

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MÉCANISMES DE RÉOLUTION ET DE PRÉVENTION  
DE LA CONTENTION POUR LES RÉSEAUX OPTIQUES  
À COMMUTATION DE RAFALES

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN INFORMATIQUE

PAR  
MARTIN LÉVESQUE

JANVIER 2010

# UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Service des bibliothèques

## Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 -Rév.01-2006). Cette autorisation stipule que «conformément à l'article **11** du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

# Remerciements

Mes premiers remerciements s'adressent à ma directrice de recherche Mme Halima Elbiaze ayant rendu possible ce travail et pour son aide et ses conseils tout au long de ce travail. Je tiens également à remercier les membres du *Laboratoire de recherche en TéléInformatique et Réseaux de l'Université du Québec à Montréal (LTIR-UQAM)* pour leur soutien et spécialement Mohamed Faten Zhani, Wael Hosny Foued Aly et Wessam Ajib pour leurs nombreuses critiques constructives.

Je remercie également ma famille et toute personne qui, de près ou de loin, m'ont soutenu dans mon travail.

# Table des matières

Remerciements . . . . .	i
Résumé . . . . .	xii
<b>1 Introduction</b>	<b>1</b>
1.1 Problématique . . . . .	2
1.2 Contributions . . . . .	2
<b>2 État de l’art</b>	<b>4</b>
2.1 Introduction . . . . .	4
2.2 Les réseaux optiques . . . . .	4
2.2.1 Introduction . . . . .	4
2.2.2 Évolution . . . . .	6
2.2.3 Hiérarchie de couches de communication . . . . .	8
2.2.3.1 IP sur ATM sur SONET sur WDM . . . . .	8
2.2.3.2 IP sur ATM sur WDM . . . . .	9
2.2.3.3 IP sur WDM . . . . .	9
2.2.4 Multiplexage . . . . .	9
2.2.4.1 Multiplexage temporel . . . . .	9
2.2.4.2 Multiplexage en longueurs d’onde . . . . .	10
2.2.5 Composantes physiques pour commuter en optique . . . . .	11
2.2.5.1 Microsystème électromécanique . . . . .	12

2.2.5.2	Amplificateur à semi-conducteur . . . . .	12
2.2.6	Paradigmes de commutation . . . . .	13
2.2.6.1	Commutation de paquets . . . . .	14
2.2.6.2	Commutation de circuits . . . . .	15
2.2.6.3	Commutation de rafales . . . . .	15
2.2.6.4	Performances . . . . .	16
2.3	Les réseaux optiques à commutation de rafales . . . . .	18
2.3.1	Introduction . . . . .	18
2.3.2	Architecture . . . . .	19
2.3.3	Signalisation . . . . .	21
2.3.3.1	Protocoles à une direction . . . . .	21
2.3.3.2	Protocoles à deux directions . . . . .	23
2.3.3.3	Résumé des protocoles de signalisation . . . . .	25
2.3.4	Méthodes de résolution de la contention . . . . .	26
2.3.4.1	Introduction . . . . .	26
2.3.4.2	Lignes à retardement . . . . .	27
2.3.4.3	Conversion de longueur d'onde . . . . .	28
2.3.4.4	Segmentation de rafales . . . . .	28
2.3.4.5	Déflexion . . . . .	29
2.3.4.6	Retransmission . . . . .	35
2.3.4.7	Déflexion et retransmission . . . . .	36
<b>3</b>	<b>Combinaison adaptative hybride de la déflexion et de la retransmission pour les réseaux OBS</b>	<b>38</b>
3.1	Introduction . . . . .	38
3.2	Description détaillée de l'algorithme AHDR . . . . .	38
3.2.1	Transfert de statistiques à travers le réseau . . . . .	40
3.2.2	Limitation des routes alternatives . . . . .	43

3.2.3	Seuil adaptatif de décision . . . . .	43
3.2.3.1	Probabilité de succès d'une route . . . . .	44
3.2.3.2	Corrélation linéaire entre le BLR et le seuil de décision . . . . .	45
3.2.3.3	Critère de sélection de résolution de contention . . . . .	47
3.2.4	Processus d'acheminement de rafales . . . . .	48
3.2.5	Table de routage AHDR . . . . .	50
3.2.6	OT adaptatif . . . . .	51
3.3	Résultats expérimentaux et discussion . . . . .	52
3.3.1	Configurations expérimentales . . . . .	52
3.3.2	Résultats et analyse . . . . .	55
3.3.2.1	Poids du seuil de décision ( $\beta_{BLR}, \beta_U$ ) . . . . .	55
3.3.2.2	Poids de mise à jour du seuil ( $\phi$ ) . . . . .	55
3.3.2.3	Comparaison des performances entre AHDR et MLHDR . . . . .	59
3.4	Conclusion . . . . .	62
<b>4</b>	<b>Modèle graphique probabiliste de prévention de la contention pour les réseaux OBS</b>	<b>66</b>
4.1	Introduction . . . . .	66
4.2	Motivation . . . . .	66
4.3	Description détaillée de MGPPC . . . . .	69
4.3.1	Description du réseau bayésien . . . . .	69
4.3.2	Table de routage . . . . .	73
4.3.3	Signalisation et paquets de notification . . . . .	76
4.4	Résultats expérimentaux et discussion . . . . .	77
4.4.1	Configurations expérimentales . . . . .	77
4.4.2	Résultats et analyse . . . . .	78
4.5	Conclusion . . . . .	83
<b>5</b>	<b>Conclusion et perspectives</b>	<b>85</b>

Bibliographie	87
Publications réalisées	93

# Liste des tableaux

2.1	Comparaison des paradigmes de commutation (Provient de [QY99]) . . . . .	16
2.2	Caractéristiques des protocoles de signalisation (Provient de [JV04]) . . . . .	25
2.3	Caractéristiques des travaux effectués sur la déflexion . . . . .	33
3.1	Informations comprises dans un paquet de contrôle AHDR . . . . .	42
3.2	Exemple de pondération pour calculer la corrélation linéaire . . . . .	45
3.3	Table de routage AHDR . . . . .	51
4.1	Table de routage MGPPC . . . . .	74

# Liste des figures

2.1	Parties d'un réseau public . . . . .	5
2.2	3 scénarios possibles de la croissance du trafic (Provient de [Pay08]) . . . . .	6
2.3	Évolution de fibres optiques WDM sous-marines (Provient de [DDF04]) . . . . .	7
2.4	Quelques architectures en couches proposées . . . . .	8
2.5	Multiplexage temporel . . . . .	9
2.6	Multiplexage en longueurs d'onde . . . . .	10
2.7	Brasseur optique 3x3 avec 2 longueurs d'onde . . . . .	11
2.8	Exemple d'architecture de réseau tout optique WDM . . . . .	12
2.9	Commutateurs tout optique . . . . .	13
2.10	Architecture d'un routeur OPS . . . . .	14
2.11	Signalisation OBS JET . . . . .	16
2.12	Évolution des tendances de développement des commutateurs optiques . . . . .	17
2.13	Architecture d'un réseau OBS . . . . .	19
2.14	Noeud principal OBS . . . . .	20
2.15	Signalisation OBS JET à une direction . . . . .	22
2.16	Signalisation OBS TAW à deux directions . . . . .	24
2.17	Deux rafales de données en contention sur un lien optique WDM . . . . .	26
2.18	Signalisation JET avec utilisation d'une FDL pour retarder une rafale . . . . .	28
2.19	Exemple illustrant l'algorithme LHDR . . . . .	36
3.1	AHDR, scénario sans congestion ni contention . . . . .	39

3.2	AHDR, scénario de congestion . . . . .	39
3.3	Signalisation AHDR, scénario avec retransmission . . . . .	41
3.4	Corrélation linéaire entre le BLR et le seuil, topologie NSFNET . . . . .	46
3.5	Variation du seuil de décision avec une charge de 0.74, topologie NSFNET . . . . .	46
3.6	Processus d'acheminement de paquets/rafales, AHDR . . . . .	49
3.7	Topologie NSFNET . . . . .	53
3.8	Topologie COST239 . . . . .	53
3.9	Variation du BLR en fonction de $\beta_{BLR}$ et $\beta_U$ , topologie NSFNET, Charge = 0.52 . . . . .	56
3.10	Variation du BLR en fonction de $\beta_{BLR}$ et $\beta_U$ , topologie NSFNET, Charge = 1.06 . . . . .	56
3.11	Variation du BLR en fonction de $\beta_{BLR}$ et $\beta_U$ , topologie NSFNET, Charge = 1.06, sans suppression du bruit . . . . .	57
3.12	Variation du BLR en fonction de $\beta_{BLR}$ et $\beta_U$ , topologie NSFNET, Charge = 1.6	57
3.13	Variation de la charge, scénario pour tester $\phi$ . . . . .	58
3.14	Variation du BLR en fonction de $\phi$ , topologie NSFNET . . . . .	58
3.15	Variation du BLR en fonction de la charge, topologie NSFNET . . . . .	60
3.16	Variation du délai de bout en bout en fonction de la charge, topologie NSFNET	60
3.17	Variation du ratio de déflexion en fonction de la charge, topologie NSFNET . .	61
3.18	Variation du OT en fonction de la charge, topologie NSFNET . . . . .	61
3.19	Variation du BLR en fonction de la charge, topologie COST239 . . . . .	63
3.20	Variation du délai de bout en bout en fonction de la charge, topologie COST239	63
3.21	Variation du ratio de déflexion en fonction de la charge, topologie COST239 .	64
3.22	Variation du OT en fonction de la charge, topologie COST239 . . . . .	64
4.1	Un exemple simple de réseau bayésien . . . . .	68
4.2	Le modèle graphique probabiliste de routage (MGPPC), un réseau bayésien . .	70
4.3	Exemple de topologie simple avec MGPPC . . . . .	70

4.4	MGPPC qui met à jour la table de routage . . . . .	71
4.5	Délimitations de trafic en utilisant le débit sortant : Faible, moyen et grand . .	72
4.6	Apprentissage de NSFNET à l'aide de MGPPC, sans information de routage initiale . . . . .	78
4.7	Topologie Pan-European (Provient de [HBB <sup>+</sup> 03]) . . . . .	79
4.8	Variation du BLR en fonction de la charge, topologie NSFNET . . . . .	80
4.9	Variation du délai de bout en bout en fonction de la charge, topologie NSFNET	81
4.10	Variation de l'utilisation en fonction de la charge, topologie NSFNET . . . . .	81
4.11	Variation du BLR en fonction de la charge, topologie Pan-European . . . . .	82
4.12	Variation du délai de bout en bout en fonction de la charge, topologie Pan- European . . . . .	82
4.13	Variation de l'utilisation en fonction de la charge, topologie Pan-European . .	83

# Liste des acronymes

<b>ACK</b>	<i>Positive-Acknowledgement</i> .....	35
<b>AHDR</b>	<i>Adaptive Hybrid Deflection and Retransmission</i> .....	38
<b>ATM</b>	<i>Asynchronous Transfer Mode</i> .....	8
<b>BHP</b>	<i>Burst Header Packet</i> .....	15
<b>DIR</b>	<i>Destination Initiated Reservation</i> .....	25
<b>FDL</b>	<i>Fiber Delay Line</i> .....	23
<b>BLR</b>	<i>Burst Loss Ratio</i> .....	36
<b>IP</b>	<i>Internet Protocol</i> .....	8
<b>ISO</b>	<i>Interconnexion de Systèmes Ouverts</i> .....	8
<b>JET</b>	<i>Just-Enough-Time</i> .....	15
<b>JIT</b>	<i>Just-In-Time</i> .....	23
<b>HDR</b>	<i>Hybrid Deflection and Retransmission</i> .....	36
<b>LAUC-VF</b>	<i>Latest Available Unscheduled Channel with Void Filling</i> .....	21
<b>LHDR</b>	<i>Limited Hybrid Deflection and Retransmission</i> .....	36
<b>MEMS</b>	<i>Micro-Electromechanical systems</i> .....	12
<b>MGPPC</b>	<i>Modèle Graphique Probabiliste de Prévention de la Contention</i> .....	66
<b>NACK</b>	<i>Negative-Acknowledgement</i> .....	35
<b>OBS</b>	<i>Optical Burst Switching</i> .....	1
<b>OCS</b>	<i>Optical Circuit Switching</i> .....	15

<b>OPS</b>	<i>Optical Packet Switching</i> .....	14
<b>OT</b>	<i>Offset Time</i> .....	2
<b>OXC</b>	<i>Optical Cross-Connect</i> .....	11
<b>RAM</b>	<i>Random-Access Memory</i> .....	1
<b>SCU</b>	<i>Switch Control Unit</i> .....	20
<b>SIR</b>	<i>Source Initiated Reservation</i> .....	25
<b>SOA</b>	<i>Semiconductor Optical Amplifier</i> .....	12
<b>SONET</b>	<i>Synchronous Optical Networking</i> .....	8
<b>TAG</b>	<i>Tell-And-Go</i> .....	23
<b>TAW</b>	<i>Tell-And-Wait</i> .....	15
<b>TCP</b>	<i>Transmission Control Protocol</i> .....	32
<b>TLO</b>	<i>Terminal de Ligne Optique</i> .....	10
<b>TTL</b>	<i>Time-To-Live</i> .....	34
<b>WDM</b>	<i>Wavelength-Division Multiplexing</i> .....	4

# Résumé

L'explosion actuelle de l'utilisation des applications en temps réel, telles que la téléphonie et la vidéo haute définition, fait en sorte qu'une demande croissante en bande passante existe. Or, la fibre optique offre un très grand potentiel en termes de bande passante allant physiquement jusqu'à 50 terabits par seconde par fibre optique en utilisant un multiplexage en longueurs d'onde. La commutation de rafales en optique est une nouvelle technique de commutation permettant de tirer profit des avantages de la commutation de paquets et de la commutation de circuits. Une problématique majeure est le taux de perte élevé à cause du problème de la contention. Plusieurs techniques de résolution de la contention existent, notamment la déflexion et la retransmission.

Dans la première partie de ce travail, on s'intéresse à combiner dynamiquement la déflexion et la retransmission d'une manière adaptative en tenant compte de l'état du réseau. L'algorithme proposé permet d'effectuer le plus de déflexions possibles tant que celles-ci ne déstabilisent pas le réseau. Les résultats démontrent également, avec des topologies fortement connectées telles que COST239, que l'algorithme proposé permet de diminuer radicalement les pertes en utilisant un ratio de déflexion très grand.

La deuxième partie de ce travail porte sur la prévention de la contention en utilisant des tables de routage optimisées. Ces tables de routage sont optimisées en utilisant un modèle graphique probabiliste utilisé en intelligence artificielle, soit un réseau bayésien.

**Mots-clés :** *fibre optique, WDM, commutation de rafales, contention, déflexion, retransmission*

# Chapitre 1

## Introduction

Internet est omniprésent dans notre quotidien. L'explosion actuelle et future des applications en temps réel, telles que la téléphonie mobile et la vidéo haute définition, fait en sorte qu'une demande croissante en bande passante existe. Or, la fibre optique offre un très grand potentiel en termes de bande passante allant physiquement jusqu'à 50 terabits par seconde par fibre optique en utilisant un multiplexage en longueurs d'onde.

Une problématique centrale dans les réseaux optiques est l'absence de mémoire. Dans les routeurs standards électroniques actuels, lorsque deux paquets tentent de réserver au même moment le même lien, la mémoire *Random-Access Memory* (RAM) est utilisée pour construire des files d'attente pour retarder l'acheminement des paquets. Puisqu'une telle technologie n'existe pas en optique, le problème d'impossibilité de réserver une ressource (qu'on nomme contention) existe.

Deux approches générales pour optimiser les réseaux existent : ajouter ou modifier des composantes physiques et optimiser les mécanismes d'acheminement de données en utilisant les composantes actuelles. Dans le cadre de ce travail, la seconde approche est investiguée.

*Optical Burst Switching* (OBS) est une technique de commutation qui comporte les avantages de la commutation de paquets et de la commutation de circuits : adaptation rapide aux fluctuations du réseau, aucune nécessité de construire des files d'attente en mémoire, etc. OBS utilise deux types de paquet : le paquet de contrôle servant à réserver les longueurs d'onde et la rafale qui sert à transférer les données entièrement dans le domaine optique. Les rafales regroupent plusieurs petits paquets clients.

Le principal aspect à améliorer dans les réseaux OBS est le taux de perte des rafales. En effet, lorsqu'un paquet de contrôle n'est pas en mesure de réserver toutes les longueurs d'onde nécessaires de la source à la destination, le problème de la contention fait en sorte que des rafales sont perdues. La rafale de données est envoyée après un certain temps, qu'on nomme *Offset Time* (OT), sans attendre un accusé de réception.

## 1.1 Problématique

Un problème majeur dans les réseaux OBS est le taux élevé des pertes de rafales. Ces pertes sont essentiellement causées par les contentions qui surviennent lorsqu'un paquet de contrôle n'est pas en mesure de réserver toutes les longueurs d'onde nécessaires pour la rafale de données. Deux approches pour diminuer les contentions sont possibles : les résoudre d'une façon réactive au moment où elles proviennent et acheminer les rafales de manière à les prévenir et les éviter.

## 1.2 Contributions

L'originalité de ce travail peut être décomposée en deux parties :

1. *Combinaison adaptative de deux méthodes de résolution de la contention (Déflexion et retransmission)*. Plusieurs méthodes existent pour résoudre la contention : la conversion de longueur d'onde, l'utilisation de lignes à retardement, la déflexion, la segmentation

et la retransmission. Lorsque la charge du réseau est faible, la déflexion permet de diminuer fortement les pertes. Par contre, lorsque la charge devient grande, la déflexion déstabilise le réseau. Un algorithme est donc proposé pour combiner efficacement la déflexion et la retransmission d'une façon adaptative.

2. Prévention de la contention en utilisant un réseau bayésien pour construire des tables de routage pour acheminer les rafales efficacement.

Pour mesurer les performances des deux modèles proposés, des simulations sont effectuées sur les topologies NSFNET, COST239 et Pan-European. Ces simulations permettent d'illustrer la variation de plusieurs métriques (Taux des pertes, délai de bout en bout, utilisation des liens, etc.) en fonction de la charge du réseau.

# Chapitre 2

## État de l'art

### 2.1 Introduction

Dans ce chapitre, les réseaux optiques sont premièrement introduits. Les principales méthodes de multiplexage utilisées dans les réseaux optiques sont ensuite décrites. Finalement, les mécanismes de commutation de paquets, de circuits et de rafales sont expliqués.

### 2.2 Les réseaux optiques

#### 2.2.1 Introduction

Très bientôt, des systèmes transmettant des données entièrement en optique verront le jour, jouissant pleinement des avantages du spectre lumineux : rapidité et efficacité des processus quantiques [Tof01]. En télécommunication, le multiplexage en longueurs d'onde (*Wavelength-Division Multiplexing* (WDM)) permet d'échanger une quantité fantastique d'informations.

Une architecture typique d'un réseau public utilisant la fibre optique est composée de deux types de réseaux [RS01] (Fig. 2.1) :

- *Réseau longue distance* (Mieux connu sous le nom de *long-haul network*) : Il permet d'in-

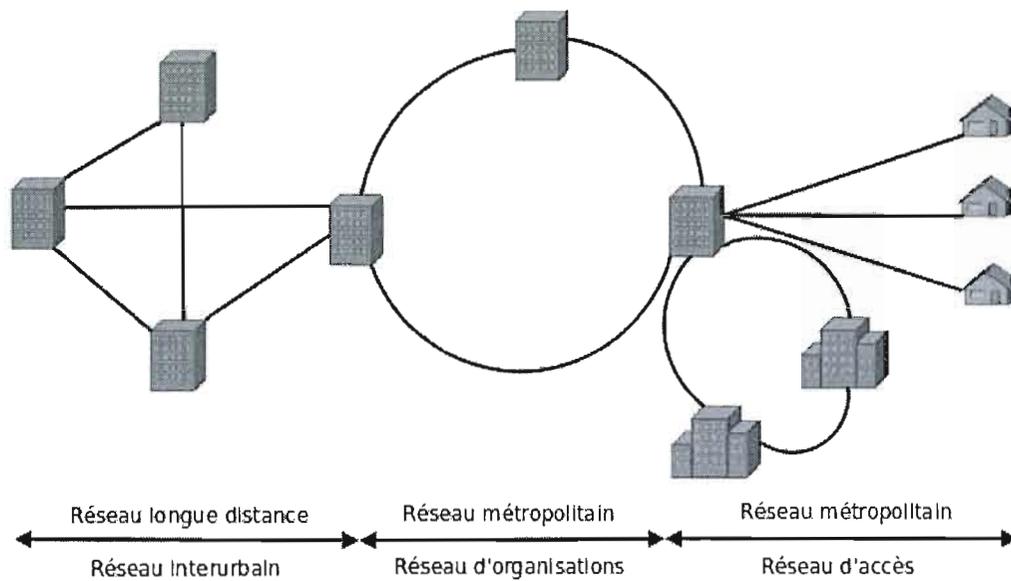


FIG. 2.1 – Parties d'un réseau public

terconnecter plusieurs villes ou régions. La distance entre les noeuds de ces réseaux varie et peut aller jusqu'à plusieurs milliers de kilomètres. Par exemple, la distance entre les villes en Europe a une moyenne de quelques centaines de kilomètres tandis qu'en Amérique du Nord les distances entre les grandes villes peuvent aller jusqu'à plusieurs milliers de kilomètres.

- *Réseau métropolitain* : Partie d'un réseau permettant d'interconnecter une grande ville. Le trafic du réseau d'accès est composé de trafic industriel et résidentiel permettant d'interconnecter des noeuds ayant des distances de quelques kilomètres. Le trafic du réseau d'accès est rassemblé au bureau central. Le réseau d'organisations permet de regrouper plusieurs bureaux centraux d'une même ville.

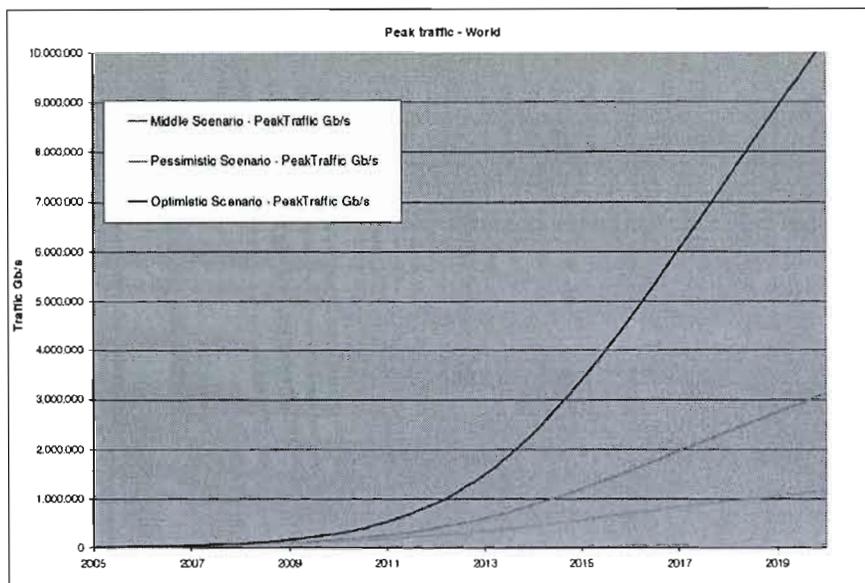


FIG. 2.2 – 3 scénarios possibles de la croissance du trafic (Provient de [Pay08])

### 2.2.2 Évolution

Selon plusieurs chercheurs, les connaissances et l'utilisation du génie électronique et du génie logiciel sont nettement supérieures aux connaissances en optique [Muk06]. La mémoire n'est présentement pas disponible en optique (outre des fibres optiques à retardement) tandis qu'en électronique la mémoire vive (RAM) est largement utilisée. Ceci est la cause primaire de la problématique du présent travail.

Selon une étude récente, l'augmentation futur du trafic pourrait croître de façon exponentielle [Pay08] (Fig. 2.2), notamment à cause de l'augmentation de la population et de l'augmentation de l'utilisation de la vidéo numérique haute définition. Le scénario pessimiste considère que les architectures réseaux présentes resteront les mêmes tandis que les deux autres scénarios considèrent que les architectures réseaux évolueront en tenant compte des nouveautés en recherche.

