g) cannot happen since v is an intermediate border node of one intra-path while we require that all potential intra-paths must be direct.

In cases (a), (b) (c): $A^{q_1 q_2}_{v} = 4$. In cases (d): $A^{q_1 q_2}_{v} = 2$. In case (e), v is on

only one intra-path and is not a border node, then $A_v^{q_1 q_2} = 2$. In case (f), v does not belong to any intra-path, then $A_v^{q_1 q_2} = 0$. Therefore, the disjointness between $q^{\mathrm{W}}_{e_1,i}$ and $q^{\mathrm{W}}_{e_2,j}$ is defined by:

$$\delta^{q^{\mathrm{W}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}} = \begin{cases} 0 & \text{if } e, e_2 \text{ have common end nodes} \\ 0 & \text{if } \exists v \in V_m \setminus V_m^{\mathrm{BORDER}} : A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{W}}_{e_2,j}} = 4 \\ 1 & \text{otherwise} \end{cases}$$

$$e_1, e_2 \in E_m^{\mathrm{VIRTUAL}}, i = 1..n^{\mathrm{W}}_{e_1}, j = 1..n^{\mathrm{W}}_{e_2}, \quad (7.34)$$

and it is equivalent to

$$\delta^{q^{\mathrm{W}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}} \begin{cases} = & 0 \text{ if } e, e_2 \text{ have common end nodes, otherwise} \\ \leq & 2 - \frac{A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{W}}_{e_2,j}}}{2}, \forall v \in V_m \setminus V_m^{\mathrm{BORDER}} \end{cases} \quad (7.35)$$

with

$$\delta^{q^{\mathrm{W}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}} \in \{0, 1\}.$$

The disjointness between a working and a backup intra-path is defined similarly. Let the two intra-paths be $q^{\mathrm{W}}_{e_1,i} \in \mathcal{P}^{\mathrm{W}}_{e_1}$ and $q^{\mathrm{B}}_{e_2,j} \in \mathcal{P}^{\mathrm{B}}_{e_2}$. We need to replace $A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{W}}_{e_2,j}}$ by $A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{B}}_{e_2,j}}$ in $\delta^{q^{\mathrm{B}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}}$ and the value of $A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{B}}_{e_2,j}}$ is computed by:

$$A_v^{q^{\mathrm{W}}_{e_1,i} q^{\mathrm{B}}_{e_2,j}} = \sum_u x^{e_1,i}_{(u,v)} + x^{e_1,i}_{(v,u)} + y^{e_2,j}_{(u,v)} + y^{e_2,j}_{(v,u)}. \quad (7.36)$$

7.6.5 Disjointness between working and backup virtual links

Working virtual link e_1 and backup virtual link e_2 are disjoint iff there exist two intra-paths $q^{\mathrm{W}}_{e_1,i} \in \mathcal{P}^{\mathrm{W}}_{e_1}$ and $q^{\mathrm{B}}_{e_2,i} \in \mathcal{P}^{\mathrm{B}}_{e_2}$ that are disjoint. Therefore, $\delta^{e_2}_{e_1}$ must satisfy:

$$\frac{1}{n^{\mathrm{W}}_{e_1} \times n^{\mathrm{B}}_{e_2}} \sum_{i=1}^{n^{\mathrm{W}}_{e_1}} \sum_{j=1}^{n^{\mathrm{B}}_{e_2}} \delta^{q^{\mathrm{B}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}} \leq \delta^{e_2}_{e_1} \leq \sum_{i=1}^{n^{\mathrm{W}}_{e_1}} \sum_{j=1}^{n^{\mathrm{B}}_{e_2}} \delta^{q^{\mathrm{B}}_{e_2,j}}_{q^{\mathrm{W}}_{e_1,i}} \quad (7.37)$$

$$\delta_{e_1}^{e_2} \in \{0,1\}. \tag{7.38}$$

7.6.6 Objective function

The objective function becomes:

$$\begin{aligned}
\min\Bigg(\quad & \mu_1 \sum_{e \in E_m^{\text{VIRTUAL}}} \sum_{i=1}^{n_e^{\text{W}}} \sum_{(u,v) \in L_m} \frac{x_{(u,v)}^{e,i}}{c_{(u,v)}^{\text{res}}} \\
+ \quad & \mu_2 \sum_{e \in E_m^{\text{VIRTUAL}}} \sum_{i=1}^{n_e^{\text{B}}} \sum_{(u,v) \in L_m} \frac{y_{(u,v)}^{e,i}}{c_{(u,v)}^{\text{res}}} \\
- \quad & \mu_3 \sum_{e_1,e_2 \in E_m^{\text{VIRTUAL}}} \sum_{i=1}^{n_{e_1}^{\text{W}}} \sum_{j=1}^{n_{e_2}^{\text{W}}} \delta_{q_{e_1,i}^{\text{W}}}^{q_{e_2,j}^{\text{W}}} \\
- \quad & \mu_4 \sum_{e_1,e_2 \in E_m^{\text{VIRTUAL}}} \delta_{e_1}^{e_2} \\
- \quad & \mu_5 \sum_{e_1,e_2 \in E_m^{\text{VIRTUAL}}} \sum_{i=1}^{n_{e_1}^{\text{W}}} \sum_{j=1}^{n_{e_2}^{\text{B}}} \delta_{q_{e_1,i}^{\text{W}}}^{q_{e_2,j}^{\text{B}}} \Bigg).
\end{aligned} \tag{7.39}$$

The coefficients $\mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 should be carefully chosen as already discussed in Section 7.5.1.

7.7 Heuristic Mapping solution

This section presents a greedy heuristic for solving the Mapping sub-problem. The main idea of the heuristic is as follows. We do not consider all possible intra-paths but only a subset $\mathcal{P}_e^{\text{CAN}} \subset \mathcal{P}_e$ of n_e^{CAN} intra-path candidates for each virtual link e. Of course, $n_e^{\text{CAN}} \geq n_e^{\text{W}}, n_e^{\text{CAN}} \geq n_e^{\text{B}}$. For satisfying Criteria 1 and 2 defined in Section 7.5, $\mathcal{P}_e^{\text{CAN}}$ will be the set of shortest intra-paths weighted by their residual capacities.

Again, the Mapping is performed independently in each domain. We start with an empty list of intra-paths for each virtual link. The list of virtual links of the domain under study is browsed. For each virtual link, we try to find several working intra-paths so that they increase the most the number of disjoint working and backup virtual link pairs (at the first iteration, it amounts to choose the working

intra-paths that lead to the largest number of disjoint working and backup virtual link pairs). Amongst these working intra-paths, we select the one that is disjoint with the largest number of the working intra-paths that are already selected in the domain. Then a backup intra-path is also identified for the virtual link as the intra-path that increases the most the number of working virtual links that are disjoint with the considered virtual link. The next virtual link will be considered in the same way. When all virtual links are visited, another round is started again and again until each virtual link receives the required number of potential working and backup intra-paths. The algorithm is detailed in the pseudo-code of Alg. 3: Greedy_mapping(N_m).

7.8 Mapping refresh

Let threshold ϵ^{res} be the smallest amount of bandwidth remaining in each intra-path before refreshing the current Mapping. In order to avoid blocking due to link saturation, ϵ^{res} must not be smaller than the smallest requested bandwidth and is not necessary to be greater than the largest requested bandwidth.

Observe that, in a domain, the smallest residual capacity of the physical links that belong to some intra-paths is equal to the smallest residual capacity of the virtual edges. Therefore, as soon as the residual capacity of a physical link, which belongs to an intra-path, gets equal to ϵ^{res}, the Mapping should be refreshed for the domain containing that physical link.

7.9 Experimental results

MaR-O and MaR-G will denote the MaR approach with the optimal and greedy heuristic mappings respectively. They will be compared with the optimal single domain OSSP solution, denoted by Opt, in [HTC04] and the multi-domain solutions GROS and DYPOS proposed in [TJ07b].

