

N° d'ordre: num

# THÈSE

présentée

devant l'université de Nice Sophia-Antipolis

pour obtenir

le grade de : DOCTEUR DE L'UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS  
Mention INFORMATIQUE

par

Corinne TOUATI

Équipe d'accueil : MISTRAL/MASCOTTE

École doctorale : UNSA

Composante universitaire : UNIVERSITÉ NICE SOPHIA-ANTIPOLIS / INRIA  
SOPHIA-ANTIPOLIS

Titre de la thèse :

*Les principes d'équité appliqués aux réseaux de télécommunications*

À soutenir le Septembre 2003 devant la commission d'examen

M. :	presP	PRESN	Président
MM. :	Véronique	VÈQUE	Rapporteurs
	Ravi	MAZUMDAR	
MM. :	Philippe	NAIN	Examineurs
	Mustapha	BOUHTOU	
	Jérôme	GALTIER	
	Eitan	ALTMAN	



*A ma mère.*



# *Remerciements*

Je remercie chaleureusement Eitan Altman, mon directeur de thèse pour sa grande disponibilité, ses compétences et la confiance qu'il m'a accordé pendant ces trois années. Je dois également une grande partie de mon travail à Jérôme Galtier qui a co-encadré ma thèse. Nos réunions régulières ont toujours guidé mes recherches.

Je tiens à remercier Véronique Vèque et Ravi Mazumdar d'avoir bien voulu accepter la charge de rapporteur. Un grand merci également à Mustapha Bouhtou et Philippe Nain d'avoir bien voulu juger ce travail.

Je remercie également tout particulièrement messieurs Zui Altman et Jean-Paul Kélif, de France Telecom R&D, pour leurs explications du fonctionnement des réseaux mobiles de troisième génération et toutes leurs précisions sur le développement de l'UMTS.

Je remercie également les professeurs Hisao Kameda et Li Jie de l'Université de Tsukuba au Japon pour m'avoir accueillie en stage pendant l'hiver 2002 et m'avoir introduite aux problèmes de théorie des jeux non coopératifs. Je remercie toute l'équipe du laboratoire OSDP pour leur accueil chaleureux et leur collaboration fructueuse. Toute ma gratitude se porte particulièrement à Simona et Adrian Vasilache pour leurs conseils avisés et leur soutien pendant cette période de ma thèse.

Les travaux que nous avons effectué sur les réseaux satellitaires n'auraient pas été menés sans la collaboration avec Benoit Fabre et Isabelle Buret de l'entreprise Alcatel Space Industries à Toulouse et sans l'aide de Tuna Toker, un étudiant stagiaire en DEA Réseau et Systèmes Distribués à l'Université de Nice Sophia-Antipolis, qui a implémenté de l'algorithme de recuit simulé dans le cadre des communications par satellite.

Un grand merci aux membres des équipes de l'INRIA qui m'ont accueillis. Je remercie tous les MISTRALIens et aux MASCOTTiens pour m'avoir introduite aux problèmes de coloriage de graphes. Je suis très reconnaissante à notre chère secrétaire, Ephie Deriche, toujours prête à nous aider dans les pires soucis administratifs. J'en profite également pour saluer les joueurs de tarot invétérés de notre étage.

Je remercie également Ludovic Henrio, de l'équipe OASIS pour son aide précieuse et sa relecture attentive de mes articles et Séverine Petat pour son aide dans l'organisation impossible des séances de TD, ses blagues dans les couloirs et pour m'avoir dit que je savais courir correctement lorsque j'étais au bord de la crise d'asthme à cinq cent mètres de l'entrée de l'INRIA.

Enfin, je remercie enfin et surtout mes parents qui ont permis de rendre cette aventure possible, à plus d'un titre. C'est avec une émotion particulière que je remercie ma mère, qui a suivi avec attention et amour mes tâtonnements pendant le début de ma thèse. Je suis également extrêmement reconnaissante à mon père n'a jamais cessé de me soutenir et m'encourager à poursuivre ma route.