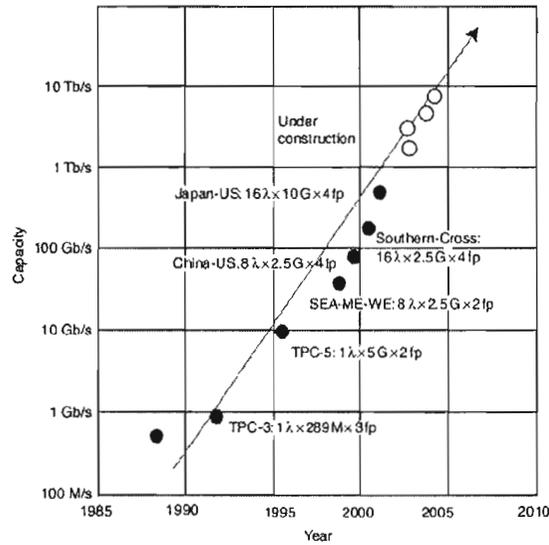


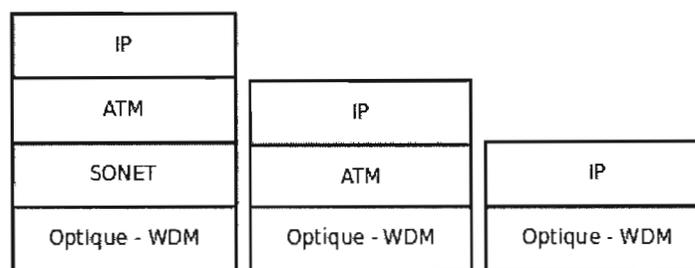
FIG. 2.3 – Évolution de fibres optiques WDM sous-marines (Provient de [DDF04])

Or, le médium optique offre un très grand potentiel pour obtenir une quantité énorme de capacité. Par exemple, une seule fibre optique utilisant un multiplexage en longueurs d'onde (WDM) a le potentiel de fournir 50 terabits par seconde de bande passante [JV04].

L'évolution de la capacité des fibres optiques WDM sous-marines (Fig. 2.3) croît de façon exponentielle tout comme le modèle de projection du trafic dans l'avenir (Fig. 2.2). Il existe deux méthodes pour optimiser le débit d'un réseau :

- Augmenter la capacité en faisant de la recherche sur les composantes physiques. Ces notions font partie du domaine de recherche de l'opto-électronique [Tof01].
- Optimiser les protocoles en utilisant les connaissances physiques actuelles et la capacité existante. Ces notions font partie du domaine de la téléinformatique.

La deuxième méthode d'optimisation des performances est utilisée dans le présent travail pour notamment augmenter le débit sortant, mais aussi pour diminuer le taux de perte.



(a) IP sur ATM sur SONET sur WDM (b) IP sur ATM sur WDM (c) IP sur WDM

FIG. 2.4 – Quelques architectures en couches proposées

### 2.2.3 Hiérarchie de couches de communication

Le modèle *Interconnexion de Systèmes Ouverts* (ISO) [Dea02] est composé de sept couches où chaque couche envoie des données à la couche immédiatement inférieure. Plusieurs architectures de couches ont été proposées en utilisant *Internet Protocol* (IP), *Asynchronous Transfer Mode* (ATM), *Synchronous Optical Networking* (SONET) [MS99] et la fibre optique avec un multiplexage WDM. Le protocole SONET utilise un multiplexage par le temps (TDM).

#### 2.2.3.1 IP sur ATM sur SONET sur WDM

Plusieurs fournisseurs déploient des réseaux ATM sur des réseaux SONET pour des utilisateurs IP (Fig. 2.4(a)). L'utilité de ATM est de permettre d'offrir aux utilisateurs de la qualité de service. Les paquets IP sont alors incorporés dans les cellules ATM et les cellules ATM sont ensuite mises dans des trames SONET.

Le coût primaire d'une telle infrastructure provient du fait que chaque protocole comporte une certaine taille pour l'en-tête ce qui fait en sorte que la proportion du trafic pour les données de contrôle devient très grande. Par exemple, l'en-tête IP est de minimalement 28 octets [Pos81] et l'en-tête ATM est de cinq octets.

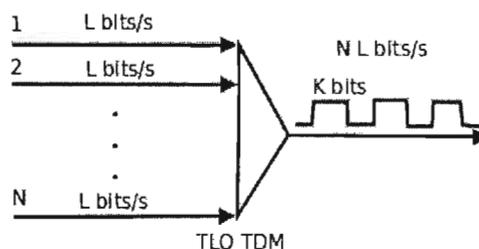


FIG. 2.5 – Multiplexage temporel

### 2.2.3.2 IP sur ATM sur WDM

L'approche IP sur ATM sur WDM (Fig. 2.4(b)) permet de diminuer le coût des en-têtes tout en permettant d'avoir une granularité fine de trafic grâce à ATM. Un inconvénient majeur est que plusieurs mécanismes comme la restauration de pannes doivent être implémentés au niveau de la couche optique WDM.

### 2.2.3.3 IP sur WDM

Selon plusieurs experts [Dix03], l'Internet de seconde génération sera basé selon le modèle IP sur WDM (Fig. 2.4(c)). Quelques applications concrètes récentes sont le *Voice IP* (VoIP) et le *Mobile IP* (MoIP) où les paquets IP des applications en temps réel passent notamment par un réseau haut débit optique WDM.

## 2.2.4 Multiplexage

Le multiplexage est l'opération qui consiste à combiner plusieurs canaux sur un même médium physique (par exemple la fibre optique). Il existe deux principales catégories de multiplexage : le multiplexage temporel et le multiplexage en longueurs d'onde.

### 2.2.4.1 Multiplexage temporel

Le multiplexage temporel consiste à séparer la fibre optique du port de sortie en  $N$  canaux de temps identiques, donnant  $N$  canaux de  $K$  bits (Fig. 2.5). Les ports d'entrées correspondent

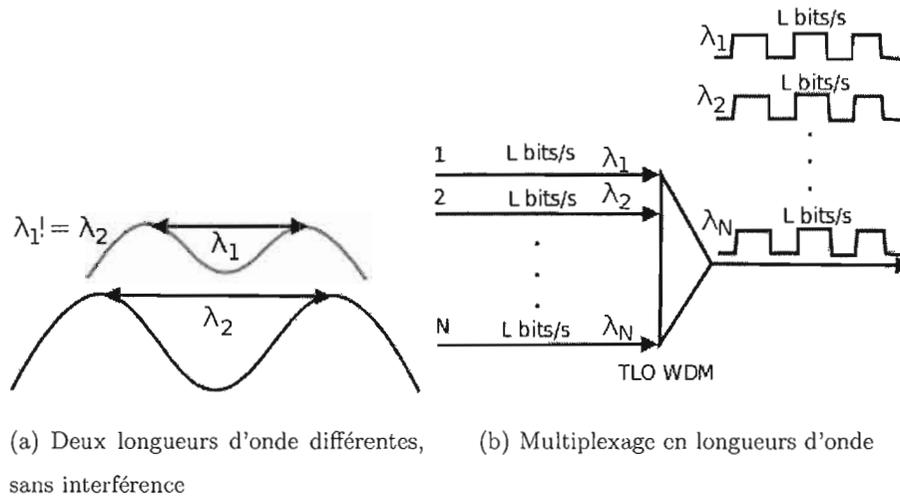


FIG. 2.6 – Multiplexage en longueurs d'onde

à des flux de données qui sont multiplexés sur le même support physique. À chaque moment donné, un seul flux envoie des données vers le port de sortie d'une façon équitable. Des expériences en laboratoire [RS01] ont expérimenté un multiplexage/démultiplexage temporel en optique (OTDM) allant jusqu'à 250 G bits/s.

Un *Terminal de Ligne Optique* (TLO) est un élément principal d'un réseau WDM, il est utilisé pour effectuer le multiplexage et le démultiplexage. Les composantes physiques d'un TLO sont le transpondeur, le multiplexeur de longueurs d'onde et l'amplificateur. Le transpondeur sert à adapter un signal provenant d'un client en un signal valide pour le réseau optique. Au besoin, le signal est amplifié par l'amplificateur. Pour plus de détails, se référer à [RS01].

#### 2.2.4.2 Multiplexage en longueurs d'onde

Le multiplexage en longueurs d'onde (WDM) consiste à transmettre plusieurs flux simultanément sur la même fibre optique en envoyant plusieurs rayons ayant des longueurs d'onde différentes (Fig. 2.6(a)). Donc, plusieurs canaux sont alors actifs simultanément ( $\lambda_1, \lambda_2, \dots, \lambda_N$ ).

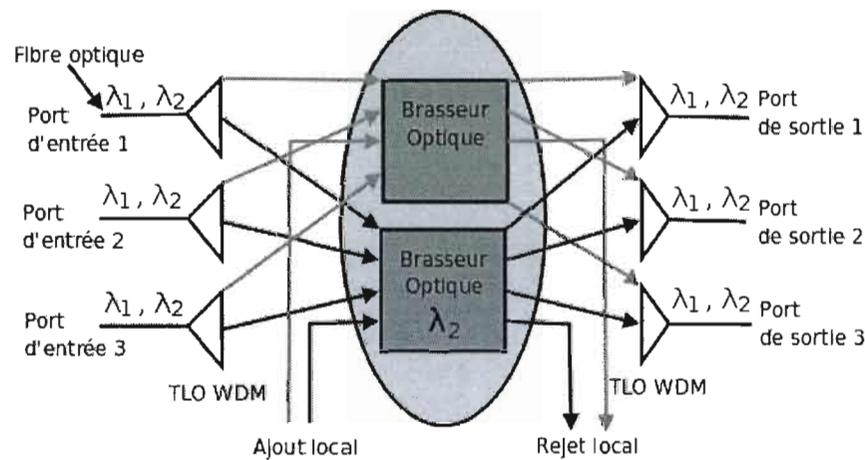


FIG. 2.7 – Brasseur optique 3x3 avec 2 longueurs d'onde

Un brasseur optique (*Optical Cross-Connect* (OXC)) complet doit être configuré pour multiplexer et démultiplexer les bons ports selon le routage voulu (Fig. 2.7). Il comporte les fonctionnalités de routage. Un brasseur optique comprend un certain nombre  $N$  de ports d'entrée et de ports de sortie ( $M$ ). Ainsi,  $N$  TLO WDM permettant le démultiplexage et  $M$  TLO WDM permettant de multiplexage sont alors nécessaires. Dans l'exemple,  $N = M = 3$ . De plus, un brasseur optique supporte un certain nombre  $W$  de longueurs d'onde. Dans l'exemple,  $W = 2$ . Un flux de données peut être localement rejeté ou ajouté, ce qui rend le brasseur optique plus flexible.

En configurant plusieurs brasseurs optiques, une connexion logique peut alors être établie à travers un réseau tout optique WDM (Fig. 2.8). Les clients peuvent typiquement être de n'importe quel type (IP, ATM, SONET, etc.) puisque le réseau tout optique WDM ne se préoccupe pas de la nature des données qu'il véhicule.

### 2.2.5 Composantes physiques pour commuter en optique

La commutation des données est une fonctionnalité primaire et a un très grand impact sur les performances d'un réseau optique en terme, notamment, de délai de bout en bout

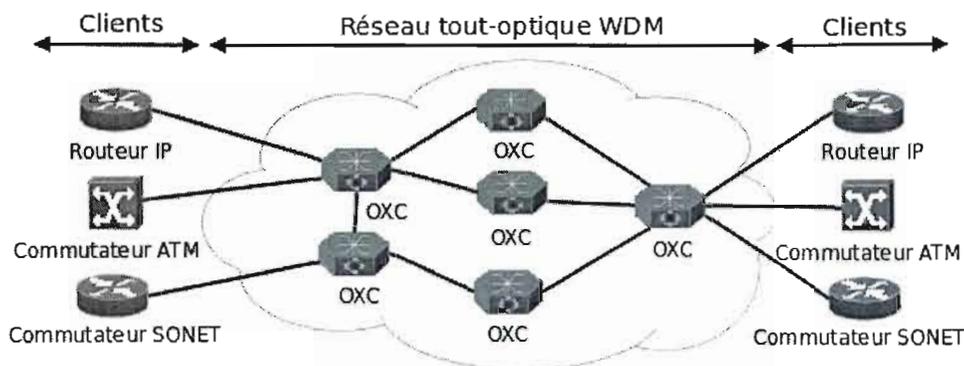


FIG. 2.8 – Exemple d'architecture de réseau tout optique WDM

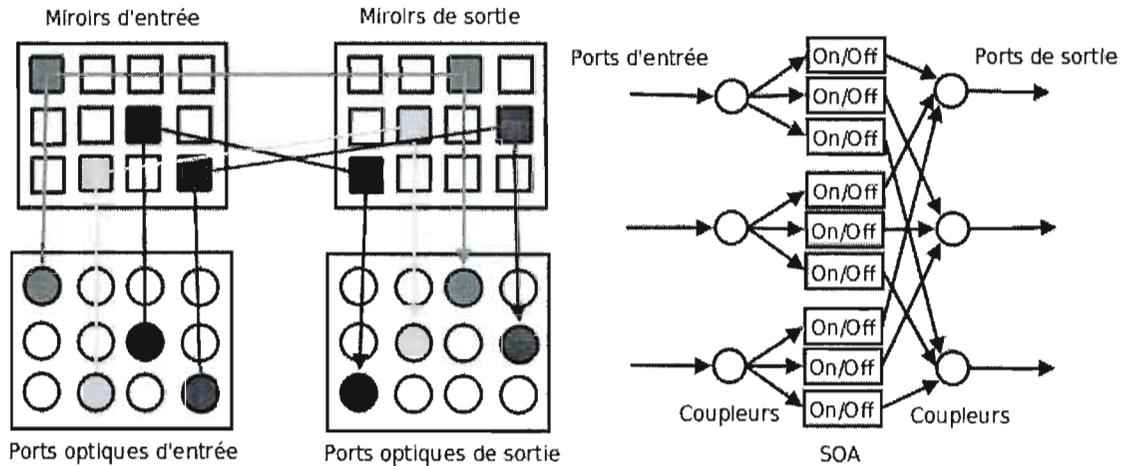
[JV04]. Il existe deux technologies matures permettant de gérer cette fonctionnalité : les microsystèmes électromécaniques et les amplificateurs à semi-conducteur.

#### 2.2.5.1 Microsystème électromécanique

Les microsystèmes électromécaniques (*Micro-Electromechanical systems (MEMS)*) sont une technologie mature composées de petits miroirs pouvant être configurés. Dans l'exemple de la Fig. 2.9(a), 4 paires de miroirs ont été configurées lorsque quatre paquets de contrôle ont spécifié chacun une certaine configuration pour permettre d'envoyer les données à la destination. Le temps de configuration est d'approximativement 50 ms.

#### 2.2.5.2 Amplificateur à semi-conducteur

Les amplificateurs à semi-conducteur (*Semiconductor Optical Amplifier (SOA)*) peuvent être utilisés pour laisser passer ou non la lumière (On/Off) (Fig. 2.9(b)). Ainsi, si un SOA est configuré à *On*, cette composante laisse passer la lumière. Un coupleur permet de combiner ou séparer un signal. Par exemple, l'exemple de la figure 2.9(b) comporte 3 coupleurs 1x3 et 3 coupleurs 3x1. Un coupleur 1x3 envoie le tiers du signal original vers chaque SOA. Ensuite le SOA va amplifier le signal au besoin. Le multiplexage se fait donc en configurant correctement les SOA.



(a) Commutateur avec microsystemes (b) Commutateur avec amplificateurs à semi-conducteur  
 électromécaniques - 4 flux de données

FIG. 2.9 – Commutateurs tout optique

## 2.2.6 Paradigmes de commutation

Un paradigme de commutation définit l'envoi, la façon de fragmenter les données et la signalisation. La commutation a un très grand impact sur les performances des réseaux optiques actuels puisque la mémoire tampon n'est présentement pas disponible. Dans les routeurs électroniques utilisant la commutation de paquets, le mécanisme *store and forward* est utilisé : les paquets qui ne peuvent pas être immédiatement acheminés sur le port de sortie désiré sont stockés dans une file d'attente dans une mémoire vive. Or, ceci ne fait pas de sens dans les réseaux optiques actuels où seulement des fibres optiques à retardement ayant des capacités limitées sont disponibles. Les paradigmes de commutation investigués dans la littérature sont les suivants :

- Commutation de paquets : Le flux de données est fragmenté en une multitude de paquets et ces paquets sont envoyés sur le réseau sans nécessairement passer par le même chemin.
- Commutation de circuits : Correspond au paradigme utilisé pour les communications téléphoniques où le circuit de la source à la destination est réservé avant de débiter la

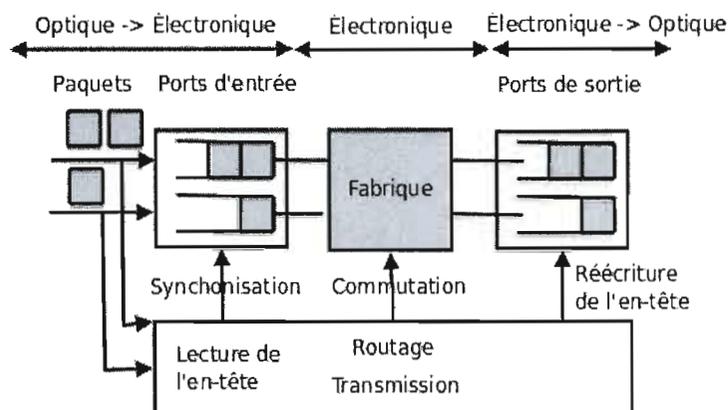


FIG. 2.10 – Architecture d'un routeur OPS

connexion.

- Commutation de rafales : Est l'intermédiaire entre la commutation de paquets et la commutation de circuits. L'idée est de tirer profit des avantages des deux paradigmes comme il sera expliqué dans la section 2.3.

### 2.2.6.1 Commutation de paquets

La commutation de paquets en optique (*Optical Packet Switching (OPS)*) utilise une granularité très fine puisque le flux de données est fragmenté en petits paquets. Les paquets, pour une même connexion, peuvent utiliser des chemins différents. L'avantage primaire de OPS est la granularité fine des paquets puisqu'une telle granularité permet aux routeurs de s'adapter facilement aux fluctuations du trafic.

Un routeur utilisant la commutation de paquets doit fournir les fonctionnalités de routage, de synchronisation, de file d'attente, de gestion des en-têtes, etc. (Fig. 2.10) [RS01]. Ainsi, tous les paquets doivent être convertis en électronique pour être traités et pour être ensuite reconvertis dans le domaine optique. Les deux inconvénients majeurs de OPS sont la nécessité d'une mémoire tampon et un très grand sur-débit causé par le temps de traitement des paquets dans le domaine électronique.

### 2.2.6.2 Commutation de circuits

En commutation de circuits en optique (*Optical Circuit Switching* (OCS)) [Aca94], seulement les paquets de contrôle doivent effectuer une conversion *Optique* – > *Électronique* – > *Optique* (O/E/O). La notion de canaux virtuels est très importante : un paquet de contrôle est envoyé à un ensemble de noeuds pour construire une route, soit un canal virtuel. Après la réservation effectuée par le paquet de contrôle, le canal virtuel complet est réservé au complet pendant un temps désiré. Une première problématique est la réservation des longueurs d'onde (*Routing and Wavelength Assignment* (RWA)) [Rou01]. Si aucun convertisseur de longueur d'onde n'est disponible, la même longueur d'onde doit être réservée pour chaque lien réservé sur le chemin virtuel. Le paradigme OCS comporte plusieurs inconvénients :

1. Faible utilisation : Une connexion peut être longue sans forcément que les longueurs d'onde soient effectivement utilisées. Les ressources ne peuvent donc pas être exploitées pleinement.
2. Grand temps de configuration : OCS utilise le protocole de signalement *Tell-And-Wait* (TAW) [JV04], où un paquet de contrôle doit réserver de la source à la destination et revenir ensuite à la source avant qu'un seul bit de données du client soit envoyé.
3. Faible adaptabilité aux fluctuations du trafic.

### 2.2.6.3 Commutation de rafales

La commutation de rafales en optique (OBS) [JV04] utilise le protocole de signalisation *Just-Enough-Time* (JET) (Fig. 2.11). Ce protocole envoie un paquet de contrôle, nommé *Burst Header Packet* (BHP), pour réserver les longueurs d'onde pour une rafale qui est envoyée après un certain temps, nommé OT.

La commutation de rafales est un paradigme qui se situe entre OCS et OPS : OBS tend à tirer profit des avantages de OCS et OPS, soit une forte utilisation de la bande passante, un faible temps de configuration, aucune nécessité d'avoir une mémoire tampon, un sur-

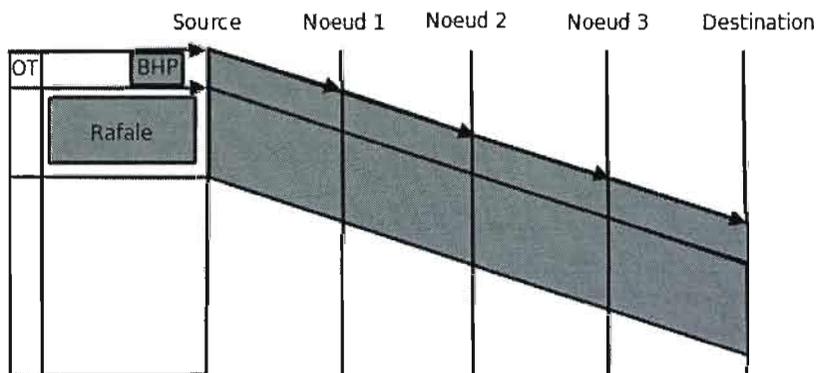


FIG. 2.11 – Signalisation OBS JET

débit faible et une forte adaptabilité aux fluctuations de l'intensité du trafic. Tout comme avec OCS, seul le paquet de contrôle doit effectuer une conversion O/E/O, les données étant transmises dans le domaine optique seulement.

Une problématique primaire dans les réseaux OBS, celle investiguée dans le cadre de ce travail, est la contention [Gau02]. Lorsqu'un BHP tente de réserver une longueur d'onde à un temps donné, mais qu'il est impossible de la réserver puisqu'un autre BHP a déjà réservé cette longueur d'onde à ce temps donné, il se produit alors une contention. Il existe plusieurs méthodes pour résoudre une contention. Ces méthodes seront revues dans la Section 2.3.4.

#### 2.2.6.4 Performances

<i>Paradigme</i>	<i>Utilisation</i>	<i>Temps de configuration</i>	<i>Tampon</i>	<i>Sur-débit</i>	<i>Adaptabilité</i>
<i>Paquet</i>	Grande	Petit	Nécessaire	Grand	Grand
<i>Circuit</i>	Petite	Grand	Non nécessaire	Petit	Petite
<i>Rafale</i>	Grande	Petit	Non nécessaire	Petit	Grand

TAB. 2.1 – Comparaison des paradigmes de commutation (Provient de [QY99])

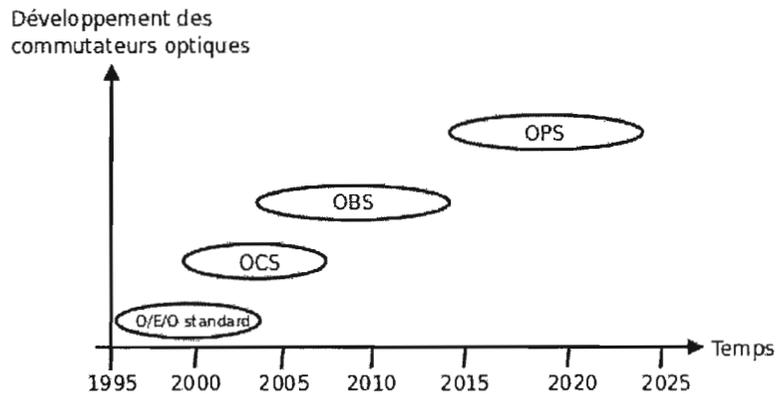


FIG. 2.12 – Évolution des tendances de développement des commutateurs optiques

Le tableau 2.1 résume les caractéristiques des performances des 3 paradigmes de commutation. OBS tire profit des avantages de OPS et OCS. Aucune mémoire tampon n'est nécessaire. Or, aucune mémoire tampon n'est présentement disponible ce qui en résulte que OCS et OBS sont les seuls paradigmes de commutation présentement viables. OBS offre également une très grande utilisation de la bande passante avec une forte adaptabilité compte tenu de la granularité fine des rafales. Le temps de configuration est également faible puisqu'une rafale est envoyée sans attendre un accusé de réception. La signalisation se fait grâce à un BHP préalablement envoyé.

Xue et al. [XYYH05] ont comparé les performances de OBS et OCS en utilisant le même trafic sur la topologie *National Science Foundation Network* (NSFNET). OBS a obtenu de meilleurs résultats en termes de taux de perte et en termes de débit sortant. OBS a obtenu un débit sortant moyen de 20 % supérieur à celui de OCS.