Since the role of working and backup intra-paths are symmetric in the Mapping, we implemented, in all experiments, the exact and greedy mapping models pre-

Algorithm 3 Greedy_mapping(N_m)

for all $e \in E_m^{\text{VIRTUAL}}$ **do**
  $\mathcal{P}_e^{\text{CAN}}$ = set of n^{CAN} shortest intra-paths weighted by residual capacity.
end for
while $\exists e \in E_m^{\text{VIRTUAL}}$ so that $|\mathcal{P}_e^{\text{W}}| < n_e^{\text{W}}$ and $|\mathcal{P}_e^{\text{B}}| < n_e^{\text{B}}$ **do**
  {—Some virtual links have not received enough potential intra-paths —}
  for all $e \in E_m^{\text{VIRTUAL}}$ **do**
    if $|\mathcal{P}_e^{\text{W}}| < n_e^{\text{W}}$ **then**
      {—Select an intra-path for e if its set of potential working intra-paths is not full—}
      for all $q \in \mathcal{P}_e^{\text{CAN}}$ **do**
        $dj_q \leftarrow$ Number of backup virtual links that are newly disjoint with e thanks to q
      end for
      $S_e \leftarrow$ Set of n intra-paths that have the highest dj_q
      $q \leftarrow$ The intra-path in S_e that is disjoint with the largest number of working intra-paths in $\bigcup_{e_1 \in E_m^{\text{VIRTUAL}}} \mathcal{P}_{e_1}^{\text{W}}$
      $\mathcal{P}_e^{\text{W}} = \mathcal{P}_e^{\text{W}} \cup \{q\}$
    end if
    if $|\mathcal{P}_e^{\text{B}}| < n_e^{\text{B}}$ **then**
      {—Select an intra-path for e if $\mathcal{P}_e^{\text{B}}$ is not full—}
      for all $q \in \mathcal{P}_e^{\text{CAN}}$ **do**
        $dj_q \leftarrow$ Number of working virtual links that are newly disjoint with e thanks to q
      end for
      $q \leftarrow$ The intra-path that has the highest dj_q
      $\mathcal{P}_e^{\text{B}} = \mathcal{P}_e^{\text{B}} \cup \{q\}$
    end if
  end for
end while

sented in Sections 7.6 and 7.7 with a unique set $\mathcal{P}_e^{\mathrm{WB}}$ of n_e^{WB} potential intra-paths for both working and backup traffic. The model for MaR-O was implemented by removing from the original one the constraints and the objective terms related to backup intra-paths, i.e., the terms weighted by μ_2 and μ_5. In the experiments, the coefficients of the objective function of MaR-O are set as follows: $\mu_1 = \max_{\ell \in L_m} c_\ell^{\mathrm{res}}$, $\mu_3 = \frac{1}{(n^{\mathrm{W}} \times |E_m^{\mathrm{VIRTUAL}}|)^2}$ and $\mu_4 = 1$. The implemented model for MaR-G is deduced from the original one by removing the computations for backup intra-paths.

For both MaR-O and MaR-G, the number of needed intra-paths per virtual link is $n^{\mathrm{W}} = 2$ and the number of intra-path candidates for MaR-G is $n^{\mathrm{CAN}} = 4$. The Mapping step of MaR-O is solved using Cplex [cpl07]. Opnet Modeler [mod07] is used to implement DYPOS for the Routing step of both MaR-O and MaR-G. Note that in MaR-O, MaR-G, GROS and DYPOS, working and backup segment lengths are limited by threshold l^{W} and l^{B} respectively.

7.9.1 Mapping evaluation

The greedy mapping of MaR-G is compared with the optimal mapping of MaR-O on 2 large multi-domain networks: LARGE-5 and LARGE-8. LARGE-5 is built using 5 real optical networks: EON [OSYZ95] (29 nodes, 39 edges), RedIris [RED05] (19, 31), Garr [GAR05] (15, 24), Renater [REN05] (18, 23), SURFnet [SUR05] (25, 34). It is also used for the experiments in [TT06], [JT06], [TJ07b]. LARGE-8 is generated using the Transit-Stub model of GT-ITM [ZCB96], a well known multi-domain network generator. The network contains 8 domains, each one has on average 4 neighboring domains in order to reflect faithfully the Internet interconnections [MP01]. The numbers of nodes and links of each domain are: $(20, 53)$, $(20, 29)$, $(21, 48)$, $(22, 41)$, $(18, 36)$, $(20, 44)$, $(17, 27)$, $(22, 47)$. Readers can visit [LAR07] for the detailed topologies of the two networks.

For comparing MaR-G and MaR-O, we compute the values of the mapping objective function (7.17) as well as each of its terms with the mapping results obtained from MaR-G and MaR-O. Table 7.1 and 7.2 give the relative gaps between the values obtained from MaR-G and those obtained from MaR-O when network

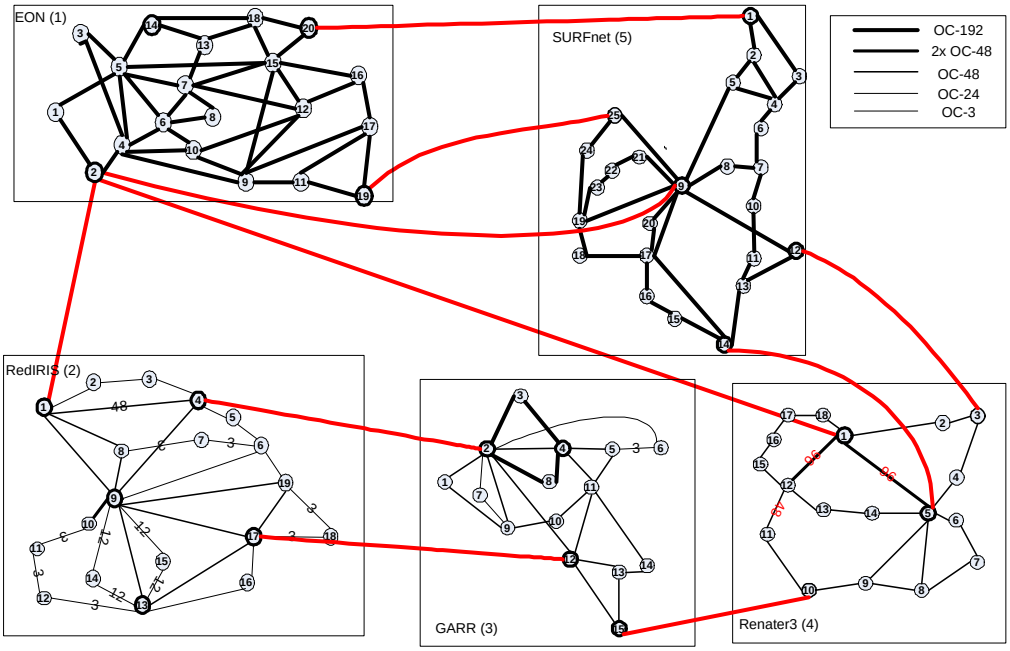

Figure 7.10: Multi-domain network LARGE-5

Domains	(μ_1) cost	($-\mu_3$) dj_{ip}	($-\mu_4$) dj_{vl}	obj
EON	-2,17	-33	0	2.52
RedIRIS	0	0	0	0
GARR	0	0	0	0
Renater	25	0	0	26.48
SURFnet	7,81	0	0	10.89

cost (%): relative gap on intra-path cost.
dj_{vl} (%): relative gap on number of disjoint virtual links.
dj_{ip} (%): relative gap on number of disjoint intra-paths.
obj (%): relative gap on the overall objective function.

Table 7.1: Relative gap of MaR-G vs. MaR-O in LARGE-5 with real link capacities.

Domains	(μ_1) cost	($-\mu_3$) dj_{ip}	($-\mu_4$) dj_{vl}	obj
EON	-2,17	-33	0	2.52
RedIRIS	0	0	0	0
GARR	0	0	0	0
Renater	20	0	0	25.03
SURFnet	7,81	0	0	10.89

Table 7.2: Relative gap of MaR-G vs. MaR-O in LARGE-5 with uniform link capacities.

links take real and uniform capacity values respectively. For the first instance, the comparison is only made on LARGE-5 due to the too high computational effort for solving MaR-O in LARGE-8. The total cost of intra-paths, the number of disjoint intra-path pairs and the number of disjoint virtual link pairs refer to the first, the third and the fourth terms of the objective function (7.17) of MaR-O respectively. The overall objective in the last columns of the two tables refers to the whole objective function.

Most differences between MaR-G and MaR-O are found in the intra-path costs. Since the intra-path cost is prioritized in the objective function, the differences reflect clearly in the overall objective gap. However the intra-path cost gaps are generally small, leading to small overall objective gaps in most domains, except for Renater domain, for both real and uniform link capacities.

In summary, the mapping results of MaR-G are close to those of the optimal Mapping MaR-O, illustrating the efficiency of the proposed greedy Mapping.