Merci donc à tous pour votre soutien et votre aide. Le travail, dont ce document est la synthèse, n'aurait pu être effectué sans vous.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Le contrôle de ressource . . . . .	4
1.1.1	Pourquoi le contrôle de congestion est-il nécessaire? . . . . .	4
1.1.2	Conclusion . . . . .	6
1.2	Définitions générales . . . . .	7
1.2.1	Ressource . . . . .	7
1.2.2	Connexion . . . . .	7
1.2.3	Allocation faisable . . . . .	7
1.2.4	Ordre parmi les allocations . . . . .	8
1.2.5	Efficacité . . . . .	8
1.2.6	Notations . . . . .	8
1.3	Allocations classiques . . . . .	9
1.3.1	Maximisation du débit total . . . . .	9
1.3.2	Équité Max-Min . . . . .	11
1.3.3	Équité proportionnelle . . . . .	16
1.3.4	Minimisation du délai potentiel . . . . .	18
1.3.5	Liens entre les équités . . . . .	18
1.3.6	Autres définitions d'équité . . . . .	19
1.4	Critère de Nash . . . . .	19
1.4.1	Deux mots sur la théorie des jeux . . . . .	20
1.4.2	Petit voyage dans le temps . . . . .	21
1.4.3	L'évolution . . . . .	24
1.4.4	Reformulation . . . . .	25
1.4.5	Adaptation pour les réseaux . . . . .	25
1.4.6	La résolution de Yaiche, Mazumdar et Rosenberg . . . . .	26
1.5	Algorithmes classiques de partage de bande passante . . . . .	27
1.5.1	Calculs explicites des taux . . . . .	27
1.5.2	Algorithmes décentralisés . . . . .	28
1.5.3	Conclusion . . . . .	33
1.6	Les fonctions d'utilité . . . . .	33
1.6.1	Les fonctions ON/OFF . . . . .	34
1.6.2	Applications élastiques . . . . .	35
1.6.3	Applications à débit variable . . . . .	35
1.6.4	Conclusion . . . . .	36
1.7	Comparaison et propriétés des critères d'équité . . . . .	37

1.7.1	Introduction . . . . .	37
1.7.2	Impact de la concavité de la fonction d'utilité . . . . .	38
1.8	La programmation semi-définie positive (SDP) . . . . .	40
1.8.1	De la programmation linéaire à la programmation Semi-Définie Positive . . . . .	40
1.8.2	Petits rappels d'algèbre linéaire. . . . .	46
1.8.3	Exemples classiques . . . . .	49
1.9	Conclusion . . . . .	50

## **I Les réseaux terrestres 51**

### **2 Introduction aux réseaux terrestres. 55**

2.1	Système de contraintes . . . . .	55
2.1.1	Notations . . . . .	55
2.1.2	Demandes des utilisateurs . . . . .	56
2.1.3	Types de routage . . . . .	56
2.2	Types d'optimisations . . . . .	58
2.2.1	L'optimisation "connexion-aware" . . . . .	58
2.2.2	L'optimisation "network-aware" . . . . .	58
2.2.3	Résumé et conclusion . . . . .	59
2.3	Conclusion . . . . .	59

### **3 Méthodes analytiques 61**

3.1	Exemples de calculs explicites . . . . .	61
3.1.1	Les allocations équitables dans le réseau linéaire . . . . .	61
3.1.2	Analyse asymptotique . . . . .	65
3.1.3	Réseau grille . . . . .	67
3.2	Méthode Lagrangienne . . . . .	68
3.2.1	Multiplicateurs de Lagrange . . . . .	68
3.2.2	Problème dual . . . . .	69
3.3	Conclusion . . . . .	70

### **4 Méthode SDP 73**

4.1	Philosophie générale de l'approche SDP . . . . .	73
4.1.1	Contraintes linéaires . . . . .	74
4.1.2	Relations entre les variables . . . . .	74
4.2	Les différentes valeurs de $\alpha$ . . . . .	75
4.2.1	Calculs d'équité pour $\alpha \neq 1$ . . . . .	76
4.2.2	Calcul du NBS (cas $\alpha = 1$ ) . . . . .	77
4.3	Un exemple simple de calcul du NBS dans le cas du routage fixé . . . . .	78
4.4	Expériences . . . . .	80
4.4.1	Calcul du NBS dans un petit réseau à routage fixé . . . . .	81
4.4.2	Le réseau COST . . . . .	83
4.5	Conclusion . . . . .	86