OBS est présentement la prochaine tendance au niveau des développements des commutateurs optiques [Muk06] (Fig. 2.12).

Puisque OBS est présentement la meilleure alternative dans les réseaux optiques au niveau de la commutation, c'est ce paradigme qui est investigué dans le cadre de ce travail.

## 2.3 Les réseaux optiques à commutation de rafales

### 2.3.1 Introduction

Les réseaux optiques à commutation de rafales (OBS) [JV04, QY99, Gau06, CQY04, Tur99, Vok04] sont une technologie prometteuse pour gérer le trafic sporadique et dynamique d'Internet. Dans les réseaux OBS, les données clientes (IP par exemple) sont assemblées dans un grand segment nommé rafale de données qui est envoyée entièrement en optique de la source à la destination. La rafale de données est précédée temporellement par un paquet de contrôle, appelé BHP, qui est envoyé sur un canal de contrôle (une longueur d'onde). À chaque noeud, une conversion O/E/O est effectuée. Le temps de différence entre la rafale de données et le BHP se nomme OT. Le BHP effectue la réservation des longueurs d'onde à chaque noeud vers la destination pour que la rafale puisse ultérieurement utiliser les ressources réservées. Lorsque le BHP arrive à un commutateur, la capacité est réservée pour la rafale. Si la capacité demandée est disponible au temps où la rafale va arriver, alors la rafale peut passer sans problème entièrement en optique sans aucune conversion vers le domaine électronique.

Puisque les rafales de données et les BHP sont envoyés sans aucun paquet d'acquittement pour confirmer la réservation, la rafale de données peut être rejetée si OT est insuffisant ou si le BHP n'a pas été en mesure de réserver toutes les ressources nécessaires, bref s'il y a eu contention. Or, il est clair que les mécanismes de résolution de la contention sont cruciaux pour diminuer les pertes.

Ce chapitre introduit les réseaux OBS. L'architecture des réseaux OBS est tout d'abord définie. Ensuite, les principaux mécanismes de signalisation sont présentés. Finalement, une revue des mécanismes de résolution de la contention sont présentés. La déflexion et la re-

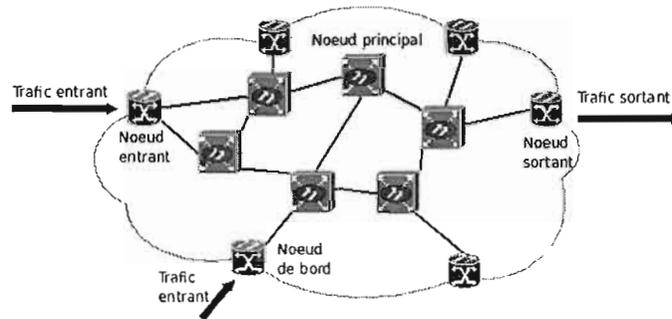


FIG. 2.13 – Architecture d'un réseau OBS

transmission sont particulièrement expliqués en détail, concepts qui sont le sujet primaire du présent travail.

### 2.3.2 Architecture

Un réseau OBS est composé de nœuds interconnectés de liens optiques utilisant le multiplexage WDM. Il existe deux types de nœud : les nœuds de bord (*Edge Nodes*) et les nœuds principaux (*Core Nodes*).

**Les nœuds de bord (Edge Nodes)** Type de nœud utilisé pour construire les rafales de données en regroupant un certain nombre de paquets provenant de clients.

Il existe deux approches pour déterminer la taille des rafales [JV04]. La première approche est celle basée sur le temps. Cette approche consiste à construire une rafale à chaque intervalle de temps, la taille de la rafale peut donc varier d'une façon significative à travers le temps selon la charge du trafic. La seconde approche consiste à utiliser un seuil du nombre de paquets à mettre dans une rafale. Le nombre de rafales envoyées par seconde varie donc selon la charge du réseau.

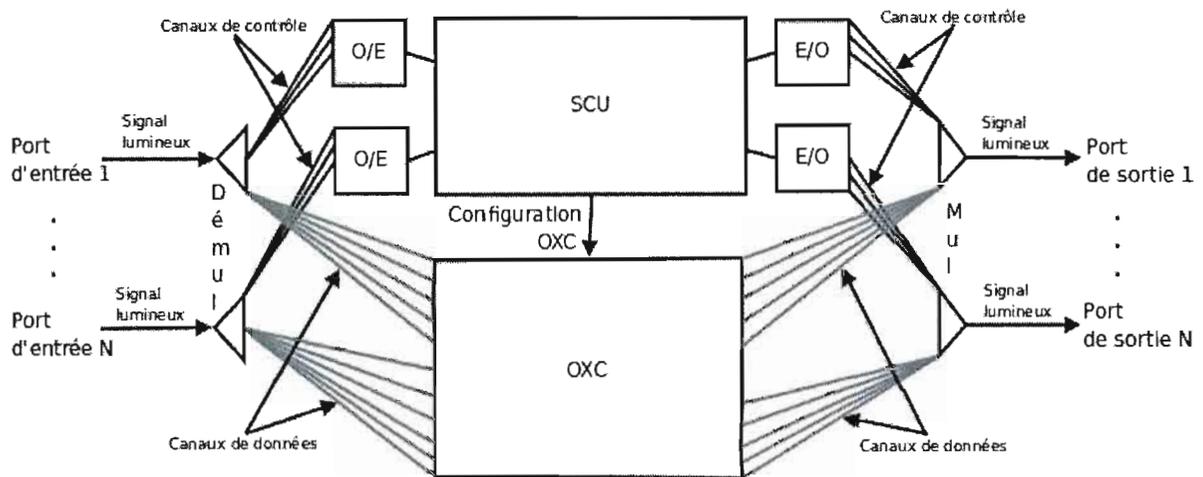


FIG. 2.14 – Noeud principal OBS

Une autre fonctionnalité primaire est l'estimation du OT. Le OT est le temps que prend le paquet de contrôle pour réserver les ressources de la source à la destination. La rafale est donc envoyée après ce temps.

**Les noeuds principaux (Core Nodes)** Type de noeud utilisé pour commuter les rafales et pour résoudre au besoin la contention [JV04]. Un noeud principal comporte deux composantes primaires (Fig. 2.14) : un OXC et un commutateur de contrôle (*Switch Control Unit* (SCU)). Lorsque le SCU reçoit un BHP, l'en-tête est interprétée et les champs sont extraits. Le SCU configure le OXC en fonction de la destination. Si le port de sortie est disponible au temps où la rafale passe, alors le port de sortie est réservé. Le port de sortie est sélectionné selon l'algorithme de routage et la destination.

Un lien optique comporte  $\eta$  longueurs d'onde ( $\lambda_1, \lambda_2, \dots, \lambda_\eta$ ). Un canal, soit un flux de données, est associé à une longueur d'onde. Or,  $K$  canaux de contrôle sont utilisés pour acheminer les BHP et  $L$  canaux de données permettent aux rafales d'être transmises sur le réseau OBS, où  $L + K = \eta$ . Normalement,  $L > K$  puisque les paquets de contrôle sont fortement plus petits que les rafales de données et conséquemment un nombre moindre de

canaux de contrôle est nécessaire comparativement au nombre de canaux de données.

**Les noeuds hybrides** Une architecture OBS nécessite que tous les noeuds sont à la fois des noeuds de bord et des noeuds principaux [JV04]. Ainsi, tous les noeuds peuvent injecter du trafic, mais aussi acheminer correctement les rafales dans le réseau OBS.

### 2.3.3 Signalisation

Un protocole de signalisation est le processus permettant de réserver les ressources [Muk06]. En OBS, un paquet de contrôle est préalablement envoyé de la source vers la destination.

À chaque noeud parcouru sur la route, un protocole de réservation de canaux de données est appliqué. Le plus considéré dans la littérature est *Latest Available Unscheduled Channel with Void Filling* (LAUC-VF) [TGCT99] qui consiste à conserver les temps d'arrivée et de départ des rafales en réservant les canaux de manière à minimiser les espaces vides en réservant les canaux seulement pendant le temps où la rafale est commutée sur le OXC. Pour plus de détails sur les protocoles de réservation de canaux, voir [JV04].

Les protocoles de signalisation peuvent être classés en trois catégories en utilisant la direction comme métrique de différenciation : à un sens, à deux sens et hybride.

#### 2.3.3.1 Protocoles à une direction

**Juste assez de temps (*Just-Enough-Time* (JET))** Le protocole JET [YQ97] est le protocole de signalisation le plus employé dans la recherche sur les réseaux OBS. Un BHP est préalablement envoyé sur un canal de contrôle pour réserver les canaux de données. Après un OT, la rafale de données est envoyée sur les canaux de données réservés. Le OT nécessaire pour le protocole JET dépend du nombre de sauts  $N$  de la source à la destination, du temps de traitement du paquet de contrôle  $t_t$  ainsi que du temps de configuration du commutateur  $t_c$  (Fig. 2.15) [Vok04] :

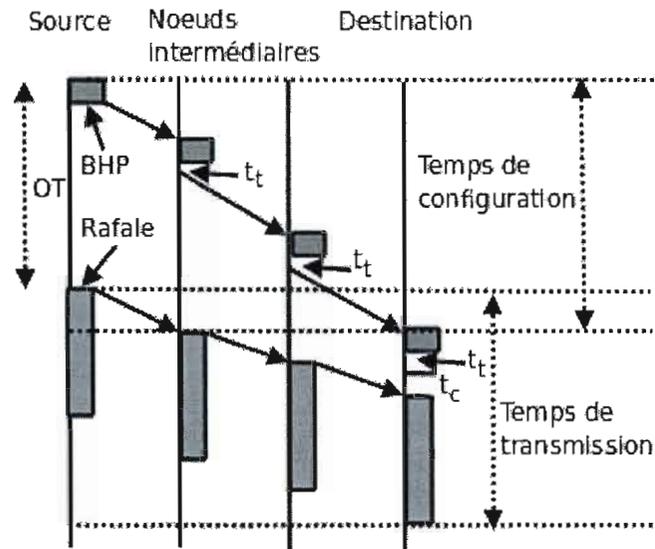


FIG. 2.15 – Signalisation OBS JET à une direction

$$OT_{JET} = N * t_t + t_c \quad (2.1)$$

Avec JET, le délai de bout en bout ( $T_{JET}$ ) est exprimé par la somme du temps d'agrégation des rafales ( $t_a$ ), le OT ( $OT_{JET}$ ), le temps de transmission ( $t_r$ ) et le temps de propagation du lien  $i$  ( $t_p^i$ ) [JV04] :

$$T_{JET} = t_a + OT_{JET} + t_r + \sum_{i=1}^N t_p^i \quad (2.2)$$

où  $N$  représente le nombre de liens de la source vers la destination.

Le taux de perte élevé est la problématique majeure de ce protocole. Une perte survient lorsqu'un BHP n'arrive pas à réserver les longueurs d'onde nécessaires.

**Juste à temps (*Just-in-Time* (JIT))**

Le protocole *Just-In-Time* (JIT) [JV04] est similaire au protocole de signalisation JET. Une première différence notable est que JIT utilise une réservation immédiate, c'est-à-dire que lorsque le BHP passe à un noeud, le canal de données est immédiatement réservé jusqu'à ce que la rafale associée ait passée, comparativement à JET où le canal est réservé seulement pendant le temps que la rafale a besoin du canal. De plus, JIT emploie une méthode de libération explicite, c'est-à-dire qu'un paquet de contrôle doit parcourir les noeuds de la destination vers la source pour libérer les ressources après que la rafale soit passée.

***Tell-and-Go* (TAG)** Le protocole *Tell-And-Go* (TAG) [DL01] se distingue de JET par le fait qu'il n'y a aucun OT. Puisqu'il n'y a pas de OT et que conséquemment les ressources ne sont pas nécessairement réservées lorsque la rafale arrive à un noeud donné, des fibres optiques de ligne à retardement (*Fiber Delay Line* (FDL)) doivent être utilisées (Voir Section 2.3.4.2 pour le fonctionnement des lignes à retardement) pour mettre la rafale dans un tampon en attendant que les ressources soient réservées. Puisque l'utilisation des lignes à retardement sont limitées, ce protocole de signalisation n'est présentement pas viable.

### 2.3.3.2 Protocoles à deux directions

***Tell-and-Wait* (TAW)** Le protocole TAW [DL01] (Fig. 2.16) utilise une direction à 2 sens : de la source vers la destination pour réserver les ressources et lors de l'envoi de la rafale et une direction de la destination vers la source pour confirmer la réservation. Il existe donc trois types de paquet de contrôle dans ce protocole : un BHP de configuration, un BHP de confirmation et un BHP de libération. Le délai de bout en bout du protocole TAW correspond à la somme du temps d'agrégation de la rafale, le temps de transmission de la rafale ainsi que le temps de 3 *aller/retour* (AR) de la source à la destination ou de la destination à la source :

$$T_{TAW} = t_a + t_r + 3 * \sum_{i=1}^N t_p^i \quad (2.3)$$

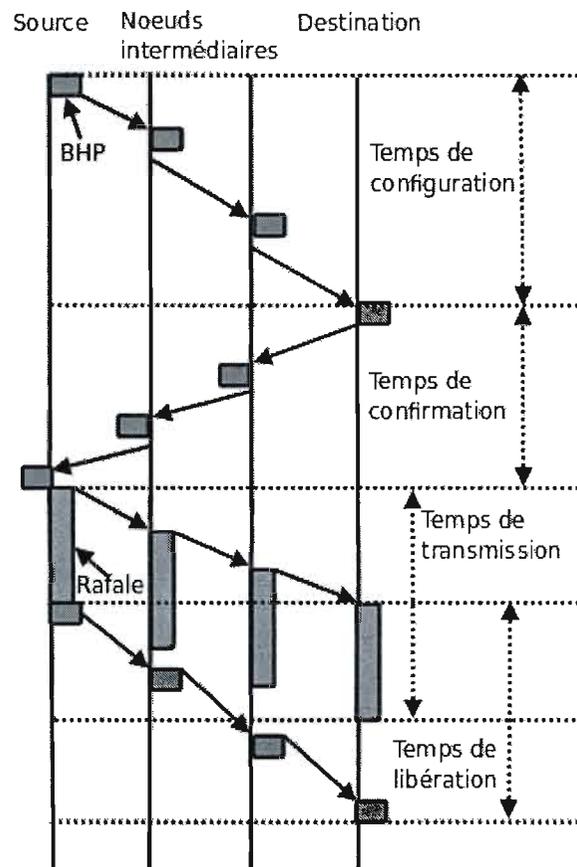


FIG. 2.16 – Signalisation OBS TAW à deux directions

Le délai de bout en bout de TAW ( $T_{TAW}$ ) est donc strictement supérieur à celui de JET ( $T_{JET}$ ). Par contre, l'avantage majeur de TAW est qu'il génère un taux de perte plus faible que JET.

### 2.3.3.3 Résumé des protocoles de signalisation

<i>Protocole</i>	<i>Direction</i>	<i>Initialisation</i>	<i>Réservation</i>	<i>Libération</i>	<i>Délai</i>	<i>Taux de perte</i>
<i>JET</i>	1 sens	Source	Implicite	Implicite	Petit	Grand
<i>JIT</i>	1 sens	Source	Explicite	Explicite	Petit	Grand
<i>TAG</i>	1 sens	Source	Implicite	Implicite	Petit	Grand
<i>TAW</i>	2 sens	Source ou Dest.	Explicite	Explicite	Grand	Petit

TAB. 2.2 - Caractéristiques des protocoles de signalisation (Provient de [JV04])

Le tableau 2.2 résume les caractéristiques des principaux protocoles de signalisation. Un protocole de signalisation peut initialiser la réservation soit par la source vers la destination (*Source Initiated Reservation* (SIR)) ou de la destination vers la source (*Destination Initiated Reservation* (DIR)). Lorsque la réservation est implicite, cela veut nécessairement dire que l'initialisation s'effectue par la source.

On remarque qu'il y a un compromis inévitable entre le délai et le taux de perte. Les protocoles à 1 sens ont un faible délai de bout en bout compte tenu qu'un seul AR est nécessaire comparativement à 3 AR pour les protocoles à 2 sens. De plus, les protocoles avec une direction à 1 sens ont un grand taux de perte puisque la probabilité qu'un paquet de contrôle n'arrive pas à réserver entièrement un chemin virtuel optique est plus grande comparativement à lorsque la source est notifiée après que le BHP ait réservé les ressources de la source à la destination. Il n'y a donc pas de protocole ayant naturellement un faible délai de bout en bout ainsi qu'un faible taux de perte, ce qui serait l'idéal.

Dans le cadre de ce travail, on s'intéresse à diminuer le taux de perte tout en conservant au maximum le faible délai de bout en bout du protocole de signalisation JET.

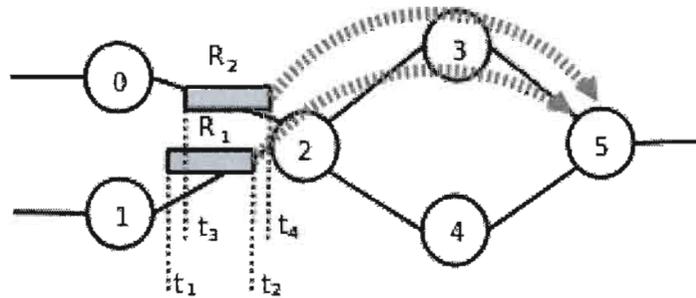


FIG. 2.17 – Deux rafales de données en contention sur un lien optique WDM

## 2.3.4 Méthodes de résolution de la contention

### 2.3.4.1 Introduction

Une problématique majeure dans les réseaux OBS est le problème de la contention [Met04]. La contention se produit lorsqu'un BHP tente de réserver une longueur d'onde  $\lambda$  pendant un certain intervalle de temps  $i_1 = [t_1..t_2]$  pour une rafale  $R_1$ , mais que la réservation ne peut pas s'effectuer puisqu'il existe un autre BHP ayant réservé la même  $\lambda$  au même noeud pour une rafale  $R_2$  pendant un certain intervalle de temps  $i_2 = [t_3..t_4]$  tel que  $i_1$  et  $i_2$  se chevauchent (Fig. 2.17) :

(2.4)

$$Contention(R_1 \rightarrow i_1, R_2 \rightarrow i_2) = \begin{cases} Vrai & \text{Si } \exists t \in \mathbb{D}t \in i_1, t \in i_2 \\ Faux & \text{Sinon} \end{cases}$$

Si la mémoire optique était présentement possible, cette problématique n'existerait pas puisque des files d'attente comme dans les routeurs électroniques actuels pourraient être utilisées. Par contre, une telle mémoire optique n'existe présentement pas (Voir Section 2.3.4.2).

Or la contention peut être résolue selon trois dimensions [Gau02] :

- *Domaine temporel* : En utilisant les lignes à retardement pour retarder les rafales.
- *Domaine spectral* : En utilisation la conversion de longueur d'onde : si une longueur d'onde n'est pas disponible, il suffit de convertir les données d'une longueur d'onde dans une autre longueur d'onde disponible.
- *Domaine spatial* : En utilisant soit la déflexion, qui consiste à envoyer une rafale vers un lien et un chemin différent en utilisant les ressources disponibles, ou soit en utilisant la retransmission, qui consiste à retransmettre une rafale de la source lorsqu'elle est perdue. La segmentation de rafales entre également dans cette catégorie, technique qui consiste à effectuer une déflexion seulement avec la partie de la rafale qui est en contention.

#### 2.3.4.2 Lignes à retardement

La mémoire n'existe présentement pas en optique. Par contre la mémoire peut être simulée en faisant circuler les données dans une fibre optique, qu'on appelle ligne à retardement (FDL). Les FDLs peuvent avoir plusieurs utilités dans les commutateurs optiques [RS01], on s'intéresse cependant à l'utilisation potentielle de retarder des rafales avec le protocole de signalisation JET. 200 mètres de FDL permettent de retarder 10 kbits pendant 1 microseconde ( $1 \mu s$ ) [RS01]. Ceci pose donc un problème : on ne peut pas utiliser cette technologie pour temporiser des milliers de rafales faisant des centaines de kbit/s.

Si l'on considère une FDL assez flexible, après la pré-configuration du BHP, si une longueur d'onde d'un lien d'un noeud intermédiaire n'a pas pu être réservé, il est alors possible de retarder la rafale pendant un certain temps et de le mettre sur le port de sortie voulue pour l'acheminer vers la destination (Fig. 2.18). Cette technique n'est cependant pas viable compte tenu de la limitation des FDLs et du nombre potentiellement très grand de rafales à temporiser.

Une technique plus intéressante consiste à retarder le BHP en utilisant une partie du OT pour tenter une nouvelle réservation ultérieurement, sans pour autant que la rafale rattrape

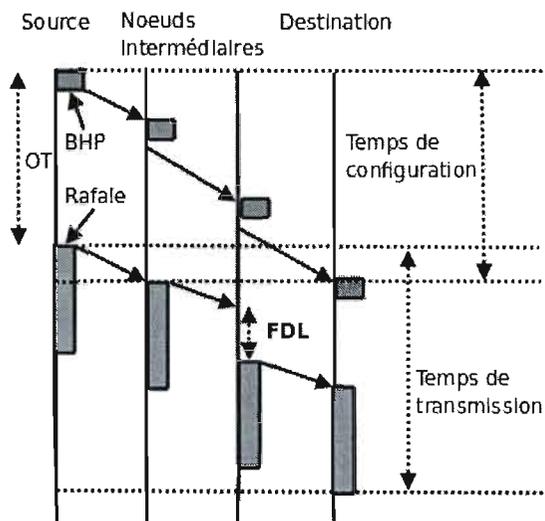


FIG. 2.18 – Signalisation JET avec utilisation d'une FDL pour retarder une rafale

le BHP [GBPS03].

### 2.3.4.3 Conversion de longueur d'onde

Si une rafale est en contention avec une certaine longueur d'onde  $\lambda_1$ , alors en utilisant un convertisseur de longueurs d'onde il est possible d'en utiliser une autre  $\lambda_2$  qui n'est pas en contention [RS01]. Selon les expériences de Gauger et al. [GKS04], lorsque la charge est élevée, la conversion seule est meilleure en termes de pertes qu'en utilisant la conversion et la déflexion (dans l'ordre). Ils ont considéré que la conversion doit toujours être la première méthode de résolution de la contention.

### 2.3.4.4 Segmentation de rafales

La segmentation de rafales consiste tout d'abord à diviser une rafale en segments [Vok01]. Initialement, la rafale entière est envoyée. Lorsqu'une contention se produit, les segments en contention sont défléchis. La déflexion peut donc être considérée comme étant une segmentation dont tous les segments sont défléchis au même endroit. Lorsque la charge est petite, la

déflexion obtient moins de pertes qu'avec la segmentation. Par contre, lorsque la charge est grande, la segmentation obtient moins de pertes comparativement à la déflexion.

#### 2.3.4.5 Déflexion

La déflexion consiste à résoudre la contention en empruntant un lien alternatif en utilisant la même longueur d'onde [WMA00b]. L'algorithme 1 définit la résolution de la contention grâce à la déflexion, sans convertisseur de longueurs d'onde. Cette section décrit la méthode de la déflexion OBS, illustre les principaux mécanismes développés dans la littérature et se termine par une catégorisation complète des travaux faits jusqu'à maintenant.

**Avantages de la déflexion** Wang et al. [WMA00b] ont dénoté plusieurs avantages d'utiliser la déflexion pour résoudre la contention :

- La déflexion est utile dans la mesure où lorsqu'une rafale est rejetée, il y a un gaspillage de la bande passante puisqu'il faut retransmettre à nouveau la rafale de la source à la destination si un mécanisme de retransmission OBS est disponible.
- Plus la distance des routes est grande, plus il est coûteux de ne pas effectuer de déflexion.
- La déflexion permet de diminuer le taux de perte, qui est le problème majeur en OBS, due à la contention et au manque de mémoire.

À cela, on peut ajouter le fait que la déflexion est une méthode de résolution de la contention gratuite en termes de composantes physiques, puisque la déflexion consiste à utiliser de la bande passante existante. Si l'on compare cette technique avec FDL ou la conversion de longueur d'onde, la FDL est coûteuse et limitée et un convertisseur de longueurs d'onde est une composante physique qui a également un certain coût.