7.9.2 Scalability in using non-border maximal SRGs

Section 7.3.1 and 7.4 show that only the non-border maximal SRGs need to be advertised amongst domains. The smaller is the number of advertised SRGs, the more scalable MaR is. Table 7.3 shows the extremely small number of SRGs that need to be advertised (denoted by adv. in Tables 7.3 and 7.4) in LARGE-5 in comparison with the number of original SRGs (denoted by org.), which would be

Domains	EON	RedIRIS	GARR	Renater	SURFnet	Total
Nb. org. SRGs (MaR-O)	16	15	13	12	21	77
Nb. org. SRGs (MaR-G)	12	16	13	16	22	79
Nb. SRG^{MAX} (MaR-O)	8	6	5	4	6	29
Nb. SRG^{MAX} (MaR-G)	7	6	5	4	6	28
Nb. adv. SRGs (MaR-O)	4	1	1	0	1	7
Nb. adv. SRGs (MaR-G)	3	1	1	0	1	6

Table 7.3: Number of SRGs in LARGE-5

Domains	1	2	3	4	5	6	7	8	Total
Nb. org. SRGs	16	16	21	18	18	20	15	17	141
Nb. SRG^{MAX}	5	12	17	9	17	19	10	13	102
Nb. adv. SRGs	1	4	8	3	8	10	4	6	44

Table 7.4: Number of SRGs of LARGE-8 with MaR-G

required in a single-domain solution, as well as with the number of maximal SRGs. Most domains require either 1 or no SRGs to be advertised.

In LARGE-8, domains are highly connected with more links and internal nodes in comparison to those of LARGE-5. This topological characteristic leads to a less drastic reduction of the number of SRGs (see Table 7.4). However, more than 68% of all SRGs are still eliminated. The number of SRGs to be advertised per domain remains small.

In conclusion, the use of only non-border maximal SRGs in backup cost computation leads to a significantly more scalable routing solution while maintaining the accuracy of the cost.

7.9.3 Routing evaluation

In this section, the Routing step is evaluated together with the Mapping step through the final routing results. Note that the dynamic programming solution used in DYPOS is used for the Routing step of MaR-G as well as of MaR-O. Let us now introduce the metrics we use for the evaluation.

The working (resp. backup) network cost is the total working (resp. backup) bandwidth used by all network links.

The *Backup overhead* is the ratio between the total working and backup network cost and the smallest working network cost minus 1. This amounts to the backup bandwidth redundancy of a protection scheme. In other words, it represents the backup bandwidth saving level of a protection scheme. The smallest working network cost can be obtained when all working paths are the shortest paths.

The *Overall blocking probability* is the percentage of the total rejected bandwidth out of the total requested bandwidth by all connections.

7.9.3.1 Comparison with optimal single domain OSSP solution

M*a*R-O, M*a*R-G, Opt, GROS and DYPOS are compared only on a small 5 domain network of 28 nodes with 70 submitted requests due to the extremely high computational effort needed for Opt. All requests remain active in the network without tearing down. The Transit-Stub model of GT-ITM is used again for generating this network instance that we denote by SMALL-5 and represent in Fig. 7.11.

Fig. 7.12 depicts the backup overhead of different solutions in SMALL-5 when the working segment length threshold l^{W} varies from 2 to 5. Due to the small size of the network, there is no need for testing with larger l^{W}. For the same reason, the backup segment length constraint is removed. M*a*R-O and M*a*R-G outperform GROS and DYPOS for most values of l^{W}. When the constraint on working segment lengths is loose, i.e., l^{W} becomes large, M*a*R-O, M*a*R-G and Opt operate under similar conditions since no segment length constraint is required in Opt. In this case, M*a*R-O and M*a*R-G provide nearly identical backup overheads to Opt, revealing their high performances in bandwidth saving.

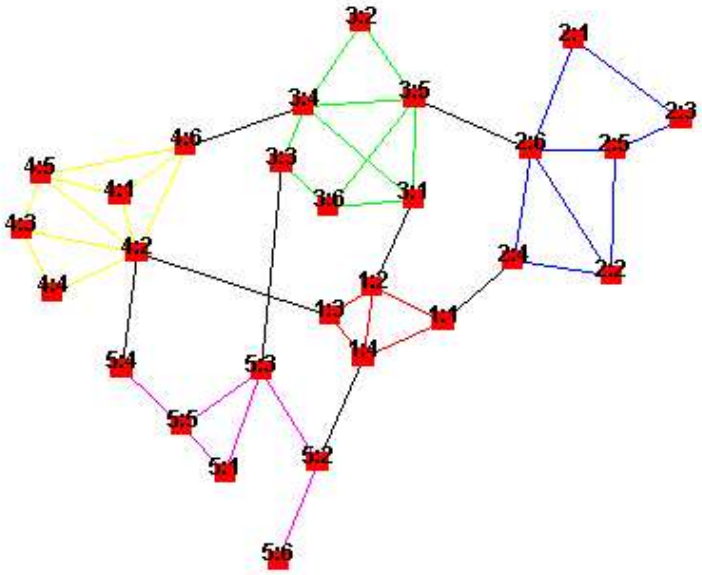

Figure 7.11: SMALL-5 network.

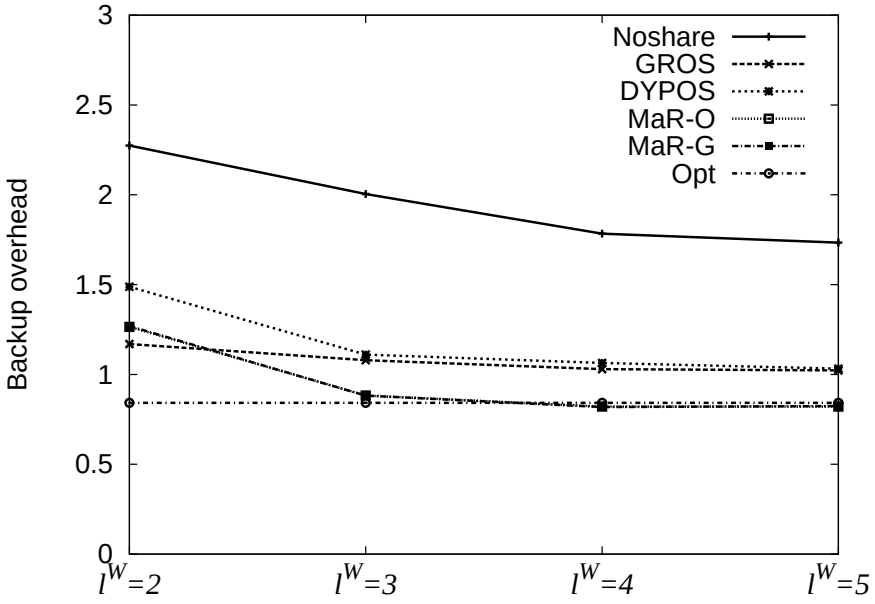

Figure 7.12: Comparison with Opt on Backup overhead in SMALL-5

7.9.3.2 Backup overhead

An advantage of MaR over RaM is that it solves a single routing optimization problem instead of multiple optimizations in inter-domain and intra-domain routings. In addition, MaR does not suffer from the approximation of working and backup cost computations in inter-domain routing as in RaM. This allows MaR to improve the quality of its final solutions. We will see through the backup overhead metric that MaR provides a better bandwidth saving.

We conducted experiments with an incremental traffic. The incremental traffic is generated by submitting subsequently 1000 connection requests to the network with all requests remaining active. This type of traffic allows keeping more requests active in the network and thus allow evaluating more accurately the bandwidth allocation characteristics of each solution scheme. Network links are uncapacitated in order to avoid the blocking cases which vary from one scheme to the other and thus would make the analysis more complex. Backup overhead is computed once after 1000 requests. No experiment is performed with MaR-O because of its high computational effort in LARGE-8 and the similarity of its routing solutions with those of MaR-G in LARGE-5.

Fig. 7.13 and Fig. 7.14 depicts backup overheads of MaR-G, GROS, DYPOS and NoShare in LARGE-5 and LARGE-8 when working and backup segment length thresholds vary. NoShare denotes the protection scheme with no backup bandwidth sharing. Obviously, MaR-G, GROS and DYPOS give better backup overheads than NoShare. As expected, MaR-G provides generally a smaller backup overhead than GROS and DYPOS.

7.9.3.3 Blocking probability

The blocking probability is examined under dynamic traffic. Requests for connection arrive and tear down after a holding time. Requests arrive according to a Poisson process with rate $r = 1$ and with an exponentially distributed holding time with mean $h = 320$. There are, on average, 320 active connections in the network.