<b>II</b>	<b>Les réseaux mobiles</b>	<b>93</b>
<b>5</b>	<b>Précédents travaux sur le contrôle de taux dans les réseaux sans fil</b>	<b>97</b>
5.1	L'équité dans les réseaux sans fils. . . . .	98
5.2	Choix des débits . . . . .	99
5.3	Contrôle de puissance . . . . .	100
5.4	Contrôle joint des débits et des puissances . . . . .	102
5.5	Conclusion . . . . .	102
<b>6</b>	<b>Allocation jointe de débit et de puissance : cas général</b>	<b>105</b>
6.1	Introduction . . . . .	105
6.1.1	Pourquoi une allocation jointe entre débit et puissance? . . . . .	105
6.1.2	Un réseau, différents liens. . . . .	106
6.1.3	La suite... . . . .	107
6.2	Discussion générale sur l'intérêt de notre proposition . . . . .	108
6.2.1	Inconvénients des techniques actuelles . . . . .	108
6.2.2	But des travaux . . . . .	109
6.3	Le modèle . . . . .	109
6.3.1	Allocation des débits . . . . .	110
6.3.2	Modèle de puissance . . . . .	110
6.3.3	Lien entre puissance et débit . . . . .	111
6.3.4	Expression analytique du SIR . . . . .	111
6.3.5	Résumé : formulation du problème d'équité . . . . .	112
6.4	Propriétés du système . . . . .	113
6.4.1	Problème équivalent et concavité de la fonction objectif . . . . .	113
6.4.2	Propriétés des vecteurs de taux acceptables . . . . .	114
<b>7</b>	<b>Application du modèle aux différents liens et approximations possibles</b>	<b>121</b>
7.1	Cas du lien montant . . . . .	121
7.1.1	Notations et hypothèses. . . . .	121
7.1.2	Utilisation des résultats précédents . . . . .	123
7.2	Solution du lien descendant . . . . .	124
7.2.1	Notations et hypothèses . . . . .	124
7.2.2	Lien avec la formulation générale . . . . .	127
7.3	Macro-diversité dans le lien descendant . . . . .	128
7.3.1	SIR d'un terminal en situation de soft-handover . . . . .	129
7.3.2	SIR des mobiles dépendants d'une station de base donnée . . . . .	129
7.3.3	Application des résultats du chapitre précédent . . . . .	130
7.3.4	Conclusion . . . . .	130
7.4	Approximations possibles . . . . .	131
7.4.1	Premier schéma : approximation de la fonction objectif . . . . .	131
7.4.2	Second schéma d'approximation : approximation de l'ensemble des contraintes . . . . .	132
7.4.3	Conclusion . . . . .	132
7.5	Quelques résultats numériques et conclusion . . . . .	133