#### Limitations de la déflexion

La déflexion comporte quelques limitations :

---

**Algorithm 1:** Réserve d'une longueur d'onde avec la déflexion, sans conversion de longueur d'onde.

---

**Data:**  $Adj$ , un ensemble de noeuds OBS adjacents.

**Data:**  $Dest$ , la destination de la rafale.

**Data:**  $\lambda$ , la longueur d'onde utilisée.

**Data:**  $OT$ , le temps de décalage entre le BHP et la rafale.

```

1 prim ← ObtenirPrimaire(Dest), prim ∈ Adj
2 nh ← prim
3 if Contention(prim, λ, OT) then
4   | nh ← null
5   | foreach a ∈ Adj do
6   |   | if Route(a, Dest) ∧ ¬Contention(a, λ, OT) then
7   |   |   | nh ← a
8   |   |   | Fin boucle
9   |   | end
10  | end
11 end
12 if nh ≠ null then
13 | Réserver(nh, λ, OT)
14 end

```

---

- Lorsque la charge devient grande, la déflexion déstabilise le réseau et augmente la probabilité de blocage des rafales [WMA00b, LSKS05, CWXQ03, ZVR<sup>+</sup>07].
- Augmentation du délai de bout en bout. Une déflexion peut faire augmenter le nombre de sauts de la source à la destination, ce qui fait augmenter le délai.

### OT, FDL et déflexion

Comme spécifié dans la Section 2.3.3.1, le OT dépend du nombre de sauts et d'un certain nombre de constantes :

$$OT_{JET} = N * t_t + t_c \quad (2.5)$$

Or, avec la déflexion, si ce  $OT_{JET}$  est inchangé, alors un nombre limité de déflexions sera possible. Il y a donc deux choix possibles [HLH02a] : ne pas changer ce OT ou estimer le  $N$  à la source.

Si l'on estime le  $N$  à la source, qui est la méthode utilisée dans le cadre de ce travail, alors on doit changer la formule 2.5 :

$$OT_{JET} = (N + d) * t_t + t_c \quad (2.6)$$

où  $d$  est le nombre de sauts supplémentaire estimé nécessaire pour une potentielle déflexion.

Si le OT est inchangé, alors des FDLs doivent être utilisés lorsqu'il y a contention, comme illustré à la Fig. 2.18.

### Protection préemptive des rafales primaires

Zalesky et al. [ZVR<sup>+</sup>07] a introduit la notion de protection préemptive des chemins primaires. Le mécanisme consiste à donner une plus forte priorité aux rafales primaires, c'est-à-dire aux rafales qui n'ont pas effectué de déflexion. Ainsi, même lorsque la charge est grande,

les rafales entrants dans le réseau ne sont pas pénalisées par la résolution de la contention. La priorité est utilisée de la façon suivante. La rafale tente une réservation sur le chemin primaire. Si une contention se produit, alors si la rafale ayant causée la contention a utilisé une déflexion et que la rafale est primaire, il y a préemption de la rafale défléchie, c'est-à-dire que la rafale primaire prend la place de la rafale défléchie. Cette technique permet de diminuer le taux de perte lorsque la charge est élevée.

### Déflexion dynamique avec assignation virtuelle des longueurs d'onde

Lee et al. [LKS<sup>+</sup>05] ont étudié la relation entre la conversion des longueurs d'onde et la déflexion. Des priorités sur l'allocation des longueurs d'onde ont été établies. La déflexion s'effectue en tenant compte de plusieurs coûts : le coût de conversion une longueur d'onde  $\lambda_k$  vers une autre longueur d'onde  $\lambda_l$ , le coût de transmettre sur un lien et finalement en tenant compte du coût de la probabilité de contention. Ce travail a permis de démontrer que la déflexion dynamique permet d'obtenir moins de pertes comparativement au chemin le plus court en utilisant, au besoin, des chemins plus longs.

### Déflexion en considérant la probabilité de contention en aval

Ogino et al. [OT05] ont considéré que la déflexion doit prendre en compte le fait qu'une contention peut se produire en aval, c'est-à-dire que la résolution de la contention ne doit pas seulement tenir compte des liens adjacents. Des statistiques sont envoyées pour prendre en compte les probabilités de contention des noeuds entre les sources et les destinations.

Une approche similaire de notification rétrograde explicite de congestion a été proposée par Tanida et al. [TOC03].

### Impacts de la déflexion sur *Transmission Control Protocol* (TCP)

Quels sont les impacts de la déflexion sur TCP? Schlosser et al. [SG05] ont investigué cette problématique. Les performances de TCP dépendent du ratio de déflexion. Si trop de déflexions sont effectuées, cela a un effet négatif sur le séquençage des paquets TCP puisque cela est interprété comme s'ils étaient perdus.

Article	Année	$T_r$	Defl. à	Sélect.	Ctrl	OT	CLO	FDL	DP	Topo.	Éval.
[Gje07]	2007	F	T	SP	Aucun	F	Oui	Non	1.7-9	CO	Sim.
[ZVR+07]	2007	V	T	SP	N/D	F	Non	Non	N/D	Simp	Sim.
[HL05]	2005	V	T	SP	S	A	Oui	Oui	N/D	AN	Sim., Ana.
[LKS+05]	2005	V	T	D	Aucun	N/D	Oui	N/D	N/D	NSF	Sim.
[OT05]	2005	V	T	D	Aucun	N/D	N/D	N/D	N/D	NSF	Sim.
[CEJM05]	2005	V	T	D,SP	Aucun	F	Non	Non	N/D	PE	Sim.
[GZ05]	2005	N/D	T	D	TTL	F,FDL	Oui	Oui	1	NSF	Sim.
[CEJ05]	2005	V	T	D,SP	Aucun	D	Oui	Non	N/D	PE	Sim.
[CZZ05]	2005	N/D	T	R	N	F	Non	Non	N/D	NSF	Sim.
[ZVR+04]	2004	V	T	SP	A	F	Oui	N/D	N/D	MSN	Sim., Ana.
[ZVZ+04]	2004	N/D	N/D	SP	N/D	N/D	Oui	Non	N/D	NSF	Sim.,Ana.
[EJM04]	2004	V	TS	D,S	TTL	N/D	Non	Non	N/D	EO,Ita	Sim.,Ana.
[CWXQ03]	2003	V	T	R	P	N/D	O-N	N/D	N/D	N/D	Sim.
[TOC03]	2003	V	T	D	Aucun	N/D	N/D	N/D	N/D	MSN	Sim.
[CDL03]	2003	N/D	N/D	SP	N/D	F	Oui	Non	N/D	N/D	Sim.,Ana.
[KKK02]	2002	F	T	SP	S	F	N/D	N/D	N/D	MSN	Sim.
[HLH02b]	2002	V	T	SP	Aucun	FDL	Oui	Oui	N/D	AN	Sim.,Ana.
[HLH02a]	2002	V	T	SP	S	A	Oui	Oui	N/D	AN	Sim.,Ana.
[WMA00b]	2000	V	TS	R	TTL	N/D	Oui	N/D	< 2	GN	Sim.
[WMA00a]	2000	V	TS	R	TTL	N/D	Oui	N/D	< 2	GN	Sim.

TAB. 2.3 – Caractéristiques des travaux effectués sur la déflexion

### Caractéristiques des travaux portant sur la déflexion

Le tableau 2.3 comprend les principales caractéristiques répertoriées dans les travaux effectués sur la déflexion pour les réseaux OBS. Les caractéristiques utilisées sont les suivantes :

- *L'année* de la publication.
- *La taille des rafales* ( $T_r$ ) : Variable (V) ou fixe (F).

- *L'endroit où la déflexion peut s'effectuer* (Defl. à) : Tous les noeuds (T), tous les noeuds sauf la source (TS).
- *La sélection des chemins* (Sélect.) : Le chemin le plus court (*Shortest Path* (SP)), Dynamiquement (D), aléatoirement (R).
- *Le contrôle de la déflexion* (Ctrl.) : Le nombre de sauts (S), *Time-To-Live* (TTL), limité avec une probabilité (P), pré calcul d'alternatives (A), nombre de déflexions (N).
- *Adaptation du OT* : Fixe (F), adaptatif (A), avec des FDLs (FDL).
- *Conversion de longueurs d'onde permise* (CLO) : Oui ou non.
- *FDL disponible* (FDL) : Oui ou non.
- *Délai de propagation des liens en ms* (DP).
- *Topologie* (Topo.) : *Manhattan Street Network* (MSN), *Grid Network* (GN), *Arpanet* (AN), *NSF Network* (NSF), *Pan European* (PE), *EONET* (EO), *Italian Network* (Ita), *COST 239* (CO), *Simple* (Simp).
- *Évaluation des performances* (Eval.) : Simulation (Sim.), modèle analytique (Ana.).

La majorité des travaux utilise la configuration suivante :

- Une taille de rafale variable.
- Les déflexions sont permises à tous les noeuds.
- Déflexion avec routage SP.
- Aucun contrôle de la déflexion.
- OT fixe.
- La conversion de longueurs d'onde est permise.
- Les fibres optiques à retardement ne sont pas disponibles.
- Une topologie irrégulière.
- Une évaluation à l'aide de la simulation.

### 2.3.4.6 Retransmission

La retransmission OBS [ZVWJ05] consiste à stocker les rafales dans des files d'attente (en électronique) aux noeuds de bord et de retransmettre ces rafales lorsqu'une contention se produit. Si une rafale est retransmise  $\eta$  fois et qu'une retransmission est effectuée lors de la réception d'un paquet *Negative-Acknowledgement* (NACK), alors le délai de bout en bout est le suivant :

$$T_\eta = \sum_{i=1}^{\eta} (2 * (N_i) * t_t) + N_{\eta+1} * t_t + t_c \quad (2.7)$$

où  $\forall_{i=1}^N$ ,  $N_i$  correspond au nombre de sauts entre la source et le noeud où la contention s'est produite. Comme on considère qu'un NACK est envoyé du noeud en contention vers la source pour demander une retransmission, un aller-retour est nécessaire pour chaque retransmission.

#### Paquets de notification

Plusieurs scénarios sont envisageables au niveau des paquets de notification (*Positive-Acknowledgement* (ACK) et NACK) utilisés pour la retransmission :

- *Avec ACK et NACK* : Le NACK est envoyé en cas de contention et sert à retransmettre une rafale. Le paquet ACK est utile pour notifier qu'un paquet s'est rendu à la destination et qu'il peut donc être enlevé de la file d'attente.
- *Avec NACK* : La rafale dans la file d'attente à la source doit alors être supprimé après un certain temps *timeout* si aucun NACK n'est reçu.
- *Avec ACK* : Les rafales dans les files d'attente sont retransmises après un temps *timeout* si aucun ACK n'est reçu de la destination.

#### Retransmission partielle

La retransmission partielle [UCCR06] consiste en une combinaison de la segmentation et de la retransmission où une rafale est composée de plusieurs segments et les segments en

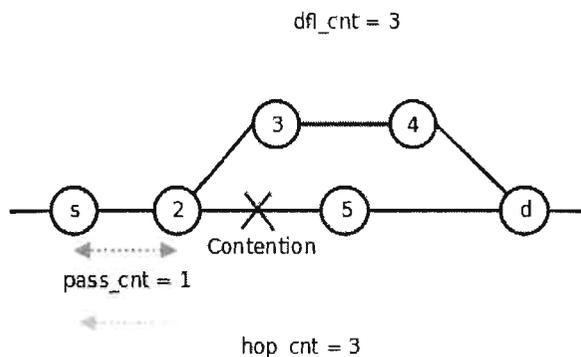


FIG. 2.19 – Exemple illustrant l’algorithme LHDR

contention sont retransmis.

### Retransmission contrôlée

Maach et al. [MBM04] ont proposé de ne retransmettre que lorsque la charge d’un réseau est en dessous d’un seuil. Lorsque la charge est en dessus de ce seuil, alors il faut contrôler le nombre de retransmission.

#### 2.3.4.7 Déflexion et retransmission

Ngo et al. [NJH06, NJHG06] ont proposé une combinaison statique de la déflexion et de la retransmission, nommée *Hybrid Deflection and Retransmission* (HDR). Les résultats de leurs simulations ont démontré que HDR ne donne pas de bonnes performances en termes de *Burst Loss Ratio* (BLR) puisque la déflexion est essayée systématiquement, même si la charge est élevée. Pour contrer cet effet négatif, les auteurs ont proposé un autre mécanisme, nommé *Limited Hybrid Deflection and Retransmission* (LHDR) dont le but est de limiter la déflexion.

### LHDR

---

**Algorithm 2:** Algorithme de résolution de contention LHDR.

---

```
if Contention then
  if  $pass\_cnt > 1 \wedge dfl\_cnt \leq hop\_cnt$  then
    | Effectuer une déflexion
  end
  else
    | Effectuer une retransmission
  end
end
```

---

L'algorithme LHDR est utilisé pour choisir entre effectuer une déflexion ou une retransmission lorsqu'il y a contention. Le choix est basé sur des critères statiques. LHDR introduit 3 variables fondamentales (Fig. 2.19) :

1. *pass\_cnt* : Le nombre de sauts de la source au noeud courant.
2. *dfl\_cnt* : Le nombre de sauts du noeud courant à la destination en passant par la route alternative.
3. *hop\_cnt* : Le nombre de sauts de la source à la destination en passant par le chemin primaire (le chemin le plus court).

L'algorithme LHDR est décrit dans l'algorithme 2. Une déflexion est donc effectuée que lorsque le noeud courant est plus loin que le premier noeud et lorsque la route défléchie est moins grande ou égale à la route primaire. Ainsi, le nombre de possibilités de déflexions est limité. L'exemple de la figure 2.19 illustre l'utilisation de l'algorithme lorsqu'un BHP arrive au noeud 2 et qu'une contention se produit. Dans cet exemple, puisque  $pass\_cnt = 1$ , la rafale est retransmise.

## Chapitre 3

# Combinaison adaptative hybride de la déflexion et de la retransmission pour les réseaux OBS

### 3.1 Introduction

Ce chapitre introduit un nouvel algorithme qui combine la déflexion et la retransmission, nommé *Adaptive Hybrid Deflection and Retransmission* (AHDR). Un seuil de décision adaptatif est utilisé pour décider dynamiquement entre effectuer une déflexion et retransmettre la rafale. Les décisions sont basées sur les connaissances locales à propos du réseau. Le OT est adapté en utilisant le seuil de décision. Pour rendre possible la connaissance locale du réseau, des paquets de notification sont utilisés. En premier lieu, l'algorithme est décrit en détails. Finalement, des résultats et une conclusion sont présentés.

### 3.2 Description détaillée de l'algorithme AHDR

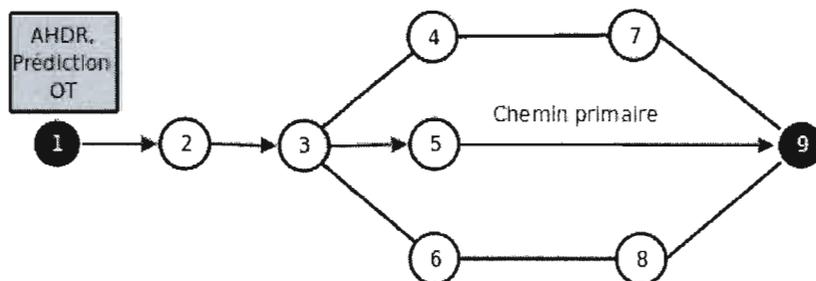


FIG. 3.1 – AHDR, scénario sans congestion ni contention

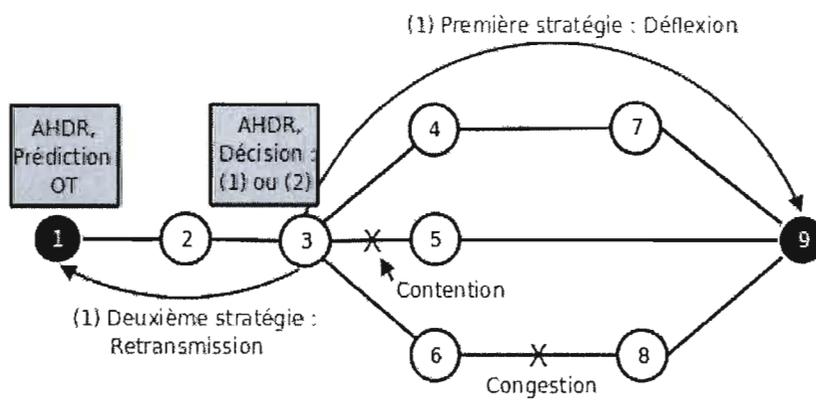


FIG. 3.2 – AHDR, scénario de congestion

L'algorithme AHDR permet d'optimiser la décision entre effectuer une déflexion ou une retransmission. Lorsque aucune contention se produit entre la source et la destination, le chemin primaire est utilisé (Fig. 3.1). Par contre, lorsqu'il y a contention, l'algorithme AHDR sélectionne la meilleure stratégie de congestion parmi la déflexion et la retransmission (Fig. 3.2). Dans l'exemple de la figure 3.2, le BHP tente de réserver entre le noeud 3 et 5, mais une contention se produit. L'algorithme décide donc de résoudre cette contention ou bien en effectuant une déflexion ou en retransmettant. Un temps est nécessaire entre le temps d'arrivé d'un paquet NACK et le temps où la rafale est retransmise. De plus, l'algorithme AHDR optimise la sélection de routes alternatives et le OT est déterminé à la source en prévoyant si une déflexion est requise pour se rendre à la destination. L'algorithme AHDR permet donc de combiner la déflexion et la retransmission de façon adaptative en considérant l'état du réseau, comparativement aux travaux de Ngo et al. (Voir section 2.3.4.7) qui font ce choix avec des critères statiques.

### 3.2.1 Transfert de statistiques à travers le réseau

Lorsqu'un BHP se rend jusqu'à destination, un paquet ACK est envoyé à la source. Lorsqu'une contention se produit, un paquet NACK est envoyé dans le but de notifier la source pour retransmettre la rafale. Dans les travaux de Ngo et al. [NJH06], des paquets de contrôle similaires ont été utilisés. Par contre, comparativement à leurs travaux, AHDR utilise ces paquets de contrôle non seulement pour notifier la source, mais aussi pour transférer des statistiques à propos de l'état du réseau (Fig. 3.3). La figure 3.3 illustre la signalisation utilisée par AHDR : le protocole de signalisation JET avec deux paquets de notification supplémentaires utilisés pour la retransmission.

Deux métriques sont utilisées pour effectuer les décisions : le BLR et l'utilisation des liens. Ces métriques sont intégrées dans les paquets de notification (ACK et NACK). Dans le cas d'un paquet NACK, les métriques intégrées sont celles entre le noeud courant de la contention et le noeud suivant. Dans le cas d'un paquet ACK, les métriques intégrées sont celles entre

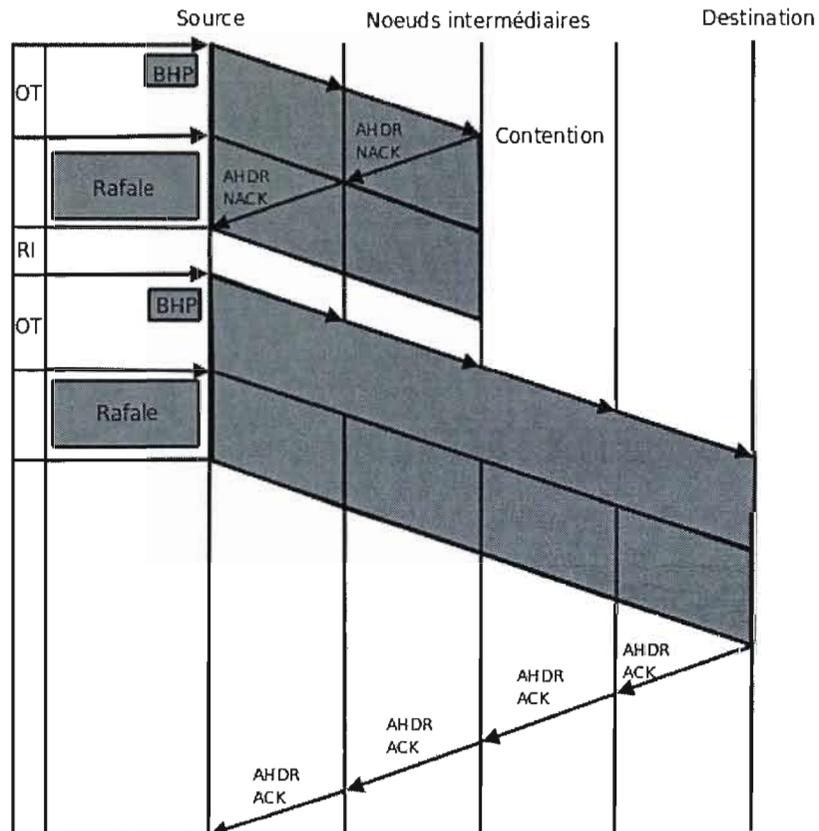


FIG. 3.3 – Signalisation AHDR, scénario avec retransmission

le noeud courant (la destination) et le noeud précédent.

Le tableau 3.1 décrit les informations contenues dans un paquet de contrôle AHDR. Cette description est en fait une augmentation du paquet de contrôle utilisé dans les travaux de Wang et al. [WMA00b].

<i>Information</i>	<i>Description</i>
<i>Type de paquet</i>	Le type du paquet de contrôle (SETUP, ACK, NACK).
<i>Id destination</i>	Identifiant du noeud de la destination de la rafale.
<i>Id source</i>	Identifiant du noeud de la source de la rafale.
<i>Id rafale</i>	Identifiant de la rafale (séquence).
<i>OT</i>	Le temps entre le BHP et la rafale.
<i>Temps départ</i>	Le temps de départ de la rafale à chaque noeud.
<i>Taille<sub>Rafale</sub></i>	La taille de la rafale.
<i>Id échec</i>	Identifiant d'un code d'erreur.
<i>Id<sub>1</sub> statistiques</i>	Noeud où la contention s'est produite, utile pour les statistiques.
<i>Id<sub>2</sub> statistiques</i>	Noeud adjacent au noeud de la contention : ACK : précédent, NACK : suivant.
<i>BLR<sub>Id<sub>1</sub>,Id<sub>2</sub></sub></i>	BLR entre le noeud <i>Id<sub>1</sub></i> et le noeud <i>Id<sub>2</sub></i> .
<i>U<sub>Id<sub>1</sub>,Id<sub>2</sub></sub></i>	Utilisation du lien entre le noeud <i>Id<sub>1</sub></i> et le noeud <i>Id<sub>2</sub></i> .

TAB. 3.1 – Informations comprises dans un paquet de contrôle AHDR

Ce paquet de contrôle diffère de celui de Wang et al. [WMA00b] sur quelques points :

- Il n'y a pas de drapeau pour différencier les rafales ayant effectuées une déflexion de celles qui n'en ont pas fait.
- Aucun champ TTL. Le champ TTL n'est pas nécessaire puisqu'il suffit de décrémenter successivement le OT à chaque noeud. Ainsi, lorsque  $OT = 0$ , alors on pose que  $TTL = 0$ .

- Les champs des métriques sont ajoutés. Ainsi, comme deux champs ont été enlevés, mais que quatre autres champs ont été ajoutés, la différence entre la taille du paquet de contrôle de Wang et al. [WMA00b] est de deux nombres en point flottant.

### 3.2.2 Limitation des routes alternatives

Pour limiter la longueur des routes alternatives possibles, un paramètre (noté  $\xi$ ) est introduit, ce qui permet d'exprimer un seuil de la taille d'une route alternative. Soit *Defl* une route possible de déflexion, *Primaire* la route primaire et  $|Route|$  le nombre de sauts de la route *Route*. L'algorithme 3 permet d'obtenir les déflexions possibles à un noeud donné.

---

**Algorithm 3:** Limitation des déflexions possibles

---

**Data:** *Primaire*, le chemin le plus court entre la source et la destination.

**Data:** *Courant*, le noeud courant.

**Data:** *D*, l'ensemble des routes possibles qui utilisent une déflexion, de la source jusqu'au noeud *Courant* en passant par le chemin primaire *Primaire*.

```

1 deflexions  $\leftarrow$  {}
2 foreach d  $\in$  D do
3   | if  $|d| \leq |Primary| * \xi$  then
4   |   | deflexions  $\leftarrow$   $\{d\} \cup deflexions$ 
5   | end
6 end

```

---

### 3.2.3 Seuil adaptatif de décision

Lorsqu'une contention se produit, le but primaire est de résoudre cette contention ou bien en effectuant une déflexion ou une retransmission. Lorsque la charge du réseau augmente, la déflexion déstabilise le réseau [WMA00b, LSKS05, CWXQ03, ZVR<sup>+</sup>07]. Par contre, la déflexion permet de maximiser l'utilisation de la bande passante tout en permettant de résoudre la contention. La problématique suivante existe donc : Quand faut-il effectuer une

déflexion sans pour autant déstabiliser le réseau? AHDR répond à cette question. L'idée générale de l'algorithme AHDR est d'effectuer des déflexions en comparant des métriques ( $M$ ) du réseau avec un certain seuil ( $TH$ ).