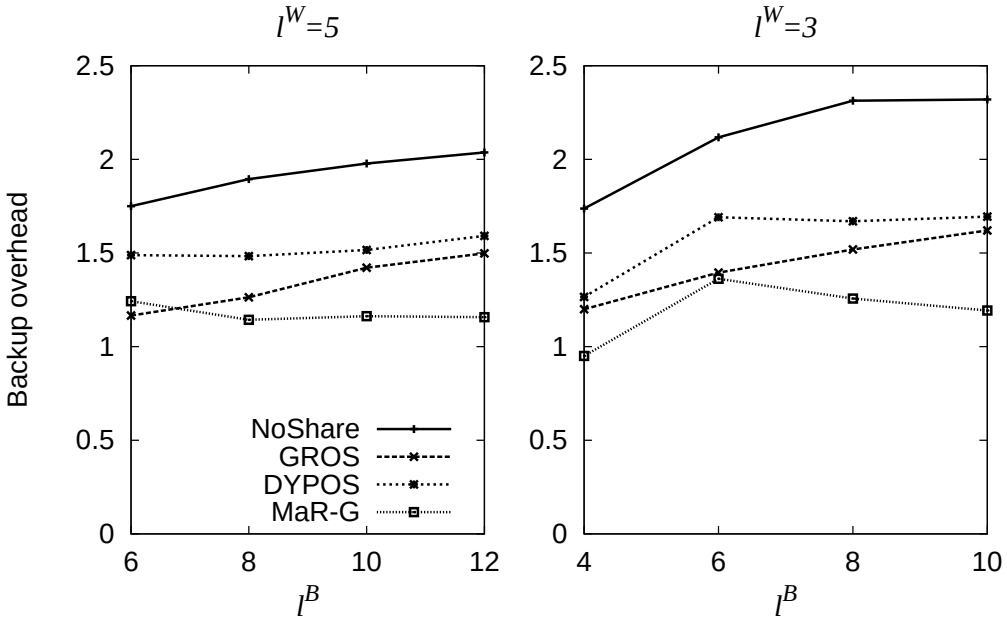

Figure 7.13: Backup overhead in LARGE-5

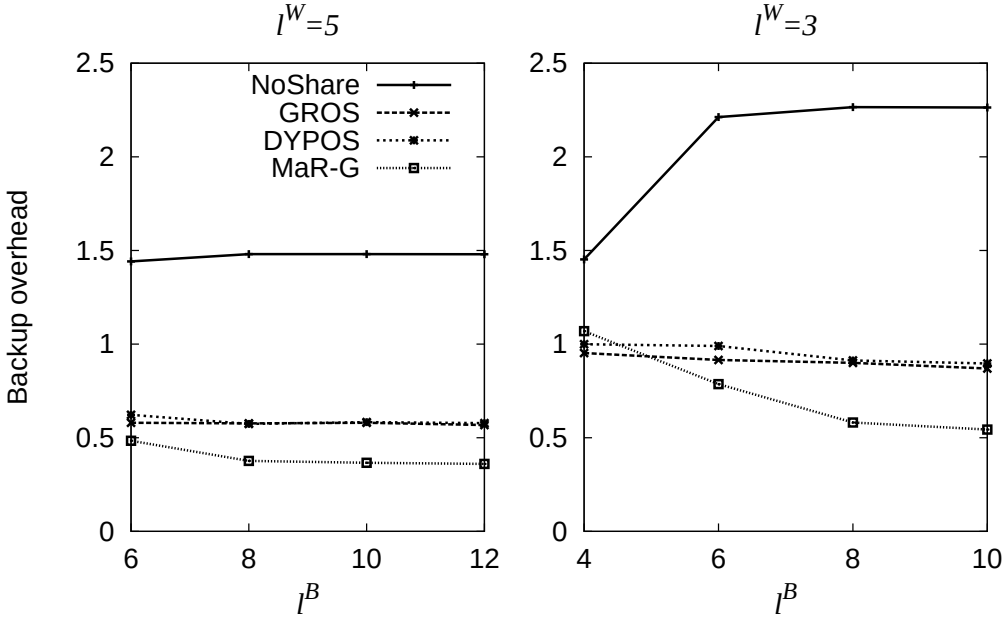

Figure 7.14: Backup overhead in LARGE-8

In general, MaR-G provides clearly smaller blocking probability than DYPOS, GROS and NoShare with $l^{\mathrm{W}} = 5$ in LARGE-5 (see Fig. 7.15) and in LARGE-8 (see Fig. 7.16). An insight in GROS and DYPOS reveals that most of their blocking is caused by bad guidances obtained from the inter-domain routing due to the cost approximation and the impossibility of mapping virtual links in the intra-domain step so that their working and backup segments are disjoint. MaR-G overcomes these weaknesses by using a unique routing based on precise working and backup costs of virtual edges as well as their disjointness indexes.

However, we observe from the results on both backup overhead and blocking probability that when segment lengths are highly limited, i.e., $l^{\mathrm{W}} = 3$ or small l^{B}, MaR-G sometimes looses its advantage. On the one hand, it is more difficult for MaR-G to build a solution satisfying segment length constraints from the restricted number of potential intra-paths, $n^{\mathrm{W}} = 2$, than GROS and DYPOS which have no restriction on the selection of intra-paths. On the other hand, the segment length constraints in DYPOS or GROS are applied on the approximation of virtual link lengths. Thus, some accepted connections may not satisfy the segment length constraints.

7.9.3.4 Number of sharing cases

Fig. 7.17 and Fig. 7.18 show the percentages of requests that benefit from backup bandwidth sharing out of the successfully routed requests in LARGE-5 and LARGE-8. The number of sharing cases are counted under dynamic traffic in order to reflect faithfully bandwidth sharing situation. Most requests share bandwidth with previously routed requests. In LARGE-5 as well as in LARGE-8, the Routing step of MaR-G encourages more requests to benefit from backup bandwidth sharing than GROS and DYPOS.