<b>8</b>	<b>Introduction aux réseaux satellitaires.</b>	<b>141</b>
8.1	Qu'est-ce qu'un réseau satellitaire? . . . . .	141
8.1.1	Orbites et empruntes. . . . .	142
8.1.2	Altitudes possibles . . . . .	143
8.1.3	Bandes de fréquence utilisées. . . . .	144
8.1.4	Spécificité des systèmes GEO . . . . .	145
8.1.5	Bilan de liaison . . . . .	145
8.1.6	Méthode d'accès . . . . .	146
8.2	Précédents travaux sur l'allocation de slot dans les systèmes TDMA et MF-TDMA . . . . .	148
8.2.1	Le problème TSA . . . . .	148
8.2.2	Allocation des burst . . . . .	155
8.2.3	Problèmes annexes . . . . .	156
8.2.4	Surveys et description de systèmes complets . . . . .	156
8.3	Notre problème . . . . .	159
8.3.1	Formulation du problème . . . . .	159
8.3.2	Système physique . . . . .	160
<b>9</b>	<b>Première partie de l'allocation : la planification radio.</b>	<b>163</b>
9.1	Introduction et notations . . . . .	163
9.1.1	Définitions générales . . . . .	164
9.1.2	Inter-Spot Compatibility Conditions (ISCC) . . . . .	164
9.1.3	Notations (résumé) . . . . .	165
9.1.4	Critères d'équité . . . . .	165
9.1.5	Formulation du problème . . . . .	166
9.2	Notre algorithme d'allocation . . . . .	167
9.2.1	Allocation des porteuses aux opérateurs dans un spot à $J$ fixé . . . . .	167
9.2.2	Obtention du vecteur $J$ minimal . . . . .	167
9.2.3	Etapas de résolutions de la solution générale . . . . .	168
9.3	Expérimentations . . . . .	170
9.3.1	Le réseau considéré . . . . .	170
9.3.2	Résultats numériques . . . . .	171
9.4	Conclusion . . . . .	172
<b>10</b>	<b>Deuxième partie de l'allocation : l'allocation des slots.</b>	<b>173</b>
10.1	Le modèle d'interférence. . . . .	173
10.2	Le recuit simulé . . . . .	174
10.2.1	Présentation générale du recuit simulé . . . . .	174
10.2.2	Recuit simulé efficace . . . . .	175
10.2.3	Améliorations de l'algorithme . . . . .	176
10.2.4	Résultats . . . . .	176
10.3	Résolution par programmation linéaire . . . . .	177
10.3.1	Présentation du problème . . . . .	177

10.3.2	Utilisation des familles . . . . .	178
10.3.3	Choix et placement des familles. . . . .	182
10.3.4	Résultats numériques . . . . .	183
10.3.5	Conclusion . . . . .	183
<b>Conclusion</b>		<b>190</b>
<b>A Programme de création des matrices d'entrées.</b>		<b>191</b>
<b>B Les réseaux satellitaires.</b>		<b>195</b>
B.1	Relation d'ordre parmi les allocations . . . . .	195
B.2	Optimalité de l'algorithme . . . . .	196

# *Objet de la thèse*

Les opérateurs de téléphones portables (GSM) basent désormais leurs publicités sur des arguments liés à la qualité de service. C'est ainsi que l'on entend fréquemment des phrases de type "notre réseau à été classé meilleur sur X critères". Alors qu'autrefois on essayait d'attirer les clients en leur proposant des formules supposées économiques (facturation à la seconde, report des minutes d'un mois à l'autre ou encore appels gratuits dans une certaine plage horaire), les arguments actuels sont désormais souvent basés sur la qualité proposée. Meilleure couverture, meilleure qualité de son, moins de blocage d'appels, accès à des services multimedia sont autant de critères de vente.

Considérons par exemple la couverture des réseaux. Le déploiement des différents réseaux de GSM en France a suivi à peu près le même schéma. Dans un premier temps, les grandes agglomérations sont couvertes. Puis viennent les stations balnéaires et les stations de ski, là où la demande est importante. Petit à petit les réseaux se développent dans les zones de population moins denses : le milieu rural et même dans les montagnes là où les coûts d'installations sont les plus grands et où les signaux radio se propagent mal. Au delà d'un improbable élan philanthrope de la part des opérateurs de téléphonie, on voit un argument commercial : le réseau ne lèse personne. Où que l'on se trouve, on peut désormais compter sur son portable. C'est l'ère de la mobilité. Désormais les gens peuvent acheter un téléphone sans crainte de ne pouvoir l'utiliser lors de leurs vacances à la campagne.

On voit ici une stratégie de marketing intéressante. Il semblerait que la politique des opérateurs de services de téléphonie mobile soit de s'adresser à une clientèle aussi large que possible, quitte éventuellement à sacrifier un peu la rentabilité du système.