---

**Algorithm 4:** Fonctionnement général de AHDR.

---

```

if Contention then
  | if  $M \geq TH$  then
  | | Effectuer une déflexion
  | end
  | else
  | | Effectuer une retransmission
  | end
end

```

---

Dans ce travail, le  $M$  utilisé est en fait la probabilité de succès calculée avec le BLR et l'utilisation d'un lien. De plus, le  $TH$  doit s'adapter selon des métriques du réseau tel qu'il sera démontré ultérieurement.

### 3.2.3.1 Probabilité de succès d'une route

AHDR incorpore des poids associés au BLR et à l'utilisation d'un lien pour mesurer la probabilité d'échec ( $P_{Echec}$ ) :

$$P_{Echec}(n_1, n_2) = \alpha_{BLR} * BLR_{(n_1, n_2)} + \alpha_U * U_{(n_1, n_2)} \quad (3.1)$$

où  $P_{Echec}$  exprime la probabilité d'échec de réserver une longueur d'onde entre les noeuds  $n_1$  et  $n_2$ ,  $\alpha_{BLR}$  est un poids associé à la métrique BLR et  $\alpha_U$  est un autre poids associé à l'utilisation d'un lien. Il est à noter que  $\alpha_{BLR} + \alpha_U = 1$ ,  $\alpha_{BLR} \in [0..1]$  et  $\alpha_U \in [0..1]$ .

La probabilité de succès d'une route  $R$  ( $P_{Succes}(R)$ ) est définie comme suit :

$$P_{Succes}(R) = \prod_{i=1}^{|R|-1} (1 - P_{Echec}(n_i, n_{i+1})) \quad (3.2)$$

La probabilité de succès du lien entre les noeuds  $n_i$  et  $n_{i+1}$  est exprimée par  $1 - P_{Echec}(n_i, n_{i+1})$ . Les probabilités de succès des liens sont alors multipliées pour obtenir une probabilité de succès globale de la route  $R$ .

### 3.2.3.2 Corrélation linéaire entre le BLR et le seuil de décision

AHDR effectue une déflexion lorsqu'une métrique  $M$  est plus grande qu'un certain seuil  $TH$ .  $M$  correspond à la probabilité de succès d'une route ( $P_{Succes}(R)$ , formule 3.2). Il existe une corrélation linéaire entre le BLR et le seuil  $TH$  (Fig. 3.4). Pour obtenir ce graphique, des simulations pour plusieurs charges avec plusieurs seuils ont été effectuées. Par exemple, pour une charge de 0.74 (Fig. 3.5), on remarque que le meilleur seuil est près de 0.18.

<i>Charge</i>	<i>Seuil</i>	<i>BLR</i>	<i>Pondération</i>
0.1	0.1	0.01	3
0.1	0.2	0.015	2
0.1	0.3	0.02	1
0.8	0.1	0.20	1
0.8	0.2	0.15	2
0.8	0.3	0.10	3

TAB. 3.2 – Exemple de pondération pour calculer la corrélation linéaire

Un exemple est présenté à l'aide du tableau 3.2 pour deux charges et une variation de trois seuils. Une simulation a donc comme paramètre la charge et le seuil, ce qui fait en sorte que cet exemple comporte six simulations. Après la simulation, le plus petit BLR est associé à la plus grande pondération (3 dans l'exemple). Ainsi, un comparateur pondéré peut être calculé

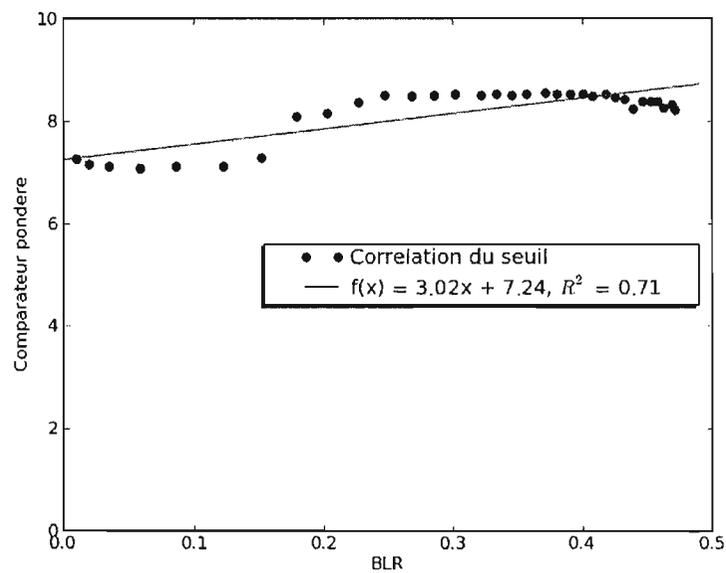


FIG. 3.4 – Corrélation linéaire entre le BLR et le seuil, topologie NSFNET

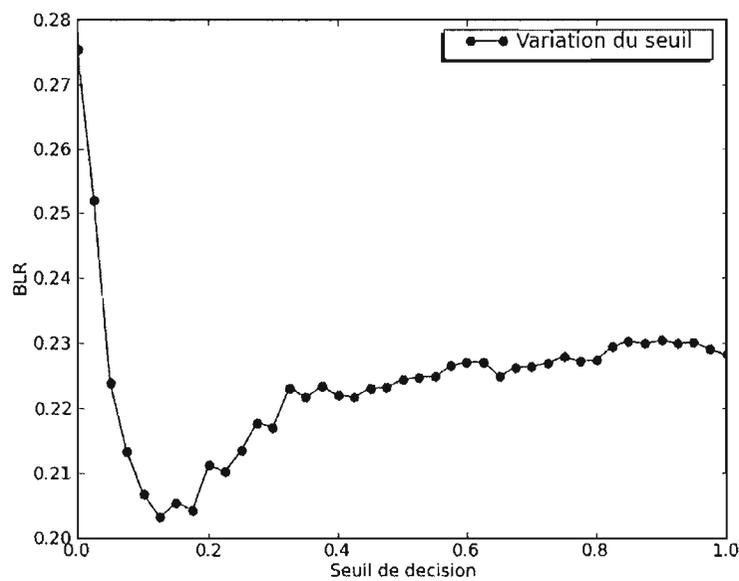


FIG. 3.5 – Variation du seuil de décision avec une charge de 0.74, topologie NSFNET

en additionnant les seuils multipliés aux pondérations. Soit les seuils  $S$ , les pondérations  $P$ , le *comparateur pondéré* ( $CP_{Charge}$ ) d'une charge donnée est exprimé par :

$$CP_{Charge} = \sum_{i=1}^{N_{Seuil}} (s_i * p_i), s_i \in S, p_i \in P \quad (3.3)$$

où  $N_{Seuil}$  correspond au nombre de seuils utilisés et

$$\forall_{i=1}^{N_{Seuil}} \neg \exists_{0 \leq j \leq N_{Seuil}, j \neq i} P_i \geq P_j \wedge BLR_i \geq BLR_j \quad (3.4)$$

Par exemple, pour  $Charge = 0.1$ , le comparateur pondéré est exprimé par :

$$CP_{0.1} = 0.1 * 3 + 0.2 * 2 + 0.3 * 1 = 1 \quad (3.5)$$

Avec NSFNET (Fig. 3.4), 33 charges avec 41 seuils ont été utilisés, totalisant  $33 * 41 = 1353$  simulations. La relation linéaire entre le BLR et le meilleur seuil de décision est forte puisque le coefficient de détermination est 71 % [MM05]. Les métriques telles que le BLR et l'utilisation d'un lien sont donc utilisées pour obtenir un bon seuil de décision.

### 3.2.3.3 Critère de sélection de résolution de contention

Pour décider entre effectuer une déflexion et retransmettre une rafale tout en considérant l'état du réseau, une fonction générant un seuil de décision est introduite :

$$Seuil_{P_{Succes}}(BLR_{Reseau}, U_{Reseau}) = \beta_{BLR} * BLR_{Reseau} + \beta_U * U_{Reseau} \quad (3.6)$$

où  $\beta_{BLR}$  est un poids appliqué au BLR du réseau (connu localement) et  $\beta_U$  est un poids appliqué à l'utilisation du réseau. Il est à noter que  $\beta_{BLR} + \beta_U \leq 1$ . Cette fonction est linéaire et une telle fonction a été conçue puisqu'il existe une relation linéaire entre le BLR et le seuil (Voir section 3.2.3.2).

Or, pour que le seuil ne change pas trop radicalement à cause de petites fluctuations de trafic, ne changeant pas le comportement de ce dernier, la mise à jour du seuil est effectuée en utilisant un poids :

$$Seuil'_{MAJ} = \phi * Seuil_{MAJ} + (1 - \phi) * Seuil_{P_{Succes}}(BLR_{Reseau}, U_{Reseau}) \quad (3.7)$$

où  $\phi \in [0..1]$  correspond à un poids sur l'ancienne valeur du seuil. Ainsi, plus  $\phi$  tend vers 0 et plus le seuil changera rapidement aux variations du trafic.

Si la probabilité de succès d'une route alternative (formule 3.2) est plus grande ou égale que le seuil (formule 3.6), alors cette déflexion est essayée. Soit *Déflexion Allouée* (*DA*), une procédure permettant de savoir si une route de déflexion *defl* doit être essayée :

(3.8)

$$DA(defl) = \begin{cases} Vrai & \text{si } P_{Succes}(defl) \geq Seuil_{P_{Succes}}(BLR_{Reseau}, U_{Reseau}) \\ Faux & \text{sinon} \end{cases}$$

Évidemment, ces formules sont pré-calculées périodiquement et une table de routage typique est périodiquement mise à jour pour ne pas pénaliser le processus d'acheminement de rafales.

### 3.2.4 Processus d'acheminement de rafales

Lorsqu'un BHP arrive à un noeud principal, le noeud suivant est sélectionné à l'aide d'une table de routage dans le domaine électronique dans le but de réserver la bande passante pour

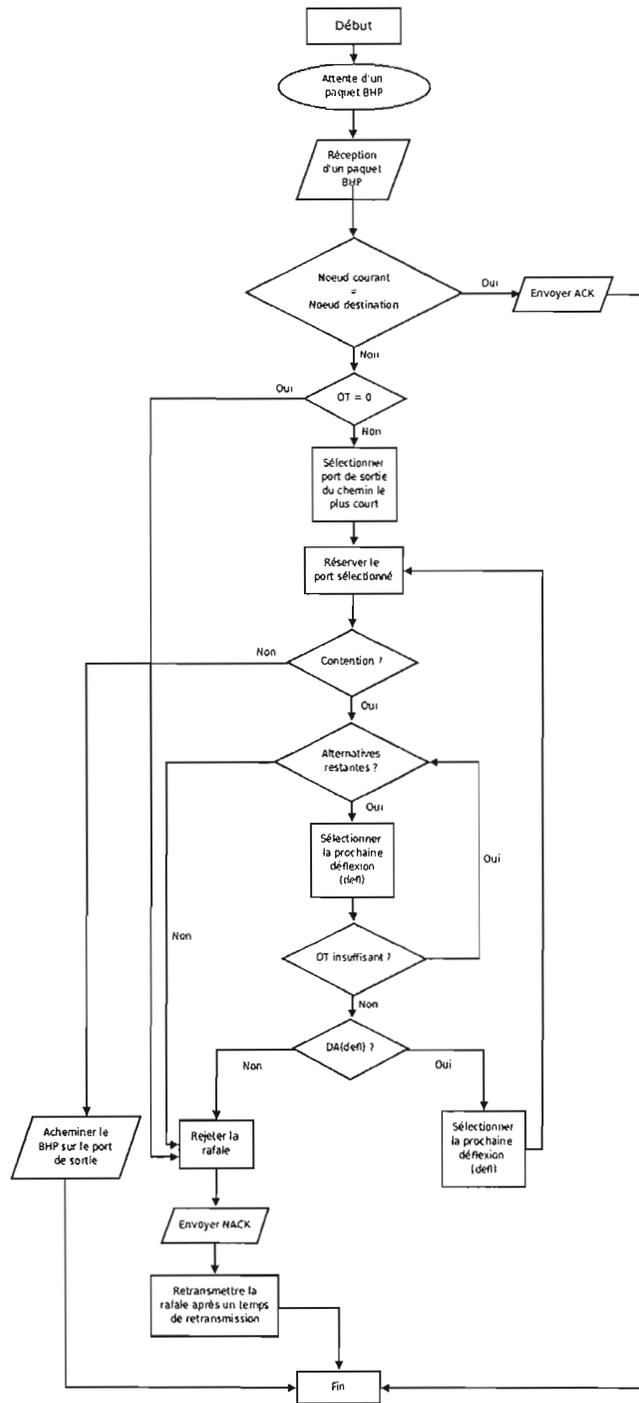


FIG. 3.6 – Processus d'acheminement de paquets/rafales, AHDR

la rafale de données. Le processus d'acheminement (Fig. 3.6) de paquets/rafales adopté par AHDR est le suivant :

- Lorsqu'un paquet de contrôle est reçu, l'identifiant du noeud courant est comparé à l'identifiant du noeud de la destination. Si le BHP est arrivé à la destination, un paquet de notification ACK est envoyé à la source.
- Ensuite, le OT est extrait dans le but de vérifier s'il est encore suffisant. Si le OT est insuffisant, un paquet de notification NACK est envoyé à la source et la rafale est retransmise après un temps de retransmission.
- Le port de sortie du chemin le plus court est sélectionné. S'il y a contention, ceci est résolu ou bien en effectuant une déflexion ou alors en retransmettant la rafale. La rafale n'est pas retransmise et rejetée après un certain nombre de retransmissions ( $N_{ret}$ ).
- En cas de contention, l'algorithme extrait successivement les routes de déflexion de façon décroissante en comparant avec la probabilité de succès d'une route (formule 3.2).
- La probabilité de succès d'une route de déflexion *defl* est comparée au seuil adaptatif de décision (formule 3.6).
- Si la probabilité de succès de *defl* est plus petite que le seuil, alors un paquet NACK est envoyé à la source et la rafale est retransmise après un temps de retransmission. Sinon, le port de sortie sélectionné est réservé.

### 3.2.5 Table de routage AHDR

Pour ne pas pénaliser le processus d'acheminement de paquets/rafales physiquement, une table de routage (tableau 3.3) est mise à jour périodiquement en utilisant la formule 3.2.  $N$  représente le nombre de destinations possibles et  $M$  représente le nombre de prochains sauts possibles d'une certaine destination  $Dest_i$ . Le coût est exprimé par :

$$Cout(ps, dest) = 1 - P_{Succes}(R(ps, dest)) \quad (3.9)$$

TAB. 3.3 – Table de routage AHDR

Destination	Prochain saut	Coût
$Dest_1$	$Saut_{1, Dest_1}$	$1 - P_{Succes}(R(Saut_{1, Dest_1}, Dest_1))$
	$\vdots$	$\vdots$
	$Saut_{M, Dest_1}$	$1 - P_{Succes}(R(Saut_{M, Dest_1}, Dest_1))$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$Dest_N$	$Saut_{1, Dest_N}$	$1 - P_{Succes}(R(Saut_{1, Dest_N}, Dest_N))$
	$\vdots$	$\vdots$
	$Saut_{M, Dest_N}$	$1 - P_{Succes}(R(Saut_{M, Dest_N}, Dest_N))$

où  $R(ps, Dest)$  est la route en passant par le prochain saut  $ps$  vers la destination de façon à ce que les prochains sauts ayant la plus grande probabilité de succès ont un faible coût. Les prochains sauts sont triés de la façon suivante :

$$\forall_{i=1}^N \forall_{j=1}^{M-1} Cout(Saut_j, Dest_i) \leq Cout(Saut_{j+1}, Dest_i) \quad (3.10)$$

### 3.2.6 OT adaptatif

Dans les réseaux OBS, la rafale de données suit un paquet de contrôle après un OT calculé initialement au noeud source de bord. Le OT doit être assez grand de façon à ce que la rafale arrive à chaque commutateur après le BHP. Avec le protocole de signalisation JET (Voir Section 2.3.3.1), le temps minimum de OT doit considérer le temps de traitement ( $t_t$ ) du paquet BHP à chaque saut, le temps de commutation et le temps de configuration ( $t_c$ ). Donc, le OT minimum est exprimé par :

$$OT = N_{Sauts} * t_t + t_c \quad (3.11)$$

où  $N_{Sauts}$  est le nombre de sauts de la source à la destination. La formule 3.11 illustre le fait que l'élément principal pour prédire le meilleur OT est le nombre de sauts puisque  $t_t$  et  $t_c$  sont des valeurs connues. Par contre, si une déflexion est nécessaire, une plus longue route peut être utilisée, ce qui peut nécessiter d'augmenter  $N_{Sauts}$ .

Le nombre de sauts ( $N_{Sauts}$ ) utilisé dans la formule 3.11 est le suivant :

$$N_{Sauts} = \begin{cases} |mdefl| & \text{Si } DA(mdefl) \\ |Chemin plus court| & \text{Sinon} \end{cases} \quad (3.12)$$

où  $|mdefl|$  correspond au nombre de sauts de la meilleure déflexion qui maximise la formule 3.2. Si  $DA(mdefl)$  est vrai (formule 3.8), alors  $|mdefl|$  est utilisé pour le nombre de sauts pour calculer le OT. Le OT est donc adapté : le OT est augmenté lorsque les conditions du réseau sont bonnes en termes de BLR et d'utilisation des liens et le OT est diminué lorsque les conditions du réseau deviennent moins bonnes pour ainsi diminuer le nombre de déflexions d'une façon adaptative.

### 3.3 Résultats expérimentaux et discussion

#### 3.3.1 Configurations expérimentales

Dans le but d'évaluer les performances de l'algorithme AHDR, des simulations sont effectuées avec les topologies NSFNET (Fig. 3.7) et COST239 (Fig. 3.8), deux topologies

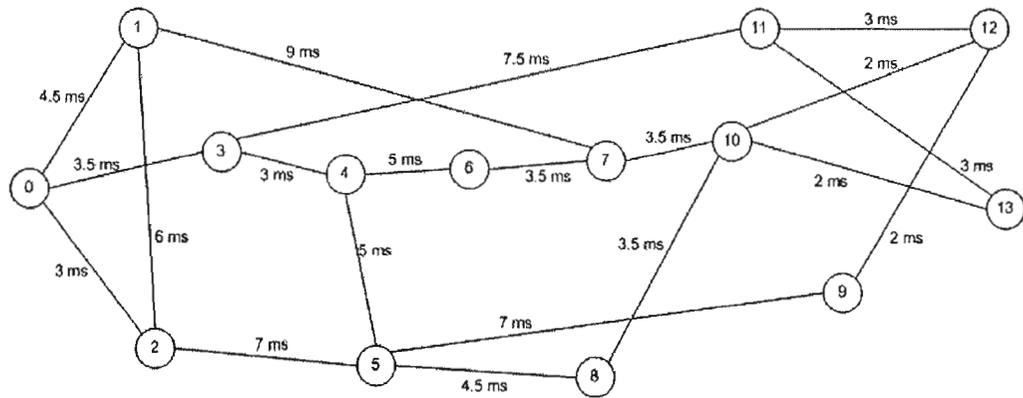


FIG. 3.7 - Topologie NSFNET

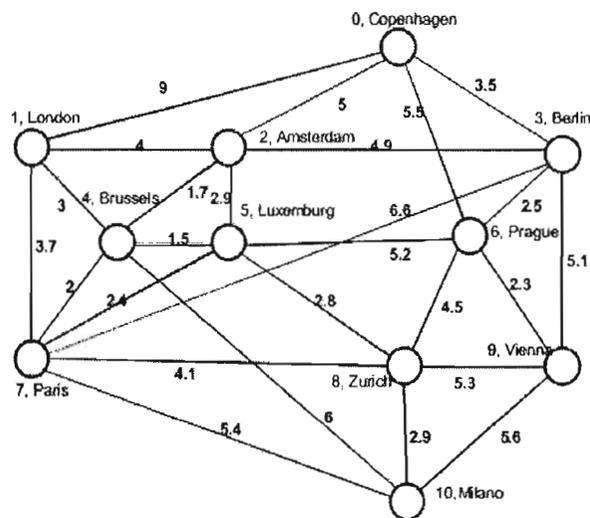


FIG. 3.8 - Topologie COST239

irrégulières. Le simulateur utilisé est le simulateur *Network Simulator 2* (ns-2) [NS] avec un module supplémentaire pour OBS.

Les configurations utilisées sont les suivantes :

- Chaque longueur d'onde a 1 Gbit/s de capacité.
- Chaque lien contient 2 canaux de contrôle et 4 canaux de données.
- La taille moyenne d'une rafale (noté  $R$ ) est de 400 KB.
- La création de rafales suit une distribution de Poisson pour le taux d'arrivée des rafales et pour la taille des rafales.
- Les connexions sont distribuées aléatoirement à travers le réseau. Chaque générateur de trafic envoie des rafales à tous les noeuds sauf à lui-même.
- Les rafales sont perdues indéfiniment après un certain nombre de retransmissions  $N_{ret}$  (retransmissions tronquées). Dans le cadre de ce travail,  $N_{ret}$  a été fixé à 1 pour ne pas augmenter significativement le temps de bout en bout.
- Les rafales à retransmettre sont retransmises après  $Rand(0, 0.05)$  secondes. Les rafales retransmises sont donc pénalisées en termes de délai de bout en bout. La fonction  $Rand$  génère un nombre aléatoire entre un nombre minimum et un nombre maximum.
- Soit  $N$  le nombre de noeuds dans la topologie,  $\zeta_{i,j}$  le nombre de connexions OBS entre les noeuds  $i$  et  $j$ ,  $\lambda_{i,j,k}$  le nombre de rafales envoyées par seconde sur la connexion  $k$  entre les noeuds  $i$  et  $j$ ,  $\mu_i$  la capacité disponible à  $i$ , la charge réseau est définie par :

$$Charge = \sum_{i=1}^N \sum_{j=1}^N \sum_k^{\zeta_{i,j}} \frac{\lambda_{i,j,k} * R}{\mu_i} \quad (3.13)$$

- Le facteur de connectivité d'une topologie est exprimé par :

$$C = \frac{L}{N * (N - 1)/2} \quad (3.14)$$

où  $L$  correspond au nombre de liens dans le réseau. Par exemple la topologie NSFNET est faiblement connectée ( $C = 0.23$ ) et la topologie COST239 a un grand facteur de connectivité

( $C = 0.47$ ).

### 3.3.2 Résultats et analyse

#### 3.3.2.1 Poids du seuil de décision ( $\beta_{BLR}$ , $\beta_U$ )

Le seuil de décision (formule 3.6) comporte deux poids :  $\beta_{BLR}$  associé au BLR et  $\beta_U$  associé à l'utilisation d'un lien. Quelles sont les meilleures valeurs de ces poids, permettant de diminuer au maximum le BLR? Ces poids ont été variés avec 11 valeurs (0, 0.1, 0.2, ..., 1.0), ce qui donne  $11 * 11 = 121$  simulations pour une charge donnée. 3 charges ont été utilisées (Fig. 3.9, 3.10 et 3.12). Lorsque  $\beta_{BLR} = 0$  et  $\beta_U = 0$  (Fig. 3.11), ceci donne un BLR hautement supérieur puisque dans ce cas la déflexion est toujours essayée en premier, ce qui déstabilise le réseau lorsque la charge augmente. Ce cas a donc été enlevé pour éliminer le bruit (Fig. 3.9, 3.10 et 3.12). On remarque que les meilleurs résultats pour le poids  $\beta_{BLR}$  sont obtenus en utilisant des valeurs entre 0.2 et 0.8 et entre 0.2 et 0.5 pour le poids  $\beta_U$ . Les meilleurs résultats sont obtenus en utilisant des valeurs de poids près de  $\beta_{BLR} = 0.2$  et  $\beta_U = 0.2$ .

#### 3.3.2.2 Poids de mise à jour du seuil ( $\phi$ )

La valeur du seuil de décision ne change pas brusquement pour éviter de changer de seuil à cause d'une courte fluctuation du trafic. Le seuil est mis à jour en accordant un poids ( $\phi$ ) sur l'ancienne valeur du seuil (Voir formule 3.7). Un scénario spécifique est donc utilisé (Fig. 3.13) : à chaque seconde la charge du réseau change de 0.2 à 2.0, de 2.0 à 0.2 et ainsi de suite, pendant 120 secondes.

Le poids  $\phi$  a été varié (Fig. 3.14) pour vérifier l'impact de ce poids sur le BLR en utilisant le scénario spécifique de la figure 3.13. Les meilleures performances en termes de BLR ont été obtenues avec  $0.2 \leq \phi \leq 0.4$  sur la topologie NSFNET.