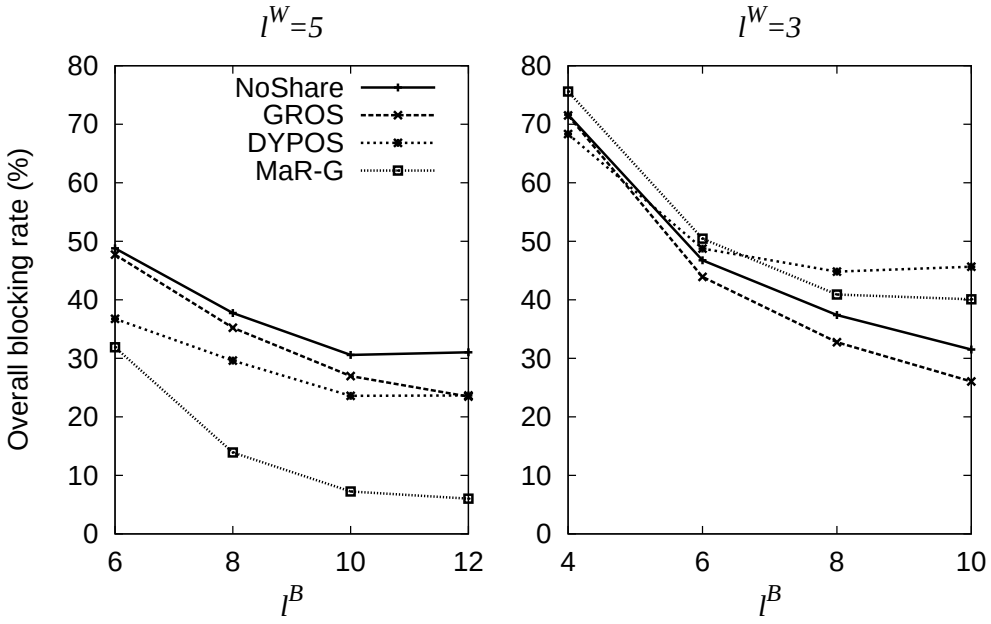

Figure 7.15: Overall bloking probability in LARGE-5

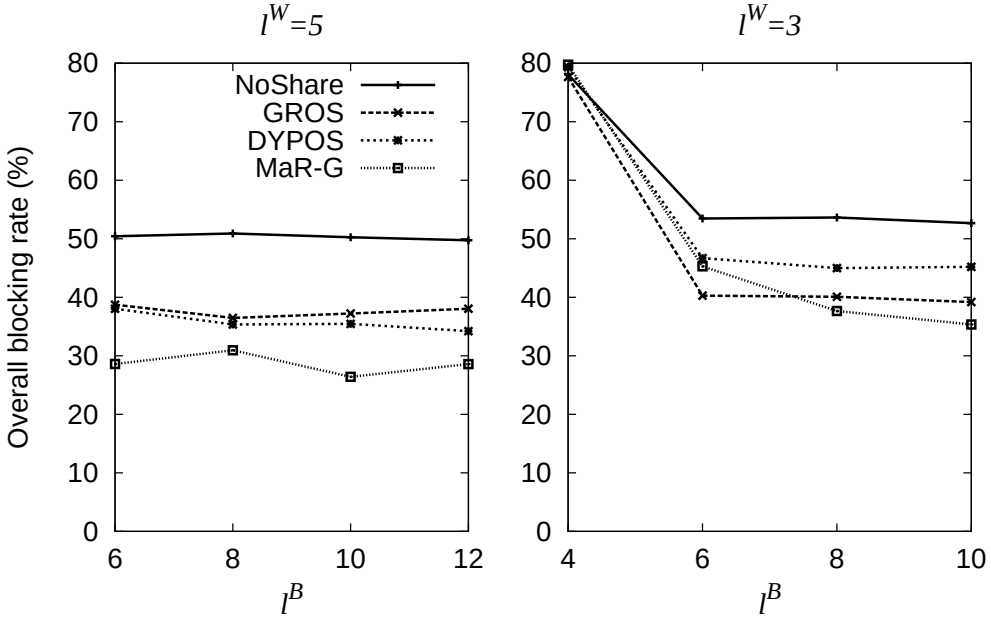

Figure 7.16: Overall bloking probability in LARGE-8

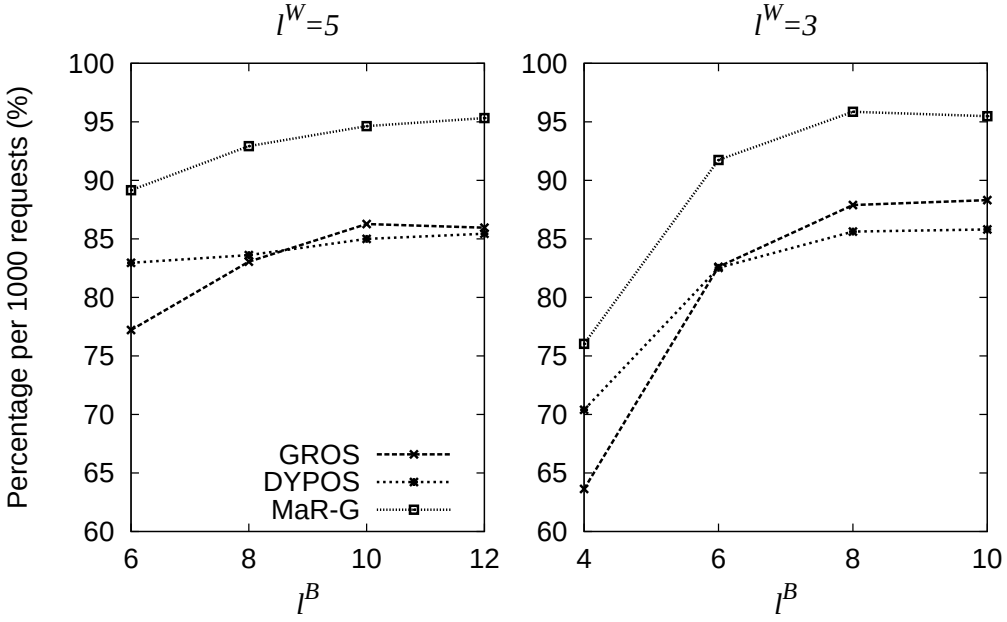

Figure 7.17: Percentage of the number of bandwidth shared requests over the number of routed requests in LARGE-5

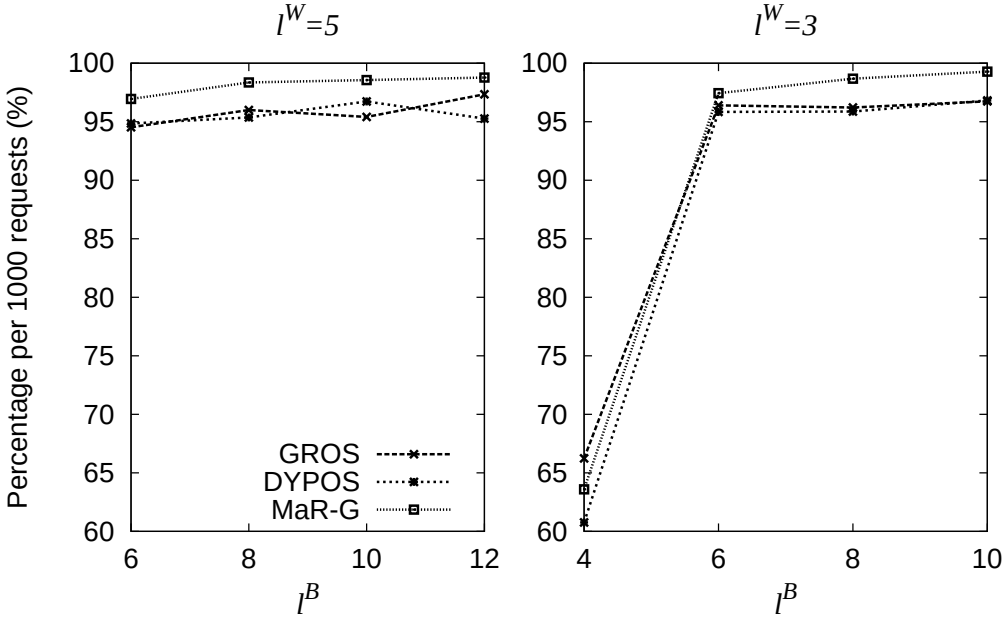

Figure 7.18: Percentage of the number of bandwidth shared requests over the number of routed requests in LARGE-8

7.10 Conclusions

In [TJ07b] we have proposed the R*a*M approach for OSSP routing in multi-domain networks. In R*a*M, approximations are used in cost computing in order to achieve the scalability. In this paper, we have proposed a new approach called M*a*R where a restricted number of potential intra-paths is selected for carrying traffic across a domain. This restriction allows M*a*R to benefit from an exact routing which is also highly scalable, although it sacrifices some small possible backup bandwidth sharing and leaves a priori less choices for building working and backup segments. Nevertheless, the Mapping with multiple well defined criteria transforms this restriction in a mechanism that orients the Routing to the best intra-paths in terms of cost, disjointness and sharing possibility. In addition, the single routing step of M*a*R allows the improvement of the quality of bandwidth optimization over the two routing steps of R*a*M.

The experimental results also confirm that M*a*R outperforms R*a*M on bandwidth saving and blocking probability. Furthermore, in bandwidth saving, M*a*R is close to the optimal single domain solution while the latter is not scalable even for a large single domain network.

M*a*R can also be applied for WDM multi-domain networks as long as the border node are wavelength conversion capable. Since intra-paths are fixed after the Mapping step, we can allocate statically one wavelength for each intra-path which becomes an optical lightpath. Wavelengths may need to be changed only at border nodes. Each network domain remains all optical without wavelength conversion at internal nodes. If the internal nodes do not have an optical bandwidth grooming capability in order to be able to perform sub-wavelength switching, a wavelength should be considered as a bandwidth unit.

CHAPTER 8

CONCLUSION

La protection joue un rôle très important dans les réseaux de transport surtout dans les réseaux multidomaines à cause de l'ampleur des impacts d'une panne simple à la fois en termes de coût et de couverture géographique (chapitre 1). Nous avons vu dans le chapitre 3 que, malgré un nombre important de recherches dans le domaine de la protection partagée, la majorité des études est restée limitée aux réseaux d'un domaine simple. Les solutions proposées exigent des informations complètes ou partielles mais toujours globales. Ces exigences en matière d'informations sont impossibles à satisfaire dans les réseaux multidomaines, en raison de la contrainte d'extensibilité. La protection d'un réseau multidomaines est plus complexe que l'ensemble des protections indépendantes des domaines simples, à cause de l'existence des noeuds de bord et des liens inter-domaines: ces derniers n'appartiennent à aucun domaine. Il y a très peu de travaux visant spécifiquement la protection des réseaux multidomaines. Ces recherches portent généralement sur des solutions de protection dédiée, ou partagée mais incomplète, qui laissent certains liens ou noeuds sans protection, ou avec une protection valable seulement pour un type de réseau spécifique.