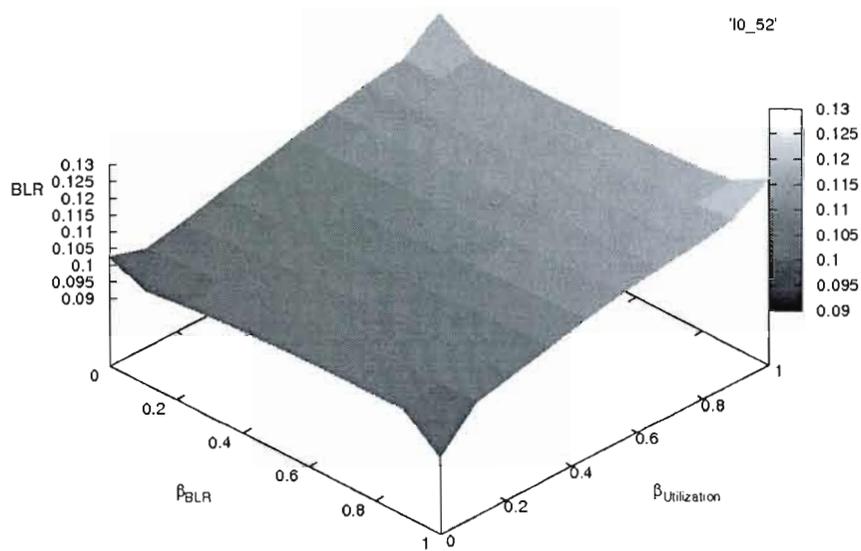


FIG. 3.9 – Variation du BLR en fonction de  $\beta_{BLR}$  et  $\beta_U$ , topologie NSFNET, Charge = 0.52

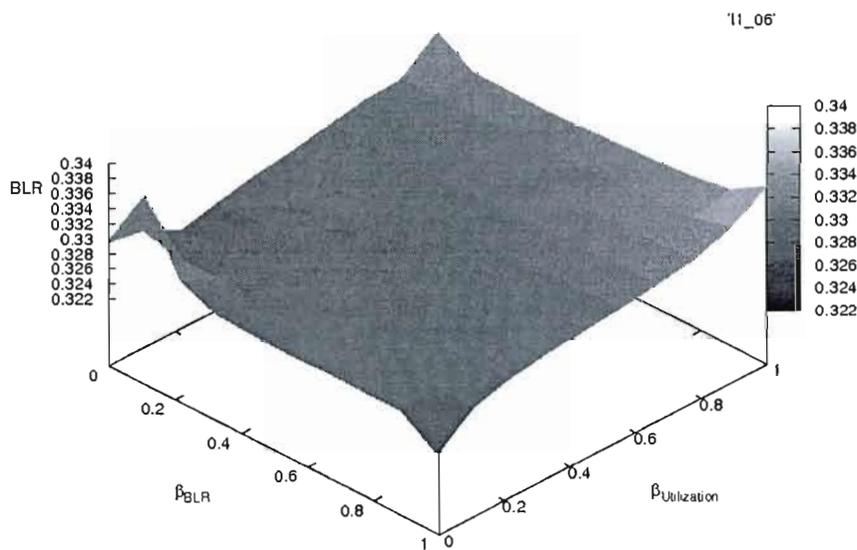


FIG. 3.10 – Variation du BLR en fonction de  $\beta_{BLR}$  et  $\beta_U$ , topologie NSFNET, Charge = 1.06

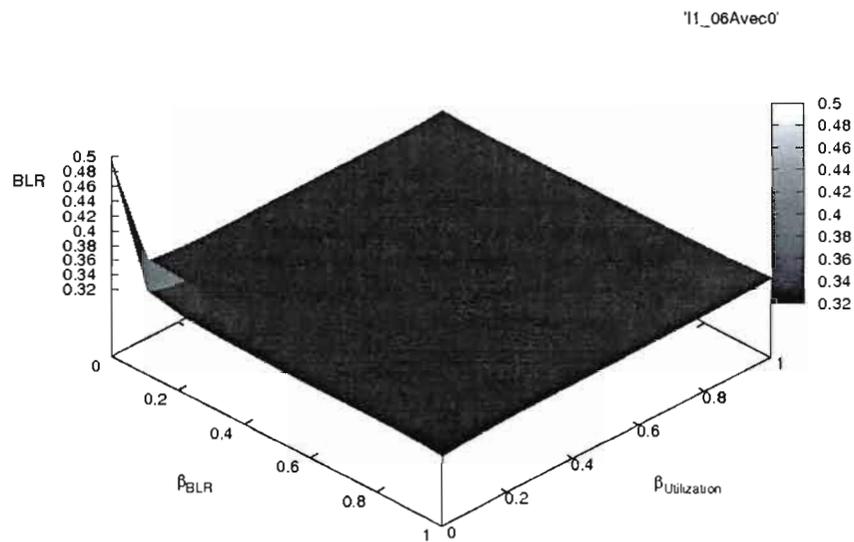


FIG. 3.11 – Variation du BLR en fonction de  $\beta_{BLR}$  et  $\beta_U$ , topologie NSFNET, Charge = 1.06, sans suppression du bruit

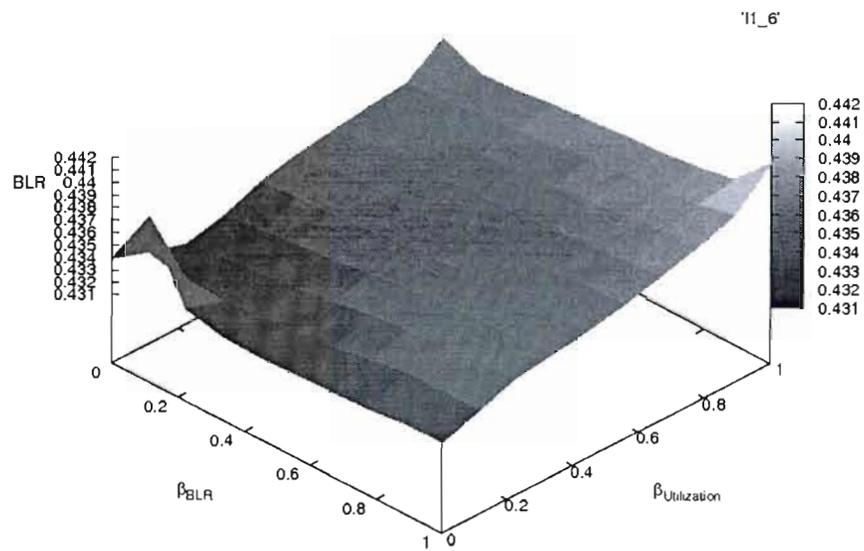


FIG. 3.12 – Variation du BLR en fonction de  $\beta_{BLR}$  et  $\beta_U$ , topologie NSFNET, Charge = 1.6

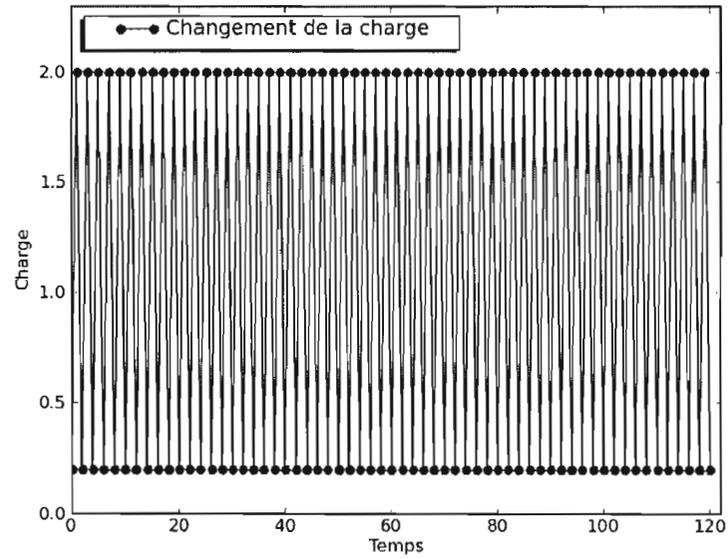


FIG. 3.13 – Variation de la charge, scénario pour tester  $\phi$

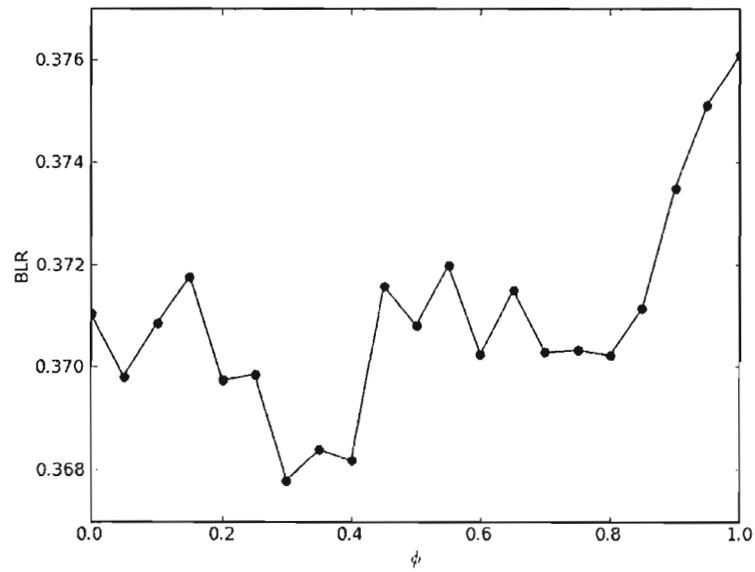


FIG. 3.14 – Variation du BLR en fonction de  $\phi$ , topologie NSFNET

### 3.3.2.3 Comparaison des performances entre AHDR et MLHDR

Cette section comporte une comparaison entre l'algorithme AHDR et l'algorithme *LHDR modifié* (MLHDR, voir la section 2.3.4.7 pour une description de LHDR). LHDR permet d'avoir seulement une alternative de déflexion tandis que AHDR permet d'en avoir plusieurs. LHDR est donc modifié dans le but d'effectuer une comparaison correctement. Lorsqu'il y a contention, MLHDR consiste à appliquer LHDR pour le deuxième chemin le plus court, ensuite pour le second chemin le plus court et ainsi de suite.

#### Topologie NSFNET

L'algorithme AHDR donne des améliorations considérables en termes de BLR même lorsque la charge est grande (Fig. 3.15). La topologie NSFNET a une faible connectivité, ce qui est un mauvais cas pour AHDR, algorithme qui sélectionne les liens les moins chargés. Lorsque peu d'alternatives sont disponibles, cela limite le gain possible dans le choix des routes. Dans ce scénario, le gain provient de deux sources :

- Un OT adaptatif est utilisé pour prendre en compte les états du réseau (Fig. 3.18).
- Le processus d'acheminement des rafales (Section 3.2.4).

On pourrait être tenté de prévoir intuitivement que le délai de bout en bout est fortement augmenté puisque AHDR peut sélectionner des routes plus longues que le chemin le plus court. Cependant, le délai de bout en bout est similaire à celui de MLHDR (Fig. 3.16). La différence maximale de AHDR du délai de bout en bout est de 2 ms, ce qui est négligeable du point de vue d'un client dans un scénario Internet où le paradigme *Best Effort* est utilisé.

Le OT adaptatif influence grandement le nombre de déflexions (*# deflexions*) par rapport au nombre de retransmissions (*# retransmissions*). Le ratio de déflexion est défini comme suit :

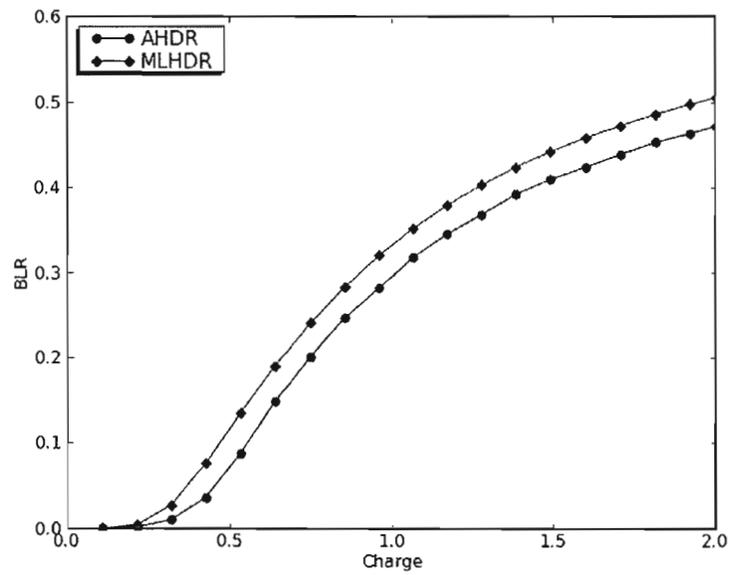


FIG. 3.15 – Variation du BLR en fonction de la charge, topologie NSFNET

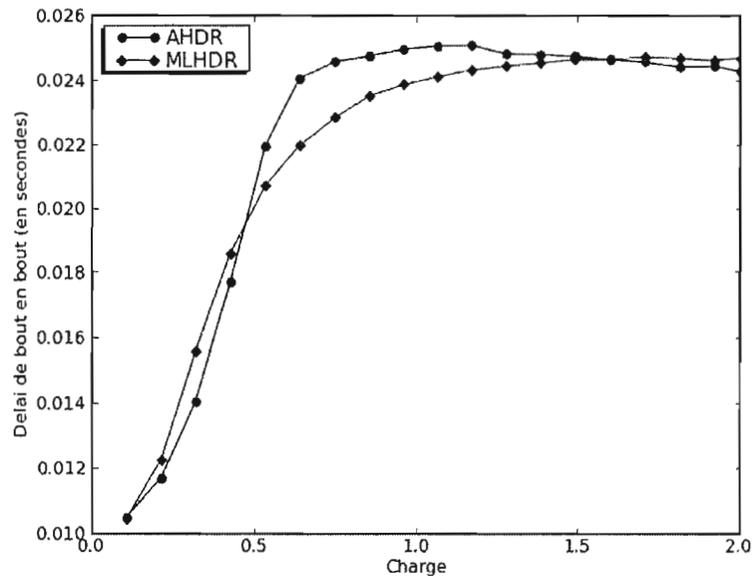


FIG. 3.16 – Variation du délai de bout en bout en fonction de la charge, topologie NSFNET

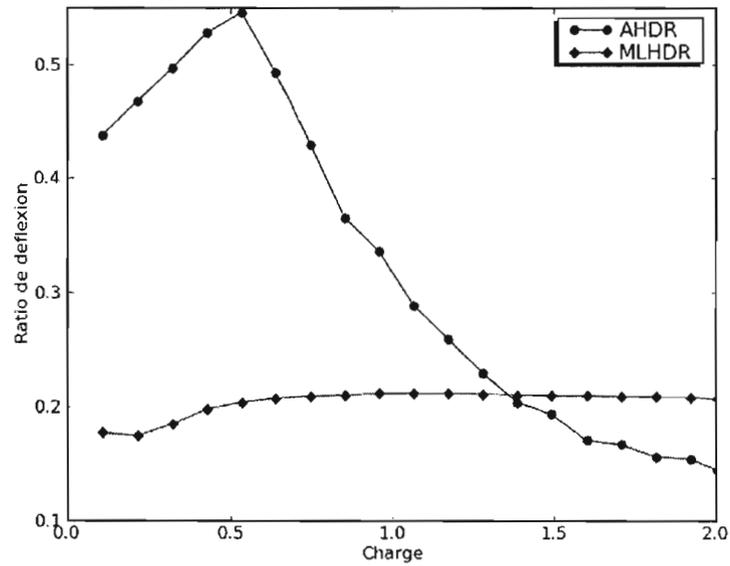


FIG. 3.17 – Variation du ratio de déflexion en fonction de la charge, topologie NSFNET

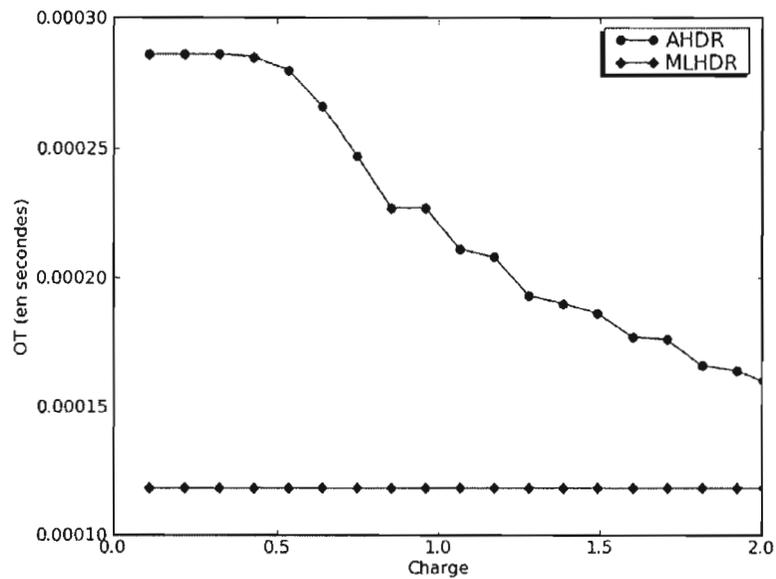


FIG. 3.18 – Variation du OT en fonction de la charge, topologie NSFNET

$$Ratio\ de\ flexion = \frac{\#\ de\ flexions}{\#\ de\ flexions + \#\ retransmissions} \quad (3.15)$$

Avec AHDR avec des charges où  $0 \leq Charge \leq 1.4$ , on peut clairement observer que les déflexions sont faites au maximum, de façon à ne pas nuire au BLR. On peut aussi clairement remarquer que pour des charges où  $0 \leq Charge \leq 1.4$ , l'algorithme MLHDR ne fait pas assez de déflexions et que pour des charges où  $1.4 < Charge \leq 2$ , trop de déflexions sont effectuées.

### Topologie COST239

Avec la topologie COST239, AHDR obtient des résultats considérablement meilleurs en termes de BLR comparativement aux résultats de MLHDR (Fig. 3.19). Notamment, lorsque la charge est inférieure à 0.4, AHDR n'obtient aucune perte. De plus, lorsque la charge est inférieure à 1.5, AHDR obtient un plus faible délai de bout en bout (Fig. 3.20). Ceci s'explique par le fait que le ratio de déflexion est très élevé avec AHDR avec la topologie COST239 (Fig. 3.21). Lorsque le ratio de déflexion est élevé, moins de retransmissions sont effectuées, ce qui fait décroître le délai de bout en bout. Lorsque la charge augmente, le OT diminue légèrement avec AHDR, tout en restant plus grand qu'avec l'algorithme MLHDR (Fig. 3.22).

## 3.4 Conclusion

L'algorithme proposé (AHDR) permet de combiner dynamiquement la déflexion et la retransmission pour diminuer les pertes. Pour rendre la décision de choisir entre effectuer une déflexion ou une retransmission plus effective, AHDR utilise un seuil de décision adaptatif. Ce seuil est adapté en utilisant le BLR et l'utilisation des liens puisqu'il existe une corrélation linéaire entre le BLR et le meilleur seuil.

Les topologies faiblement connectées comme NSFNET nécessitent d'adapter les déflexions et les retransmissions pour réduire les pertes. De plus, les approches statiques telles que

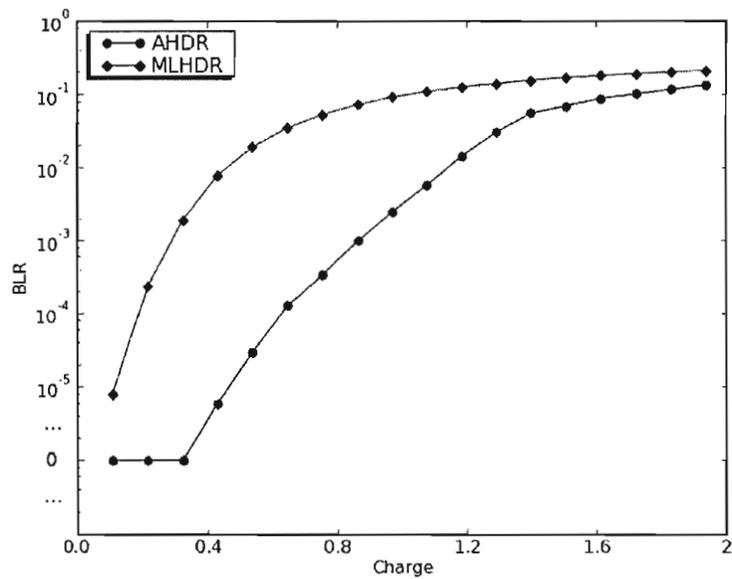


FIG. 3.19 – Variation du BLR en fonction de la charge, topologie COST239

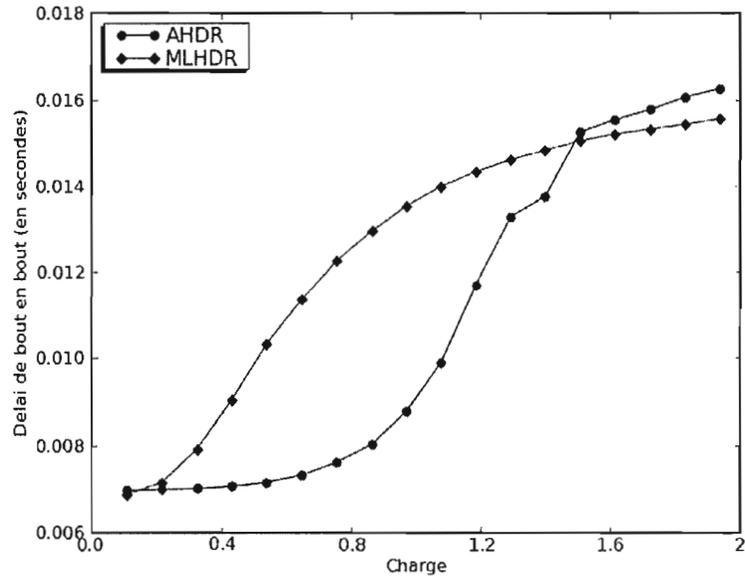


FIG. 3.20 – Variation du délai de bout en bout en fonction de la charge, topologie COST239

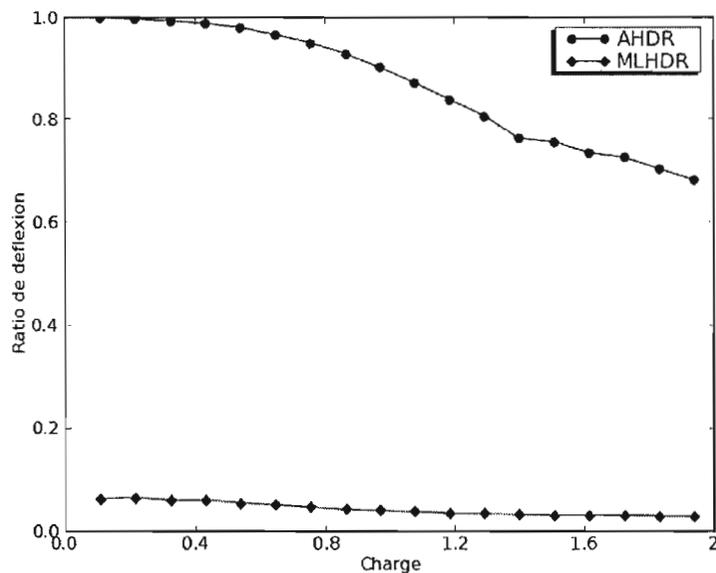


FIG. 3.21 – Variation du ratio de déflexion en fonction de la charge, topologie COST239

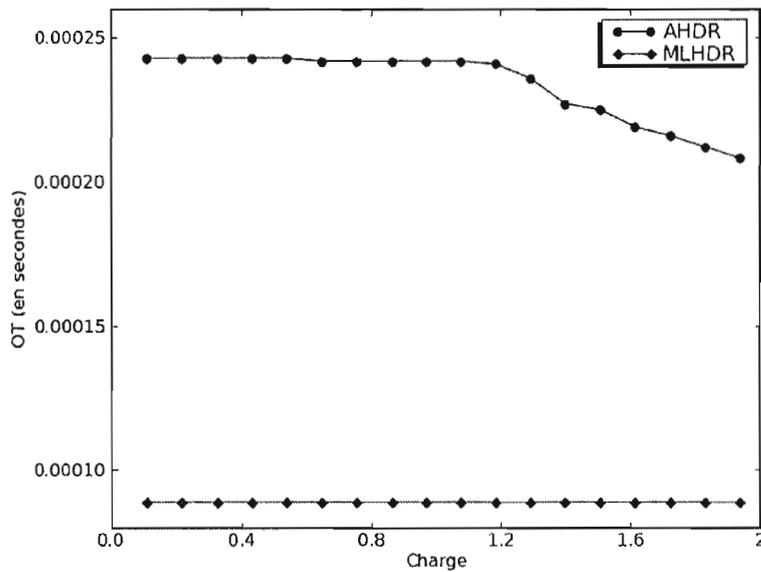


FIG. 3.22 – Variation du OT en fonction de la charge, topologie COST239

MLHDR effectuent pas assez de déflexions lorsque la charge est faible et trop de déflexions lorsque la charge augmente.

Les topologies fortement connectées telles que COST239 offrent la possibilité d'avoir un grand ratio de déflexion pour réduire grandement le BLR et le délai de bout en bout.

## Chapitre 4

# Modèle graphique probabiliste de prévention de la contention pour les réseaux OBS

### 4.1 Introduction

Ce chapitre introduit un nouveau *Modèle Graphique Probabiliste de Prévention de la Contention* (MGPPC) pour permettre d'effectuer un routage plus efficace pour prévenir la contention. En premier lieu, les réseaux bayésiens sont introduits pour démontrer l'utilité de ce modèle pour les réseaux OBS. Ensuite, le modèle MGPPC et des algorithmes associés sont présentés. Finalement, les résultats des topologies NSFNET et Pan-European avec un trafic réaliste sont présentés et analysés.

### 4.2 Motivation

Le trafic entre deux villes dépend fortement de la population, du nombre d'employés et le nombre d'hôtes [HBB<sup>+</sup>03]. Or, dans la réalité, le trafic n'est pas distribué uniformément

dans le temps. Un réseau doit être régulé en tenant compte de deux catégories de stratégie de régulation de trafic :

- Stratégies *réactives*, utiles pour résoudre la contention :
  - Déflexion.
  - Conversion de longueurs d'onde.
  - Retransmission.
  - Etc.
- Stratégies *proactives*, utiles pour réguler le trafic de façon préventive :
  - Routage.
  - Estimation du temps de OT.
  - Estimation du nombre de retransmissions et du temps de retransmission.
  - Etc.

L'idée principale est d'effectuer ces stratégies de façon intelligente pour s'adapter à l'environnement du réseau. Un modèle graphique probabiliste permet de spécifier un ensemble de distributions de probabilités avec la représentation d'un graphe [Nea03]. Un réseau bayésien est un modèle graphique probabiliste et un graphe acyclique orienté. Les noeuds représentent des variables et les liens représentent les dépendances.

La figure 4.1 présente un réseau bayésien simple composé de 3 noeuds (les variables) et 2 liens orientés (les dépendances). Les composantes d'une variable sont ses états et sa distribution conditionnelle (qu'on nomme table des probabilités conditionnelles). Dans l'exemple, la variable  $V$  comporte 2 états possibles ( $v_1$  et  $v_2$ ) et une table des probabilités conditionnelles (composée de 4 lignes). Un état est une valeur possible que peut prendre une variable. La variable  $C$  est une évidence puisque c'est une information connue. Les variables  $V$  et  $S$  dépendent de la variable  $C$ .  $C$  influence  $V$  et  $S$ , ce qui explique le sens des liens. On dit que la variable  $C$  est un noeud parent. Plus généralement, une variable dépend de ses variables parentes connectées. La table des probabilités conditionnelles d'une variable  $VA$  donne les

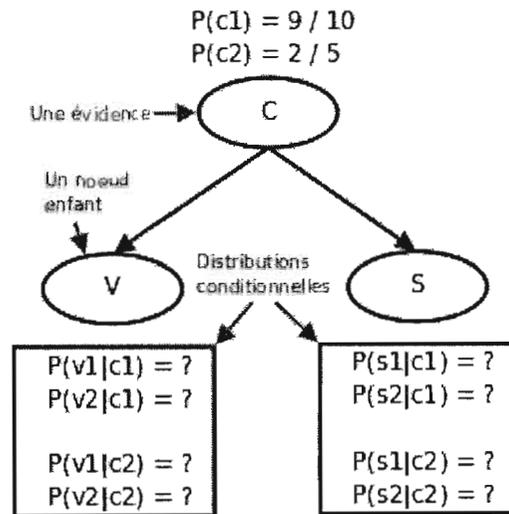


FIG. 4.1 – Un exemple simple de réseau bayésien

probabilités associées à l'ensemble des combinaisons des états de  $V A$  et des états des variables parentes de  $V A$ . Plus formellement, la distribution jointe d'un réseau bayésien contenant  $N$  variables  $X_1, X_2, \dots, X_N$  correspond au produit de la distribution d'une variable et de ses parents :

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^N P(X_i | \text{parents}(X_i)) \quad (4.1)$$

Si une variable n'a aucun parent, elle est inconditionnelle. Un autre concept important est la notion d'inférence. Obtenir le résultat d'une inférence est effectué en appliquant le théorème de Bayes [Nea03]. Si une certaine variable  $X_i$  a  $M$  évidences, la probabilité que la variable  $X_i$  soit dans l'état  $x_i$  est obtenu en effectuant l'inférence :

$$P(X_i = x_i | E_1 = e_1, E_2 = e_2, \dots, E_M = e_M) \quad (4.2)$$

Une inférence consiste à déduire des caractéristiques inconnues à partir d'informations connues. Cette application d'inférence est simple. Plusieurs autres applications avancées peuvent

être faites [Nea03].

Un réseau bayésien est un modèle qui permet de représenter des connaissances ainsi qu'un calculateur de probabilités conditionnelles. C'est un modèle qui permet l'apprentissage, par exemple l'apprentissage des paramètres (des tables de probabilités conditionnelles). L'utilisation d'un réseau bayésien est motivée par plusieurs critères :

- La formulation des inférences bayésiennes est naturelle : on obtient la probabilité qu'une variable soit dans un certain état en considérant ses variables parentes.
- Une probabilité est calculée selon plusieurs évidences parentes ce qui donne une certaine forme d'intelligence : une décision est prise en tenant compte des états de plusieurs variables selon l'historique. Dans le cadre des réseaux OBS, des exemples d'évidences sont : le BLR, le délai de bout en bout, la charge, le OT, etc.
- Le concept d'interconnecter des informations à l'aide de flèches est naturel pour l'être humain. Ainsi, un système peut être modélisé et modifié graphiquement à l'aide des experts du domaine.
- La capacité d'apprentissage des réseaux bayésiens permet au réseau de s'adapter automatiquement à son environnement.

### 4.3 Description détaillée de MGPPC

Cette section décrit le *Modèle Graphique Probabiliste de Prévention de la Contention* (MGPPC) pour les réseaux OBS.

#### 4.3.1 Description du réseau bayésien

Le modèle proposé pour effectuer le routage (qu'on nomme MGPPC) est composé d'évidences et de noeuds de décision (Fig. 4.2). Chaque noeud OBS contient un réseau bayésien (Fig. 4.3). La fonctionnalité primaire de MGPPC est la sélection du prochain saut dans le processus d'acheminement des BHP.

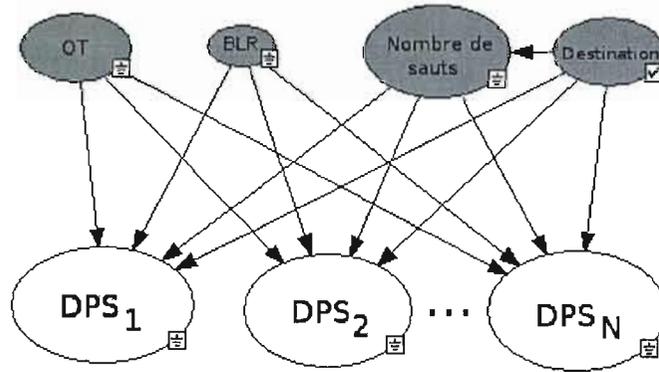


FIG. 4.2 – Le modèle graphique probabiliste de routage (MGPPC), un réseau bayésien

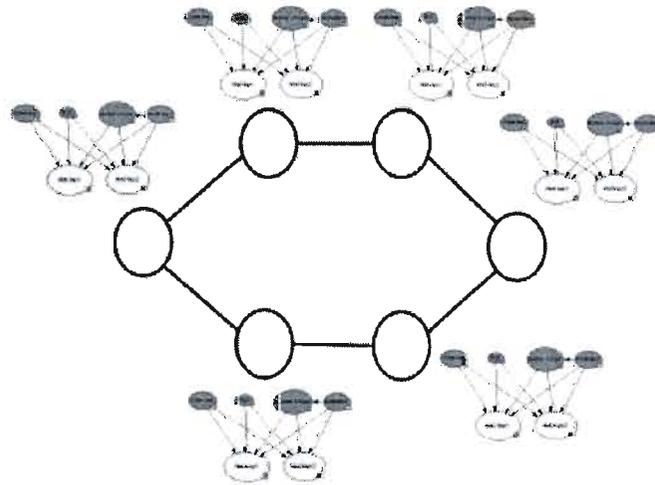


FIG. 4.3 – Exemple de topologie simple avec MGPPC

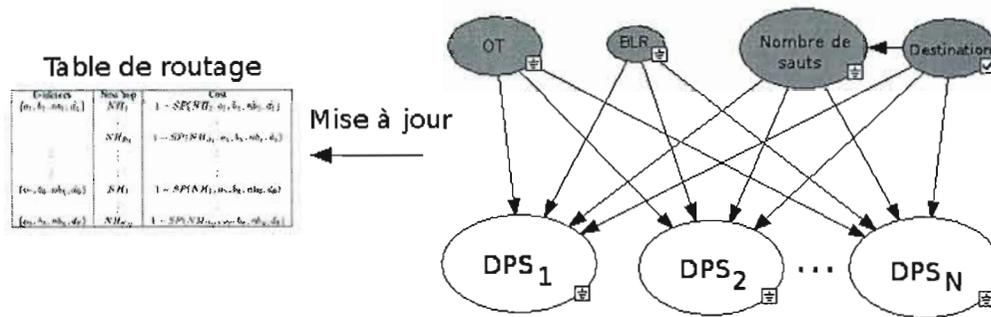


FIG. 4.4 – MGPPC qui met à jour la table de routage

Évidemment, ce processus d'acheminement doit être rapide et une table de routage standard est mise à jour périodiquement par le modèle proposé (Fig. 4.4). Cette table de routage est utilisée lorsque le BHP tente, dans le domaine électronique, de réserver une longueur d'onde pour la rafale de données.

MGPPC est composé de quatre évidences et d'un noeud de décision pour chaque prochain saut adjacent. Un noeud bayésien de décision d'un prochain saut ( $DPS_i$ ) est utile pour obtenir la probabilité de succès de réserver une longueur d'onde vers un noeud  $i$ . Une recherche dans la table de routage est effectuée en tenant compte des évidences dans le but d'obtenir le meilleur prochain saut en termes de probabilité de succès d'acheminer une rafale à la destination. MGPPC comporte les évidences suivantes :

- Offset time ( $OT$ ) : Le  $OT$  a un impact significatif sur le taux de perte puisque si le  $OT$  est insuffisant, la rafale est rejetée. Par conséquent, le  $OT$  est une métrique importante pour sélectionner le prochain saut. Les états de cette évidence correspondent aux nombres de sauts pour acheminer une rafale à la destination. 0 à 15 a été utilisé dans le cadre de ce travail.
- BLR ( $B$ ) : Le BLR est utilisé pour catégoriser les statistiques. Dans ce travail, cette variable peut être dans 3 états différents : faible, moyen et grand (Fig. 4.5).
- Nombre de sauts ( $NS$ ) : Les états de cette variable dépendent de la destination.

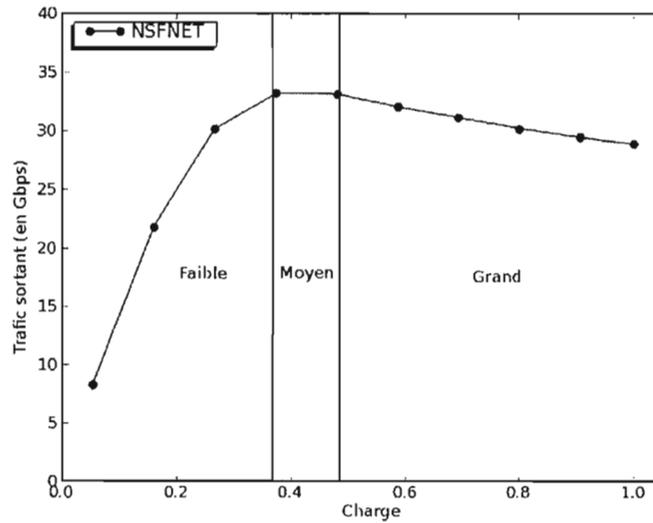


FIG. 4.5 – Délimitations de trafic en utilisant le débit sortant : Faible, moyen et grand

– Destination ( $D$ ) : Les états possibles de cette évidence sont les identifiants des noeuds OBS.

MGPPC inclut un noeud de décision pour chaque prochain saut possible. Chaque noeud de décision a deux états possibles : Succès (noté  $\oplus$ ) et Échec (noté  $\ominus$ ).

Soit  $k$  un identifiant de noeud OBS,  $DPS_k$  un noeud bayésien de décision du prochain saut  $k$ , la fonction de la probabilité jointe de MGPPC est définie comme suit :

$$\begin{aligned}
 P(OT, B, NS, D, DPS_1, \dots, DPS_N) = & \\
 & \left( \prod_{i=1}^N P(DPS_i | OT, B, NS, D) \right) * \\
 & P(OT) * P(B) * P(NS|D) * P(D)
 \end{aligned} \tag{4.3}$$

Le maximum a posteriori de  $DPS_k$  est défini par :

$$MAP_{DPS_k} = \arg \max_{\varphi} P(\varphi|ot, b, ns, d) \quad (4.4)$$

où  $\varphi \in \{\oplus, \ominus\}$  est une valeur possible du noeud  $DPS_k$  et où  $ot, b, ns, d$  sont des valeurs possibles des évidences dans le réseau bayésien ( $ot$  par exemple est un état possible de la variable OT).

$$MAP_{DPS_k} = \arg \max_{\varphi} \frac{P(\varphi)P(ot, b, ns, d|\varphi)}{P(ot, b, ns, d)} \quad (4.5)$$

$$MAP_{DPS_k} = \arg \max_{\varphi} P(\varphi)P(ot, b, ns, d|\varphi) \quad (4.6)$$

Si on assume que les évidences sont indépendantes, alors le maximum à posteriori peut être approximé par :

$$MAP_{DPS_k} \approx \arg \max_{\varphi} P(\varphi)P(ot|\varphi)P(b|\varphi)P(ns|\varphi)P(d|\varphi) \quad (4.7)$$

On définit  $P_{Succes}(DPS_k, ot, b, ns, d)$  (probabilité de succès,  $DPS_k = \oplus$ ), qui sera utilisé plus loin, comme étant l'approximation de  $MAP_{DPS_k}$  quand  $DPS_k = \oplus$ .

### 4.3.2 Table de routage

MGPPC utilise une table de routage différente comparativement aux approches classiques utilisant par exemple les chemins les plus courts. Une table de routage spécifique est nécessaire pour tenir compte des informations du réseau bayésien. L'approche courante pour construire une table de routage consiste à utiliser les métriques  $\langle Destination, Prochain\ saut, Cout \rangle$ . MGPPC utilise  $\langle Evidences, Prochain\ saut, Cout \rangle$  où les évidences ajoutent de la granularité dans le but d'acheminer les rafales plus efficacement. La table de routage MGPPC est définie dans le tableau 4.1.

TAB. 4.1 – Table de routage MGPPC

Évidences	Prochain saut (PS)	Coût
$\{ot_1, b_1, ns_1, d_1\}$	$PS_1^1$	$1 - P_{Succes}(DPS_{PS_1^1}, ot_1, b_1, ns_1, d_1)$
	$\vdots$	$\vdots$
	$PS_{\beta_1}^1$	$1 - P_{Succes}(DPS_{NH_{\beta_1}^1}, ot_1, b_1, ns_1, d_1)$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$\{ot_\gamma, b_\delta, ns_\eta, d_\theta\}$	$PS_1^{PE}$	$1 - P_{Succes}(DPS_{PS_1^{PE}}, ot_\gamma, b_\delta, ns_\eta, d_\theta)$
	$\vdots$	$\vdots$
	$PS_{\beta_{PE}}^{PE}$	$1 - P_{Succes}(DPS_{PS_{\beta_{PE}}^{PE}}, ot_\gamma, b_\delta, ns_\eta, d_\theta)$

$\beta_i$  exprime le nombre de prochains sauts selon les évidences de la  $i$  permutation,  $\gamma, \delta, \eta, \theta$  sont les nombres d'états des variables d'évidence.  $PS_i^j$  représente le  $i^{th}$  prochain saut de la  $j^{th}$  permutation d'évidence.  $PS_1^j$  représente le prochain ayant le coût minimal de la  $j^{th}$  permutation d'évidence. Par exemple, si on a la permutation d'évidence  $\{ot_1, b_1, ns_1, d_1\}$  et  $\beta_1 = 3$ , les prochains sauts pourraient être  $\{1, 4, 8\}$ . Les lignes de la colonne *Évidences* représentent les combinaisons possibles de tous les états de toutes les évidences. Le nombre de permutations d'évidences (PE) est exprimé par :

$$PE = \gamma * \delta * \eta * \theta \quad (4.8)$$

Donc, le nombre de lignes dans la table de routage est définie par :

$$N_{Lignes} = \sum_{i=1}^{PE} \beta_i \quad (4.9)$$

Le coût est exprimé par :

$$Cout(PS, ot, b, ns, d) = 1 - P_{Succes}(PS, ot, b, ns, d) \quad (4.10)$$

Les prochains sauts sont triés comme suit :

$$\forall_{i=1}^{PE} \forall_{j=1}^{\beta_i-1} Cout(PS_j^i, ot, b, ns, d) \leq Cout(PS_{j+1}^i, ot, b, ns, d) \quad (4.11)$$

L'extraction du meilleur prochain saut en termes probabiliste s'effectue à l'aide de l'algorithme 5.

---

**Algorithm 5:** Extraction du meilleur saut selon les évidences à l'aide de la table de routage MGPPC.

---

**Data:**  $N$ , un ensemble des identifiants des noeuds OBS.

- 1 À chaque noeud  $i \in N$
  - 2 **foreach**  $BHP$  **do**
  - 3     **Extraire les évidences**  $\{ot, b, ns, d\}$  dans BHP
  - 4     **Trouver** la ligne correspondante ( $k$ ) dans la table de routage MGPPC selon les évidences.
  - 5     **Sélectionner**  $PS_1^k$ .
  - 6 **end**
-

L'algorithme 6 permet de transformer le réseau bayésien (Section 4.3.1) dans une table de routage rapide mise à jour périodiquement.

---

**Algorithm 6:** Transformation du réseau bayésien MGPPC vers la table de routage MGPPC.

---

**Data:**  $N$ , un ensemble des identifiants des noeuds OBS.

**Data:**  $T$ , l'intervalle de temps de mise à jour.

```

1  À chaque noeud  $i \in N$ 
2  À chaque temps  $T$ 
3  foreach Permutation d'évidence  $j^{th}$   $\{ot, b, ns, d\}$  do
4  |   Localiser  $\{ot, b, ns, d\}$  dans la table de routage.
5  |   Obtenir les identifiants des prochains sauts ( $ids$ ) selon les évidences  $\{ot, b, ns, d\}$ .
6  |    $index \leftarrow []$ 
7  |   foreach  $id \in ids$  do
8  |   |   Ajouter  $(id, 1 - P_{Succes}(DPS_{id}, ot, b, ns, d))$  par ordre de coût dans  $index$ .
9  |   end
10 |   Associer  $\{ot, b, ns, d\}$  à  $index$  dans la table de routage.
11 end
```

---

### 4.3.3 Signalisation et paquets de notification

Le modèle proposé utilise une signalisation et des paquets de notification similaires à ceux de l'algorithme AHDR (Section 3.2.1). La seule différence est que les paquets de notification contiennent les évidences pour mettre à jour les réseaux bayésiens lors de la réception de ces paquets de contrôle.

L'algorithme 7 décrit la réception d'un paquet de notification et la mise à jour du réseau bayésien selon les évidences.

---

**Algorithm 7:** Réception d'un paquet de notification et mise à jour du réseau bayésien selon les évidences.

---

**Data:**  $N$ , un ensemble des identifiants des noeuds OBS.

```

1 À chaque noeud  $i \in N$ 
2 foreach Réception ACK/NACK do
3   Extraire les évidences  $\{ot, b, ns, d\}$ 
4   Extraire le dernier saut ( $DS$ ).
5   Extraire le type du paquet de notification ( $TPN$ ).
6    $SP \leftarrow P_{Succes}(DPS_{DS}, ot, b, ns, d)$ 
7    $P_{Succes}(DPS_{DS}, ot, b, ns, d)' \leftarrow \alpha SP + (1 - \alpha)A$ 
8   Où  $\alpha \in [0..1]$  et  $A$  est défini par :
9
10
11 end

```

$$A = \begin{cases} 1 & \text{if } TPN = ACK \\ 0 & \text{if } TPN = NACK \end{cases}$$

10 Mettre à jour le noeud du réseau bayésien  $DPS_{DS}$  tel que  
 $P_{Succes}(DPS_{DS}, ot, b, ns, d)'$  est la nouvelle valeur dans la table de probabilités  
conditionnelles selon les évidences.

## 4.4 Résultats expérimentaux et discussion

### 4.4.1 Configurations expérimentales

Des simulations ont été effectuées à l'aide du simulateur ns-2. La librairie *Structural Modeling, Inference, and Learning Engine* (SMILE) [lie] est utilisée pour les réseaux bayésiens. Des configurations similaires à celles utilisées dans les simulations de AHDR ont été utilisées

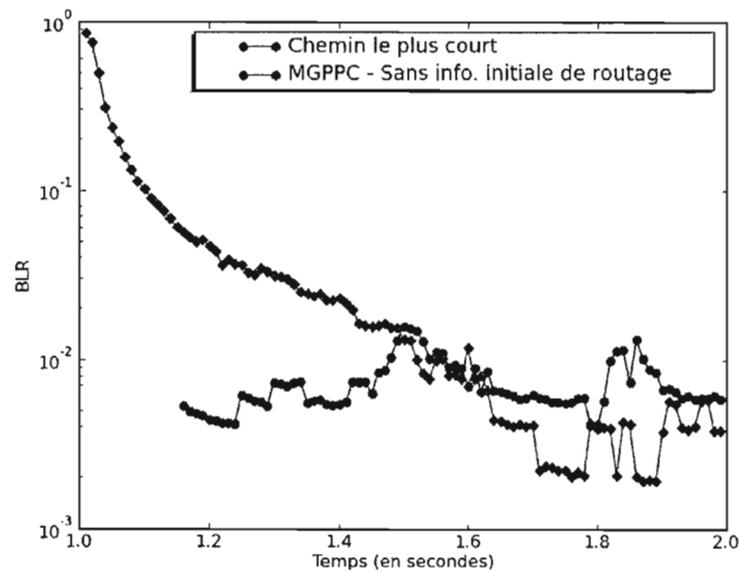


FIG. 4.6 – Apprentissage de NSFNET à l’aide de MGPPC, sans information de routage initiale

(Section 3.3.1). Les différences seront discutées dans les sections qui suivent.