À notre connaissance, les travaux présentés dans cette thèse sont les seuls qui proposent des solutions de protection partagée complètes pour les réseaux multidomaines. Nous avons proposé des solutions de protection par chemins, JDP et WPF (chapitre 4); ainsi que des solutions complémentaires pour celles-ci, RRGlobal, RRLocal, LeastRRLocal (chapitre 5), qui consistent à ré-optimiser les chemins de protection existants permettant d'accepter plus de demandes et de réduire la capacité de protection. Pour satisfaire la nécessité d'une rapide restauration après une panne, nous avons proposé un ensemble de solutions de protection avec des segments se chevauchant GROS, GROS-BGB, DYPOS, DYPOS-BGB (chapitre 6). Ce modèle de protection nous permet de toujours protéger tous les noeuds et

liens du réseau.

À travers ces solutions, nous avons apporté deux contributions techniques principales. La première contribution se trouve dans notre technique d'agrégation de la topologie, y compris d'informations du réseau multidomaines, qui permet d'obtenir une image simple enrichie d'informations agrégées du réseau. La deuxième contribution provient des routages à deux niveaux qui utilisent la topologie et les informations agrégées avec les réseaux initiaux. L'utilisation des informations agrégées au lieu des informations complètes ou globales durant le routage, rend nos solutions extensibles dans le contexte des réseaux multidomaines.

Jusqu'au chapitre 6, nous avons utilisé des approximations pour le calcul des coûts dans nos algorithmes de routage afin d'éliminer les demandes d'informations complètes et détaillées. Dans le chapitre 7, nous avons proposé une approche différente, MaR, pour laquelle les calculs de coût se basent sur des informations exactes. Cette nouvelle approche se résume à sacrifier de petits partages de la bande passante et à représenter des chemins internes potentiels par des arêtes virtuelles dans le but de maintenir un routage exact qui soit toujours extensible sans avoir besoin des informations complètes et globales. Comme les expérimentations l'ont montré, cette nouvelle approche améliore nettement les résultats.

Chacune de nos solutions a été comparée avec celles de problèmes similaires dans les réseaux d'un domaine simple, y compris les solutions optimales. Il faut noter que nos solutions sont pénalisées dans ces comparaisons puisqu'elles prennent en compte la contrainte d'extensibilité, ce qui n'est pas le cas des solutions concurrentes. Cette contrainte nous oblige à utiliser des informations locales, incomplètes et inexactes ou à négliger certains partages de bande passante alors que les solutions concurrentes utilisent des informations globales et complètes ou exactes. En dépit de ce possible handicap, lors des comparaisons avec des solutions optimales d'un domaine simple dans les chapitres 4 et 6, nous avons montré que nos solutions ne sont pas loin de la solution optimale, voire très proches comme dans le chapitre 7. Malgré la grande taille des réseaux multidomaines, l'effort de calcul de nos algorithmes de routage est toujours très raisonnable, de l'ordre de quelques

millisecondes, ce qui convient parfaitement aux exigences du routage en ligne. De plus, les analyses concernant la mise à jour des informations de routage ainsi que les processus de signalisation pour le calcul et l'établissement des chemins/segments d'opération et de protection montrent que la quantité de messages à échanger entre les domaines est limitée de façon à satisfaire la contrainte d'extensibilité.

La comparaison entre nos approches montre que la protection par segments offre une restauration bien plus rapide. Cependant, la protection par segments peut mener à une redondance de capacité de protection un peu plus grande que celle constatée dans le cas de la protection par chemins. Le choix définitif d'un des deux modèles dépend du critère considéré comme étant le plus important: la rapidité de restauration ou de l'économie de ressources.

Bien que les exemples, les expérimentations et les termes que nous avons employés dans la thèse privilégient parfois les réseaux à base optique, les solutions que nous avons proposées sont génériques pour tous les réseaux multidomaines avec connexions à bande passante garantie tel que MPLS, ATM, SONET/SDH et WDM avec conversion de longueurs d'onde. Nous estimons donc que les possibilités d'application de nos solutions sont très prometteuses.

Beaucoup de travail reste à faire pour rendre nos solutions prêtes à utiliser en pratique. Nous croyons qu'il faut d'abord définir en détail les messages d'information agrégée à échanger entre les domaines afin de mettre à jour les informations de routage. Il faut ensuite spécifier plus concrètement les procédures de signalisation entre les domaines durant le routage, l'établissement de chemins/segments d'opération et de protection, les procédures de notification de panne et d'activation des chemins/segments de protection. Cette thèse a cependant montré qu'il est possible de réaliser un routage dynamique pour la protection partagée dans les réseaux multidomaines. Ce routage rend ces réseaux capables de survivre lors d'une panne simple intervenant sur un lien (fibre) ou à un noeud (commutateur) et il économise des ressources en tenant compte des partages possibles de la bande passante entre les chemins/segments de protection.

Le temps limité d'une thèse ne nous permet pas d'explorer toutes les pistes

de recherche qui ont été entrevues. Nous les réservons pour des études futures. Des techniques d'optimisation plus sophistiquées pourraient être considérées pour améliorer la qualité de nos algorithmes de routage. Il faut cependant maintenir un court temps d'exécution des algorithmes afin qu'ils conviennent toujours à une utilisation en ligne.

Dans le cadre de cette thèse, nous sommes restés dans le contexte d'une panne simple. L'étude du cas de pannes multiples en général, et de pannes doubles en particulier, permettra de renforcer la tolérance aux pannes des réseaux multidomaines. Ce sujet sera une de nos futures avenues de recherche.

Dans le but de réduire le coût des conversions optique-électrique-optique (O/E/O), les réseaux DWDM tout optique, dans lesquels les signaux restent toujours dans le domaine optique, deviennent de plus en plus populaires. L'absence de conversions O/E/O aide à éliminer les cartes électriques ainsi que les traitements nécessaires et les protocoles additionnels qu'il faut installer dans les commutateurs optiques. Cependant, certaines difficultés restent à surmonter. D'abord, une contrainte de continuité de longueur d'onde doit être respectée à tous les noeuds du réseau. Des adaptations avec la prise en compte de cette contrainte seront nécessaires pour pouvoir étendre l'application de nos solutions aux réseaux multidomaines DWDM tout optique. Ce problème est complexe compte tenu de la difficulté du problème de routage dynamique dans les réseaux DWDM d'un domaine simple et compte tenu que les solutions actuelles de ce dernier problème ne sont pas très efficaces en termes de qualité. Ensuite, les réseaux DWDM se prêtent très mal à une protection par segments (ainsi que par liens) puisque les noeuds intermédiaires d'un chemin ne sont pas capables d'interpréter les signaux qui les traversent pour éventuellement signaler une panne dans le réseau. À moins de bénéficier de systèmes de surveillance de la couche physique dans les noeuds intermédiaires (ce qui implique des coûts supplémentaires), les réseaux DWDM tout optique sont condamnés à la protection par chemins. Enfin, une protection par liens ou par segments dans ces réseaux implique que plusieurs noeuds intermédiaires possèdent un organe de commutation des longueurs d'onde dans le domaines optique, ce qui est plutôt

dispendieux avec la technologie d'aujourd'hui.

BIBLIOGRAPHY

[AHSG05] C. Assi, W. Huo, A. Shami et N. Ghani : Analysis of Capacity Reprovisioning in Optical Mesh Networks. *IEEE Communication Letters*, 9(7):658–660, juil. 2005.

[AMO93] R. K. Ahuja, T. L. Magnanti et J. B. Orlin : *Network flows : theory, algorithms, and applications*. Prentice Hall, 1993.

[ASL+02] A.A. Akyamac, S. Sengupta, J.-F. Labourdette, S. Chaudhuri et S. French : Reliability in Single domain vs. Multi domain Optical Mesh Networks. *Dans Proc. National Fiber Optic Engineers Conference*, Dallas, Texas, sept. 2002.

[Bha99] R. Bhandari : *Survivable Networks: Algorithms for Diverse Routing*. Kluwer Academic, 1999.

[BL04] E. Bouillet et J.-F. Labourdette : Distributed computation of shared backup path in mesh optical networks using probabilistic methods. *IEEE/ACM Transactions on Networking*, 12(5):920–930, oct. 2004.

[Bou05] A. Bouffard : Dimentionnement GRWA et protection par segment dans les réseaux optiques WDM. Mémoire de maîtrise, Département d'Informatique et de recherche opérationnelle, Université de Montréal, nov. 2005.

[BSO02] G. Bernstein, V. Sharma et L. Ong : Interdomain Optical routing. *OSA Journal of Optical Networking*, 1(2):80–92, févr. 2002.

[can05] CAnet, 2005. `http://www.canarie.ca`.

[CGYL07] J. Cao, L. Guo, H. Yu et L. Li : A novel recursive shared segment protection algorithm in survivable WDM networks. *Journal of Network and Computer Application*, 30(2):677–694, avr. 2007.

[Cis03] Cisco System : Cisco Multiservice over SONET/SDH Product Migration and Strategy. Rapport technique, 2003.

[cpl07] CPLEX, 2007. `http://www.ilog.com/products/cplex/`.

[DLM+04] A. D'Achille, M. Listanti, U. Monaco, F. Ricciato et V. Sharma : Diverse Inter-Region Path Setup/Establishment. Rapport technique, IETF Internet-Draft, draft-dachille-diverse-interregion-path-setup-00.txt, juin 2004.

[GAR05] Consortium GARR, 2005. `http://www.garr.it`.

[GMM00] P. K. Gummadi, J. P. Madhavarapu et C. S. R. Murthy : A Segmented Backup Scheme for Dependable Real Time Communication in Multihop Networks. *Dans Proc. 15th Workshops on Parallel and Distributed Processing*, pages 678–684, 2000.

[Gro03] W. D. Grover : *Mesh-based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*. Prentice Hall, première édition, sept. 2003.

[GS98] W. Grover et D. Stamatelakis : Cycle-Oriented Distributed Preconfiguration: Ring-like Speed with Mesh-like Capacity for Self-planning Network Restoration. *Dans Proc. IEEE ICC*, Atlanta, USA, juin 1998.

[HC02] H. Huang et J.A. Copeland : Multi-domain Mesh Optical Network Protection using Hamiltonian cycles. *Dans Proc. High Perfer-formance Switching and Routing, Workshop on Merging Optical and IP Technologies*, pages 26–29, mai 2002.

[HM02] P.-H. Ho et H. T. Mouftah : A framework for service-guaranteed shared protection in WDM mesh networks. *IEEE Communications Magazine*, 40(2):97–103, févr. 2002.

[HM03] P.-H. Ho et H. T. Mouftah : Spare capacity allocation for WDM mesh networks with partial wavelength conversion capacity. *Dans Workshop on High Performance Switching and Routing*, pages 195–199, juin 2003.

[HM04a] P.-H. Ho et H. T. Mouftah : A Novel Survivable Routing Algorithm for Shared Segment Protection in Mesh WDM Networks With Partial Wavelength Conversion. *IEEE/ Journal on Selected Areas in Communications*, 22(8):1548–1560, oct. 2004.

[HM04b] P.-H. Ho et H. T. Mouftah : Reconfiguration of Spare Capacity for MPLS-Based Recovery in the Internet Backbone Networks. *IEEE Transactions on Networking*, 12(1):73–84, févr. 2004.

[HM04c] P.-H. Ho et H. T. Mouftah : Shared protection in Mesh WDM networks. *IEEE Communications Magazine*, 42(1):70–76, janv. 2004.

[HTC04] P.-H. Ho, J. Tapolcai et T. Cinkler : Segment shared protection in mesh communications networks with bandwidth guaranteed tunnels. *IEEE/ACM Transactions on Networking*, 12(6):1105–1118, déc. 2004.

[Int06] International Herald Tribune : Telephone and Internet services cut off after Taiwan earthquake, 26 déc. 2006. `http://www.iht.com/articles/ap/2006/12/27/asia/AS_GEN_Taiwan_Quake.php`.

[JO01] B. G. Jozsa et D. Orincsay : Shared backup path optimization in telecommunication networks. *Dans Proc. Dynamic of Reliable Communication Networks, DRCN'2001*, pages 251–257, oct. 2001.

[JT06] B. Jaumard et D. L. Truong : Backup Path Re-optimizations for Shared Path Protection in Multi-domain Networks. *Dans Proc. IEEE Globecom 2006*, San Francisco, USA, déc. 2006.

[KL00] M. Kodialam et T.V. Lakshman : Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration. volume 2, pages 902–911, mars 2000.

[KL03] M. Kodialam et T.V. Lakshman : Dynamic Routing of Restorable Bandwidth-Guaranteed Tunnels Using Aggregated Network Resource Usage Information. *IEEE/ACM Transactions on Networking*, 11(3): 399–410, 2003.

[Kle96] J. Kleinberg : *Approximation Algorithms for Disjoint Paths Problems*. Thèse de doctorat, Dept. of EECS, MIT, 1996.

[Lab04] J.-F. Labourdette : Shared mesh restoration in optical networks. *Dans Proc. Optical Fiber Communication Conference (OFC)*, volume 1, pages 23–27, févr. 2004.

[LAR07] LARGE-5 and LARGE-8 networks, 2007. `http://www.iro.umontreal.ca/~truongtd/topo/`.

[Liu01] Y. Liu : *Spare Capacity Allocation: Model, Analysis and Algorithm*. Thèse de doctorat, School of Information Sciences, Univ. Pittsburgh, 2001.

[LLWL06] C. Lu, H. Luo, S. Wang et L. Li : A Novel Shared Segment Protection Algorithm for Multicast Sessions in Mesh WDM Networks. *ETRI Journal*, 28(3):329–336, juin 2006.

[LMDL92] C. Li, S. T. McCornick et D.Simchi-Levi : Finding Disjoint Paths with Different Path Costs: Complexity and Algorithms. *Networks*, 22:653–667, 1992.

[LR01] G. Liu et K. G. Ramakrishnan : A*Prune: An Algorithm for Finding K Shortest Paths Subject to Multiple Constraints. *Dans Proc. Infocom 2001*, volume 1, pages 743–749, févr. 2001.

[LRVB04] J. L. Le Roux, J. P. Vasseur et J. Boyle : Requirements for Inter-area MPLS Traffic Engineering. Rapport technique, IETF Internet-Draft, draft-ietf-tewg-interarea-mpls-te-req-02.txt, juin 2004.

[LWKD03] G. Li, D. Wang, C. Kalmanek et R. Doverspike : Efficient Distributed Restoration Path Selection for Shared Mesh Restoration. *IEEE/ACM Transactions on Networking*, 11(5):761–771, oct. 2003.

[LYL06] H. Luo, H. Yu et L. Li : A heuristic algorithm for shared segment protection in mesh WDM networks with limited backup path/segments length. *ScienceDirect/Computer Communications*, 29(16):3197–3213, oct. 2006.

[MK98] K. Murakami et H. S. Kim : Optimal Capacity and Flow Assignment for Self-Healing ATM Networks Based on Line and End-to-End Restoration. *IEEE/ACM Transactions on Networking*, 6(2):1223 – 1232, avr. 1998.

[MKAM04] T. Miyamura, T. Kurimoto, M. Aoku et A. Misawa : An Inter-area SRLG-disjoint Routing Algorithm for Multi-segment Protection in GMPLS Networks. *Dans Proc. ICBN Conference*, Kobe, Japan, avr. 2004.

[mod07] Opnet Modeler, 2007. `http://www.opnet.com/solutions/network_rd/modeler.html`.

[MP01] D. Magoni et J. J. Pansiot : Analysis of the autonomous system network topology. *SIGCOMM Computer Communication Review*, 31(3):26–37, juil. 2001.

[Muk06] B. Mukherjee : *Optical WDM Networks*. Springer, 2006.

[nsf05] NSF network, 2005. `http://www.nsf.gov`.

[OMZ01] C. Ou, B Mukherjee et H Zang : Sub-Path Protection for Scalability and Fast Recovery in WDM Mesh Networks. *Dans Proc. OSA Optical Fiber Communication Conference (OFC)*, volume 54, page ThO6, France, févr. 2001.

[ORM05] C. S. Ou, S. Rai et B. Mukherjee : Extension of segment protection for bandwidth efficiency and differentiated quality of protection in optical/MPLS networks. *Optical Switching and Networking*, pages 19–33, janv. 2005.

[OSYZ95] M.J. O'Mahony, D. Simeonidu, A. Yu et J. Zhou : The Design of the European Optical Network. *Journal of Ligthwave Technology*, 13(5):817–828, mai 1995.