#### 4.4.2 Résultats et analyse

##### Apprentissage de la topologie sans information initiale de routage

Un bénéfice implicite de MGPPC est la capacité qu’un noeud OBS a, sans information initiale de routage, d’apprendre ses prochains sauts dans le but de distribuer le trafic efficacement. Cela peut être utile lorsque des fautes arrivent dans une topologie pour que des chemins alternatifs soient utilisés ou lorsque la topologie change. L’algorithme du chemin le plus court utilise typiquement des informations initiales à propos du routage, comme par exemple les prochains sauts et les nombres de sauts pour atteindre les destinations. L’algorithme MGPPC se stabilise après environ 1 seconde. Après ce temps de stabilisation, l’algorithme

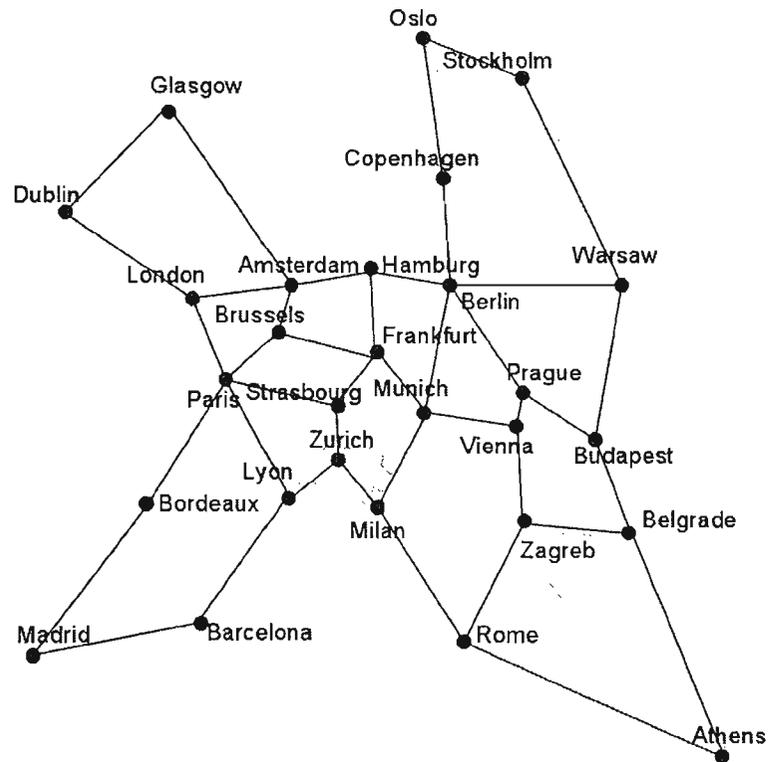


FIG. 4.7 – Topologie Pan-European (Provient de [HBB<sup>+</sup>03])

distribue le trafic au moins aussi efficacement qu'avec l'algorithme du chemin le plus court (Fig. 4.6).

### Comparaison de MGPPC et le chemin le plus court

Cette section comporte une comparaison de MGPPC et du chemin le plus court avec les topologies NSFNET (Fig. 3.7) et Pan-European (Fig. 4.7). Pour rendre plus réalistes les simulations, les connexions sont distribuées à travers le réseau proportionnellement aux matrices de trafic définies par Hülsermann et al. [HBB<sup>+</sup>03] Pour ces simulations, on assume que les

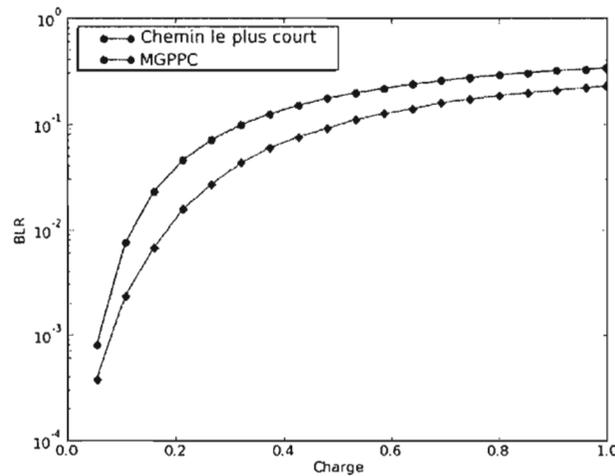


FIG. 4.8 – Variation du BLR en fonction de la charge, topologie NSFNET

informations de routage initiales (nombre de sauts, prochains sauts, etc.) sont disponibles. Il est à noter que MGPPC n'utilise ni la déflexion ni la retransmission pour résoudre la contention. Des meilleurs chemins sont utilisés dans le but de prévenir ces contentions.

Des améliorations notables sont obtenues en termes de BLR avec les topologies NSFNET (Fig. 4.8) et Pan-European (Fig. 4.11). On remarque que les gains en termes de BLR pour la topologie Pan-European sont beaucoup plus grands. Ceci s'explique par le fait que cette topologie est fortement connectée comparativement à NSFNET. Cette forte connectivité offre à MGPPC plus de chemins possibles tandis que l'algorithme du chemin le plus court utilise toujours le même chemin, indépendamment du trafic.

Avec MGPPC, le réseau est plus utilisé (Fig. 4.10 pour NSFNET et Fig. 4.13 pour Pan-European) pour deux raisons principales :

- Des chemins plus longs peuvent être utilisés, ce qui fait augmenter globalement l'utilisation des liens du réseau.
- Moins de rafales sont rejetées à cause des contentions, ce qui fait augmenter l'utilisation

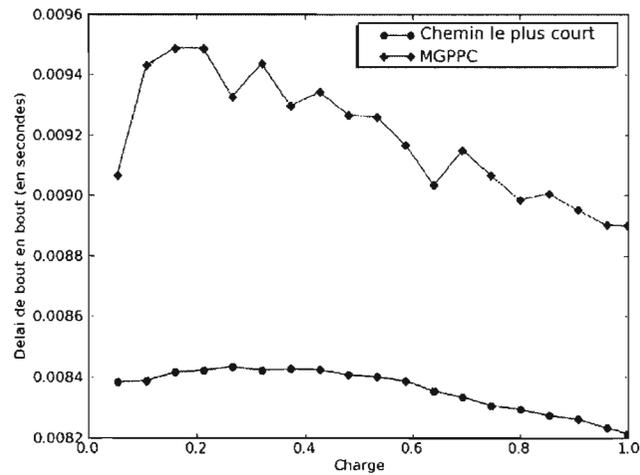


FIG. 4.9 – Variation du délai de bout en bout en fonction de la charge, topologie NSFNET

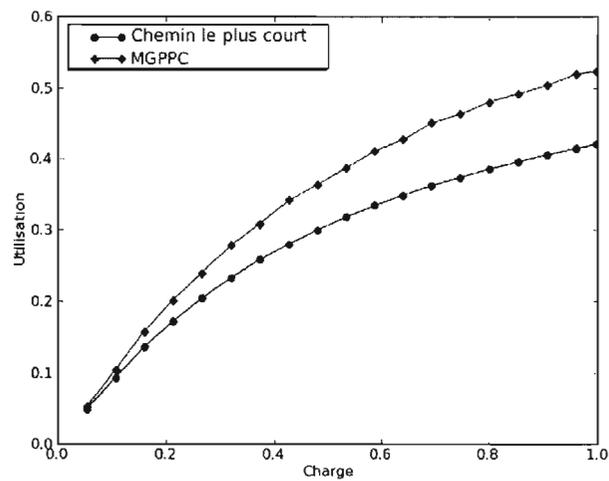


FIG. 4.10 – Variation de l'utilisation en fonction de la charge, topologie NSFNET

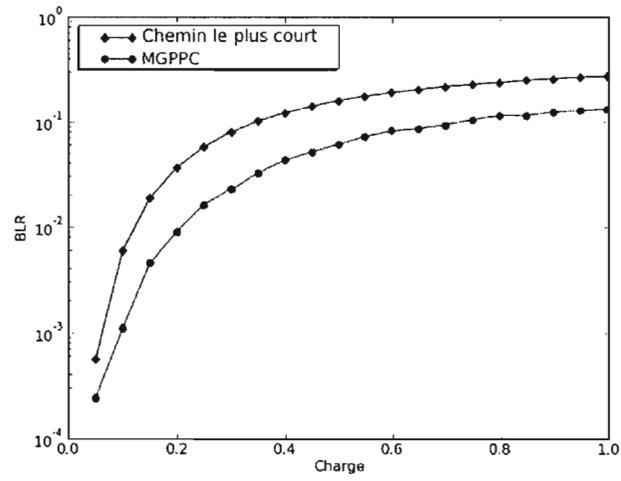


FIG. 4.11 – Variation du BLR en fonction de la charge, topologie Pan-European

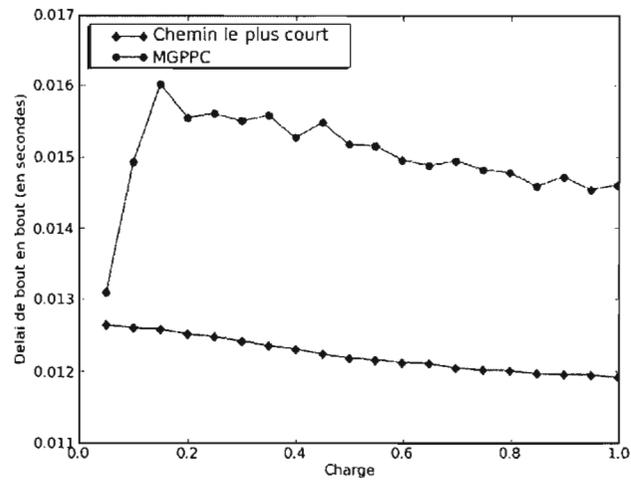


FIG. 4.12 – Variation du délai de bout en bout en fonction de la charge, topologie Pan-European

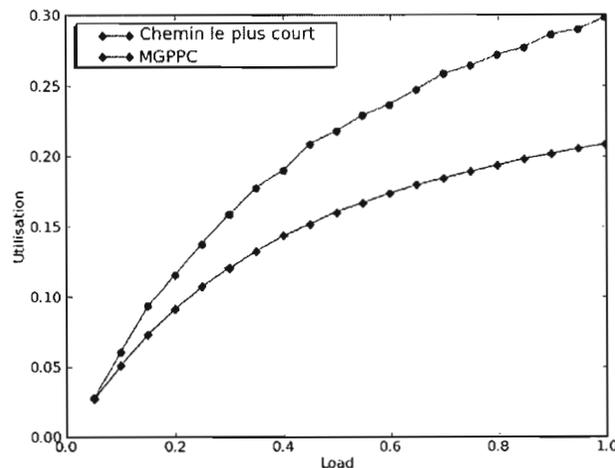


FIG. 4.13 – Variation de l'utilisation en fonction de la charge, topologie Pan-European des liens.

Des gains en termes d'utilisation des liens allant jusqu'à 10 % sont observés pour les topologies NSFNET et Pan-European.

Quel est le coût de diminuer les pertes et d'augmenter l'utilisation des liens à l'aide de MGPPC? Le délai de bout en bout est supérieur comparativement au plus court chemin. Environ 1 ms de plus est utilisé par MGPPC avec la topologie NSFNET (Fig. 4.9) et 3 ms supplémentaires avec la topologie Pan-European (Fig. 4.12). 1 ou 3 ms supplémentaire de délai de bout en bout n'a aucun impact sur les protocoles de transport tels que TCP.

## 4.5 Conclusion

Le modèle MGPPC propose une méthode préventive de routage pour diminuer la probabilité qu'une contention arrive. Ce modèle utilise un réseau bayésien permettant de construire des tables de routage *intelligentes* pour sélectionner les meilleures routes possibles pour

diminuer le BLR.

Le modèle bayésien permet de notablement diminuer le BLR sur des topologies faiblement ou fortement connectées telles que NSFNET et Pan-European avec un trafic réaliste où la charge des liens n'est pas uniformément distribuée. Un coût en termes de délai de bout en bout de 1-4 ms est observé. Ce coût ne pose pas de problème pour la majorité des applications actuelles.

## Chapitre 5

### Conclusion et perspectives

Ce travail montre premièrement que lorsqu'une contention se produit dans un réseau OBS, le choix entre effectuer une déflexion et une retransmission, pour résoudre la contention, est très important. L'algorithme AHDR proposé utilise une méthode dynamique qui tient compte de l'état du réseau. Les approches statiques proposées dans la littérature telles que LHDR avec des topologies faiblement connectées telles que NSFNET effectuent pas assez de déflexions lorsque la charge est faible et trop de déflexions lorsque la charge augmente. L'algorithme AHDR proposé permet de contrer cet effet en effectuant des déflexions tant que celles-ci ne déstabilisent pas le réseau tout en diminuant les pertes. Sur des topologies fortement connectées telles que COST239, l'algorithme AHDR permet de diminuer drastiquement les pertes en utilisant un ratio de déflexion très grand.

Le modèle bayésien (MGPPC) proposé dans ce travail démontre le potentiel en réseau des modèles graphiques probabilistes utilisés en intelligence artificielle. Le modèle MGPPC permet de construire des tables de routage avec une granularité d'information beaucoup plus fine, rendant le routage plus efficace pour prévenir les contentions pour diminuer les pertes.

Les travaux futurs liés à ce présent travail sont l'intégration de la qualité de service dans les modèles AHDR et MGPPC, l'adaptation dynamique du nombre de retransmissions et

du temps de latence entre les retransmissions, la détermination du OT pour adapter les déflexions en utilisant une méthode comme les réseaux bayésiens et l'étude de la combinaison des approches réactives et préventives pour diminuer les pertes.

# Bibliographie

- [Aca94] A. S. Acampora. The Scalable Lightwave Network. *IEEE Communications Magazine*, 32(12) :36–42, 1994.
- [CDL03] M.-T. Choy, Y. Deng, and T. T. Lee. Design of Optical Burst Switches based on Dual Shuffle-exchange Network and Deflection Routing. *Symposium on High Performance Interconnects*, pages 102–107, 2003.
- [CEJ05] T. Coutelen, H. Elbiaze, and B. Jaumard. Performance Comparison of OCS and OBS Switching Paradigms. *Proceedings of 2005 7th International Conference on Transparent Optical Networks*, 2005.
- [CEJM05] T. Coutelen, H. Elbiaze, B. Jaumard, and A. Metnani. Measurement-Based Alternative Routing Strategies in Optical Burst-Switched Networks. *Proceedings of 7th International Conference Transparent Optical Networks*, 1, 2005.
- [CQY04] Y. Chen, C. Qiao, and X. Yu. Optical Burst Switching : A New Area in Optical Networking Research. *Network, IEEE*, 18 :16–23, 2004.
- [CWXQ03] Y. Chen, H. Wu, D. Xu, and C. Qiao. Performance Analysis of Optical Burst Switched Node with Deflection Routing. *International Conference on Communications (ICC)*, 2 :1355–1359, 2003.
- [CZZ05] C. Cameron, A. Zalesky, and M. Zukerman. Prioritized Deflection Routing in Optical Burst Switching Networks. *IEICE Transactions on Communication*, E88-B(5) :1861–1867, 2005.

- [DDF04] A. K. Dutta, N. K. Dutta, and M. Fujiwara. *WDM Technologies : Optical Networks*. Academic Press, 2004.
- [Dea02] T. Dean. *Réseaux Informatiques*. Reynald Goulet, 2002.
- [Dix03] S. Dixit. *IP over WDM*. Wiley, 2003.
- [DL01] A. Detti and M. Listanti. Application of Tell & Go and Tell & Wait Reservation Strategies in a Optical Burst Switching Network : a Performance Comparison. *Proceedings of IEEE International Conference on Telecommunication (ICT)*, 2 :540–548, 2001.
- [EJM04] H. Elbiaze, B. Jaumard, and A. Metnani. Network Topology Impact on Optical Burst Deflection Routing. Technical report, Université du Québec à Montréal, 2004.
- [Gau02] C. Gauger. Performance of Converter Pools for Contention Resolution in Optical Burst Switching. *OptiComm 2002 : Optical Networking and Communications*, 2002.
- [Gau06] C. Gauger. *Novel Network Architecture for Optical Burst Transport*. PhD thesis, Universität Stuttgart, Stuttgart, Germany, 2006.
- [GBPS03] C. Gauger, H. Buchta, E. Patzak, and J. Saniter. Performance meets Technology - an Integrated Evaluation of OBS Nodes with FDL Buffers. *Proceedings of the 1st International Workshop on Optical Burst Switching*, 2003.
- [Gje07] S. Gjessing. On Burst Loss in Optical Burst Switched Networks with Hot Potato Deflection Routing. *Conference on Digital Society*, pages 19–25, 2007.
- [GKS04] C. Gauger, M. Köhn, and J. Scharf. Performance of Contention Resolution Strategies in OBS Network Scenarios. *Proceedings of the 9th Optoelectronics and Communications Conference/3rd International Conference on the Optical Internet (OECC/COIN2004)*, 2004.
- [GZ05] D. Gao and H. Zhang. Reservation Information Sharing Enhancement for Deflection Routing in OBS Network. *Optics Express*, 13(5) :1702–1709, 2005.

- [HBB<sup>+</sup>03] R. Hülsermann, S. Bodamer, M. Barry, A. Betker, C. Gauger, M. Jäger, M. Köhn, and J. Späth. Reference Transport Network Scenarios. *MultiTeraNet Report*, 2003.
- [HL05] C.-F. Hsu and T.-L. Liu. On Deflection Routing in Optical Burst-Switched Networks. *Conference on Digital Society*, 14(4) :341–362, 2005.
- [HLH02a] C. Hsu, T. Liu, and N. Huang. Performance Analysis of Deflection Routing in Optical Burst-Switched Networks. *Proceedings of IEEE INFOCOM*, 2002.
- [HLH02b] C.-F. Hsu, T.-L. Liu, and N.-F. Huang. On the Deflection Routing in QoS Supported Optical Burst-Switched Networks. *International Conference on Communications*, 5 :2786–2790, 2002.
- [JV04] J. P. Jue and V. M. Vokkarane. *Optical Burst Switched Networks*. Springer, 2004.
- [KKK02] S. Kim, N. Kim, and M. Kang. Contention Resolution for Optical Burst Switching Networks Using Alternative Routing. *Proceedings of IEEE ICC*, 5 :2687–2691, 2002.
- [lie] Structural Modeling, Inference, and Learning Engine (SMILE), une librairie C++ pour les modèles graphiques probabilistes et les modèles de décisions théoriques, <http://genie.sis.pitt.edu/>.
- [LKS<sup>+</sup>05] S. Lee, L. Kim, J. Song, D. Griffith, and K. Sriram. Dynamic Deflection Routing with Virtual Wavelength Assignment in Optical Burst-Switched. *Photonic Network Communications*, 9(3) :347–356, 2005.
- [LSKS05] S. Lee, K. Sriram, H. Kim, and J. Song. Contention-Based Limited Deflection Routing Protocol in Optical Burst-Switched Networks. *Journal on Selected Areas in Communications, IEEE*, 23(8) :1596–1611, 2005.
- [MBM04] A. Maach, G. V. Bochman, and H. Mouftah. Congestion Control and Contention Elimination in Optical Burst Switching. *Telecommunication Systems*, 27 :115–131, 2004.

- [Met04] A. Metnani. Routage par Déflexion dans les Réseaux Tout Optique à Commutation de Bursts. Master's thesis, Université de Montréal, 2004.
- [MM05] David S. Moore and George P. McCabe. *Introduction to the Practice of Statistics*. Springer, 2005.
- [MS99] A. Malis and W. Simpson. PPP over SONET/SDH. RFC 2615 (Proposed Standard), 1999.
- [Muk06] B. Mukherjee. *Optical WDM Networks*. Springer, 2006.
- [Nea03] R. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 2003.
- [NJH06] S.-H. Ngo, X. Jian, and S. Horiguchi. Hybrid Deflection and Retransmission Routing Schemes for OBS Networks. *Workshop on High Performance Switching and Routing*, 2006.
- [NJHG06] S.-H. Ngo, X. Jiang, S. Horiguchi, and M. Guo. Performance Evaluation of Hybrid Deflection and Retransmission Routing for OBS Networks. *International Symposium on Broadband Access Technologies*, 2006.
- [NS] Network Simulator 2 (ns-2), un simulateur de réseaux événementiel, <http://www.isi.edu/nsnam/ns/>.
- [OT05] N. Ogino and H. Tanaka. Deflection Routing for Optical Bursts Considering Possibility of Contention at Downstream Nodes. *IEICE Transactions on Communication*, E88-B(9) :3660–3667, 2005.
- [Pay08] D. Payne. World bandwidth growth over the next decade - is it viable? Technical report, Institute of Advanced Telecommunications, Swansea University, 2008.
- [Pos81] J. Postel. Internet Protocol. RFC 791 (Standard), 1981. Updated by RFC 1349.
- [QY99] C. Qiao and M. Yoo. Optical Burst Switching - A New Paradigm for an Optical Internet. *Journal of High Speed Networks*, 8(1) :69–84, 1999.
- [Rou01] G. N. Rouskas. Routing and Wavelength Assignment in Optical WDM Networks. *Wiley Encyclopedia of Telecommunications*, 2001.

- [RS01] R. Ramaswami and K. Sivarajan. *Optical Networks : A practical perspective*. Morgan Kaufmann, 2001.
- [SG05] M. Schlosser and E. Patzak P. Gelpke. Impact of Deflection Routing on TCP Performance in Optical Burst Switching Networks. *Proceedings of 2005 7th International Conference on Transparent Optical Networks*, 1 :220–223, 2005.
- [TGCT99] L. Tancevski, A. Ge, G. Castanon, and L. Tamil. A New Scheduling Algorithm for Asynchronous, Variable Length IP Traffic Incorporating Void Filling. *Optical Fiber Communication Conference (OFC)*, 3 :180–182, 1999.
- [TOC03] H. Tanida, K. Ohmae, and Y. Choi. An Effective BECN/CRN Typed Deflection Routing for QoS Guaranteed Optical Burst Switching. *Global Telecommunications Conference*, 2003.
- [Tof01] Z. Toffano. *Optoélectronique : Composants photoniques et fibres optiques*. Ellipses Marketing, 2001.
- [Tur99] J. S. Turner. Terabit Burst Switching. *Journal of High Speed Networks*, 8 :3–16, 1999.
- [UCCR06] T.-W. Um, J. K. Choi, S. G. Choi, and W. Ryu. Performance of a Partial Burst Retransmission Mechanism in OBS Networks. *International Conference on Networking and Services (ICNS)*, pages 108–114, 2006.
- [Vok01] V. M. Vokkarane. Burst Segmentation : A New Approach To Reduce Packet Loss in Optical Burst Switched Networks. Master's thesis, University of Texas at Dallas, 2001.
- [Vok04] V. M. Vokkarane. *Design and Analysis of Achitectures and Protocols for Optical Burst-Switched Networks*. PhD thesis, The University of Texas, Texas Dallas, 2004.
- [WMA00a] X. Wang, H. Morikawa, and T. Aoyama. Burst Optical Deflection Routing Protocol for Wavelength Routing WDM Networks. *Proceedings of OptiComm*, pages 120–129, 2000.

- [WMA00b] X. Wang, H. Morikawa, and T. Aoyama. Deflection Routing Protocol for Burst Switching WDM Mesh Networks. *Proceedings of SPIE/IEEE Terabit Optical Networking : Architecture, Control, and Management Issues*, pages 242–252, 2000.
- [XYYH05] F. Xue, S. J. Ben Yoo, H. Yokoyama, and Y. Horiuchi. Performance Comparison of Optical Burst and Circuit Switched Networks. *Technical Digest of IEEE/OSA Optical Fiber Communication Conference*, 2005.
- [YQ97] M. Yoo and C. Qiao. Just-Enough-Time (JET) : a High Speed Protocol for Bursty Traffic in Optical Networks. *IEEE/LEOS Technologies for a Global Information Infrastructure*, 1997.
- [ZVR<sup>+</sup>04] A. Zalesky, H. L. Vu, Z. Rosberg, E. W. M. Wong, and M. Zukerman. Modelling and Performance Evaluation of Optical Burst Switched Networks with Deflection Routing and Wavelength Reservation. *Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, 3 :1864–1871, 2004.
- [ZVR<sup>+</sup>07] A. Zalesky, H. L. Vu, Z. Rosberg, E. W. M. Wong, and M. Zukerman. Stabilizing Deflection Routing in Optical Burst Switched Networks. *Journal on Selected Areas in Communications*, 25(6) :3–19, 2007.
- [ZVWJ05] Q. Zhang, V. M. Vokkarane, Y. Wang, and J. P. Jus. Evaluation of Burst Retransmission in Optical Burst-Switched Networks. *Conference on Broadband Networks*, 1 :276–282, 2005.
- [ZVZ<sup>+</sup>04] A. Zalesky, H. L. Vu, M. Zukerman, Z. Rosberg, and E. W. M. Wong. Evaluation of Limited Wavelength Conversion and Deflection Routing as Methods to Reduce Blocking Probability in Optical Burst Switched Networks. *Conference on Communications*, 3 :1543–1547, 2004.

## Publications réalisées

- [1] Martin Lévesque and Halima Elbiaze, Graphical Probabilistic Routing Model for OBS Networks with Realistic Traffic Scenario, *IEEE Globecom 2009*, Hawaii, United States, November 30 - December 4, 2009.
- [2] Martin Lévesque, Halima Elbiaze, and Wael Hosny Fouad Aly, Adaptive Threshold-based Decision for Efficient Hybrid Deflection and Retransmission Scheme in OBS Networks, *13th Conference on Optical Network Design and Modeling (ONDM 2009)*, Braunschweig, Germany, February 18-20, 2009.
- [3] Martin Lévesque, Halima Elbiaze, and Wael Hosny Fouad Aly, Adaptive Hybrid Deflection and Retransmission Routing for Optical Burst-Switched Networks, *The 7th International Conference on Optical Communications and Networks (ICOON 2008)*, Singapore, Dec 8-11, 2008.