[OZZ+04] C. Ou, J. Zhang, H. Zang, L. H. Sahasrabuddhe et B. Mukherjee : New and Improved Approaches for Shared-Path Protection in WDM Mesh Networks. *Journal of Lightwave Technology*, 22(5):1223 –1232, mai 2004.

[QX01] C. Qiao et D. Xu : Distributed partial information management (DPIM) schemes for survivable networks-Part I. *Dans Proc. Twentieth Annual Joint Conference of the IEEE Computer and Com. Societies*, pages 302–311, avr. 2001.

[Ram99] B. Ramamurthy, S.; Mukherjee : Survivable WDM Mesh Networks, Part I - Protection. *Dans Proc. IEEE INFOCOM*, volume 2, pages 744–751, New York, NY, mars 1999.

[RED05] REDIrid, 2005. `http://www.rediris.es/red/index.en.html#red%20troncal`.

[REN05] RENATER-4 network, 2005. http://www.renater.fr.

[Reu07] Reuters : Chinese Web users lose 10,000 domain names in quakes. janv. 2007. http://www.reuters.com/article/technologyNews/idUSSHA15067820070105.

[ris07] RISQ network, 2007. http://www.risq.ca.

[RKM02] G. Ranjith, G. P. Krishna et C. S. Ram Murthy : A distributed primary-segmented backup scheme for dependable real-time communication in multihop networks. *Dans Proc. International Parallel and Distributed Processing Symposium*, pages 139–146, avr. 2002.

[RMD04] F. Ricciato, U. Monaco et A. D'Achille : A Novel Scheme for End-to-End Protection in a Multi-Area Network. *Dans IPS '04*, Budapest, Hungary, mars 2004.

[RS02] R. Ramaswami et K. N. Sivarajan : *Optical Networks: A practical perspective*. Elsevier/Morgan Kaufmann, deuxième édition, 2002.

[RSM03] S. Ramamurthy, L. Sahasrabuddhe et B. Mukherjee : Survivable WDM mesh networks. *J. Lightwave Technol.*, 21(4):870–883, avr. 2003.

[Sch01] J. Schallenburg : Is 50 ms Restoration Necessary? *Dans IEEE Bandwidth Management workshop IX*, Montebello, Quebec, Canada, juin 2001.

[SG03] G. Shen et W. D. Grover : Extending the p-Cycle Concept to Path Segment Protection for Span and Node Failure Recovery. *IEEE JSAC Optical Communications and Networking*, 21(8):1306–1319, oct. 2003.

[SG04] G. Shen et W. D. Grover : Segment-based approaches to survivable translucent network design under various ultra-long-haul system reach capabilities. *OSA Journal of Optical Networking*, 3(1):1–24, janv. 2004.

[SGA02] D. A. Schupke, C. G. Gruber et A. Autenrieth : Optimal configuration of p-cycles in WDM networks. *Dans Proc. IEEE ICC*, Atlanta, USA, juin 2002.

[SR01] S. Sengupta et R. Ramamurthy : From Network Design for Dynamic Provisioning and Restoration in Optical Cross-Connect Mesh Networks: An Architectural and Algorithmic Overview. *IEEE Network*, 15(4):46–54, juil. 2001.

[SS01] X. Su et C.-F. Su : An online distributed protection algorithm in WDM networks. *Dans Proc. IEEE ICC 2001*, volume 5, pages 1571–1575, Helsinki, Finland, juin 2001.

[ST84] J. W. Suurballe et R. E. Tarjan : A quick methods for finding shortest pairs of disjoint paths. *Networks*, 14(2):325–336, 1984.

[SUR05] Surfnet, 2005. `http://www.surfnet.nl`.

[Suu74] J. W. Suurballe : Disjoint paths in a network. *IEEE Networks*, 1974.

[TC03] J. Tapolcai et T. Cinkler : On-line Routing Algorithm with Shared Protection in WDM Networks. *Dans Proc. Optical Network Design and Modeling*, pages 351–364, Budapest, Hungary, févr. 2003.

[TH04] J. Tapolcai et P.-H. Ho : A deeper study on segment shared protection. *Dans Proc. 7th International Symposium on Parallel Architectures, Algorithms and Networks*, pages 319–325, mai 2004.

[The06] The Register : Taiwan earthquake shakes internet. 27 déc. 2006. `http://www.theregister.co.uk/2006/12/27/boxing_day_earthquake_taiwan/`.

[THVC05] J. Tapolcai, P.-H. Ho, D. Verchere et T. Cinkler : A novel shared segment protection method for guaranteed recovery time. volume 1, pages 117–126, oct. 2005.

[TJ06] D. L. Truong et B. Jaumard : Overlapped Segment Shared Protection in Multi-domain Networks. *Dans Proc. APOC*, volume 6354, pages 63541K–1–63541K–10, Gwangju, Korea, sept. 2006.

[TJ07a] D. L. Truong et B. Jaumard : A novel approach for Overlapping Segment Shared Protection in Multi-domain Networks. soumis pour publication, 2007.

[TJ07b] D. L. Truong et B. Jaumard : Using Topology Aggregation for Efficient Segment Shared Protection Solutions in Multi-domain networks. *IEEE Journal of Selected Areas in Communication/ Optical Communications and Networking series*, 25(9), déc. 2007. à paraître.

[TN94] M. To et P. Neusy : Unavailability analysis of long-haul networks. *IEEE Journal on Selected Areas in Communications*, 12(1):100–109, janv. 1994.

[TT06] D. L. Truong et B. Thiongane : Dynamic routing for Shared Path Protection in Multidomain optical mesh networks. *OSA Journal of Optical Networking*, 5(1):58–74, janv. 2006.

[Tuo06] Tuoi Tre Online : Coupure des fibres optiques à cause du tremblement de terre, Internet du Vietnam reste à 30% bande passante. 27 déc. 2006. `http://www.tuoitre.com.vn/Tianyon/Index.aspx?ArticleID=179791&ChannelID%=16`.

[XCX+04] D. Xu, Y. Chen, Y. Xiong, C. Qiao et X. He : On finding disjoint paths in single and dual link cost networks. *Dans Proc. IEEE Infocom*, volume 1, page 715, mars 2004.

[XM99] Y. Xiong et L. G. Mason : Restoration strategies and spare capacity requirements in self-healing atm networks. *IEEE/ACM Transactions on Networking*, 7(1):98–110, févr. 1999.

[XQX02] D. Xu, C. Qiao et Y. Xiong : An Ultra-Fast Shared Path Protection Scheme Distributed Partial Information Management, Part II. *Dans Proc. 10th IEEE International Conference in Network Protocols*, pages 344–353, France, nov. 2002.

[XXQ02] D. Xu, Y. Xiong et C. Qiao : Protection with Multi-Segments (PROMISE) in Networks with Shared Risk Link Groups (SRLG). *Dans Proc. The 40th Annual Allerton Conference on Communication*, pages 1320–1331, oct. 2002.

[XXQ03a] Y. Xiong, D. Xu et C. Qiao : Achieving fast and bandwidth-efficient shared-path protection. *Journal of Lightwave Technology*, 21(2):365 – 371, févr. 2003.

[XXQ03b] D. Xu, Y. Xiong et C. Qiao : Novel Algorithms for Shared Segment Protection. *IEEE/Journal on Selected Areas in Communications*, 21(8):1320–1331, oct. 2003.

[XYDQ01] C. Xin, Y. Ye, S. Dixit et C. Qiao : A Joint Working and Protection Path Selection Approach in WDM Optical Networks. *Dans Proc. IEEE Globecom*, volume 4, pages 2165–2168, San Antonio, CA, USA, nov. 2001.

[YZW06] O. Yong, Q. Zeng et W. Wei : Segment protection algorithm based on an auxiliary graph for wavelength-division multiplexing optical networks. *Journal of Optical Networking*, 5:15–25, janv. 2006.

[ZCB96] E. W. Zegura, K. L. Calvert et S. Bhattacharjee : How to Model an Internetwork. *Dans IEEE Infocom*, volume 2, pages 594–602, San Francisco, CA, mars 1996.

[ZM04a] J. Zhang et B. Mukherjee : A review of Fault Management in WDM Mesh networks: Basic Concept and Research Challenges. *IEEE Network*, 18(2):41–48, mars 2004.

[ZM04b] J. Zheng et H. T. Mouftah : *Optical WDM Networks: Concepts and Design Principles*. Wiley-Interscience/IEEE Press, première édition, 2004.