Nous arrivons tout naturellement au concept d'équité. L'idée est qu'aucun utilisateur ne doit être pénalisé arbitrairement. Ainsi, dans un système parfaitement égalitaire, un berger du massif central aurait accès aux mêmes services qu'un businessman de la Défense. Évidemment, un tel système risquerait cependant d'être moins rentable.

On pourrait ajouter à celui-ci de nombreux autres exemples. D'une façon générale, plusieurs stratégies sont possibles. Soit l'on peut décider de se concentrer sur quelques utilisateurs, se trouvant ainsi pieds et poings lié à un petit nombre de clients, soit l'on peut tendre à toucher un grand nombre d'utilisateurs potentiels, quitte à sacrifier un peu la rentabilité d'un système.

L'idée sur laquelle repose ce travail est que tout opérateur de réseau se confronte au choix de l'équilibre entre rentabilité et équité.

Des critères d'équités ont été définis, dès les années 50. C'est tout naturellement dans le domaine de l'économie et des mathématiques qu'ils ont vu le jour. Évidemment, il n'existe pas de formulation universelle de l'équité, comme nous le verrons dans la suite. Mon but est de tenter de montrer comment ces critères d'équités peuvent être adaptés afin de proposer de nouveaux schémas d'allocations de ressources dans les réseaux.



# Chapitre 1

## Introduction

Comme déjà énoncé, le but de cette thèse est de mieux comprendre quels sont les différents concepts d'équité et comment les adapter aux problèmes d'allocations de ressources dans les réseaux.

L'intérêt est donc double. D'une part, il s'agit de voir comment ces critères d'équités (définis de façon très générale) peuvent être utilisés de façon à prendre en compte les particularités inhérentes aux réseaux de télécommunications, et en particulier répondre de façon naturelle aux applications temps-réels. D'autre part l'on désire proposer des méthodes de résolution centralisées et adaptées à chaque type de réseaux (terrestres, mobiles et satellitaires).

Ce premier chapitre a pour but d'introduire les notions nécessaires à la compréhension du manuscrit. Il présente comment nous proposons d'adapter et d'utiliser les critères d'équités déjà existants dans les réseaux de télécommunications. Nous commençons notre étude par montrer la nécessité d'instaurer des contrôles de ressources dans les télécommunications. Puis nous introduisons les notations et définitions utilisées tout au long de la thèse (Section 1.2), nous présentons alors les critères classiques déjà utilisés dans le cadre des réseaux (Section 1.3) ainsi que les algorithmes qui ont été développés pour les obtenir (Section 1.5). Nous présenterons ensuite un autre critère, né de la théorie des jeux non-coopératifs, appelé Nash Bargaining Solution (Section 1.4). Nous verrons que l'intérêt majeur de ce critère est l'introduction de fonctions d'utilités, qui représentent l'intérêt relatif ou la satisfaction qu'apporte l'allocation aux différents utilisateurs. Nous présenterons ensuite comment ces fonctions d'utilités peuvent être choisies dans le cas des réseaux de télécommunications (Section 1.6). Nous présenterons alors l'idée majeure de la thèse qui est l'utilisation de ces fonctions d'utilités dans une grande famille de critères d'équité et nous commenterons brièvement l'impact de ces fonctions d'utilité sur les allocations obtenues (Section 1.7). Nous présentons enfin une nouvelle méthode de résolution, la Programmation Semi-Définie Positive (SDP) (Section 4), qui est une extension de la programmation linéaire et que nous utiliserons dans les chapitres suivants pour résoudre le problème d'allocation équitable et optimale dans les réseaux.

## 1.1 Le contrôle de ressource

D'une façon générale, le problème d'allocation de ressource dans les réseaux est le suivant. On considère :

- un réseau, avec une certaine topologie, c'est-à-dire une architecture, et un ensemble de ressources,
- un ensemble d'utilisateurs avec des besoins éventuellement différents les uns des autres (on conçoit bien par exemple que dans le domaine de l'Internet, une personne consultant ses mails n'a pas les mêmes besoins qu'une autre utilisant une application vocale ou audio).

Le problème est alors de choisir une façon de partager la ressource du réseau entre les utilisateurs.

On peut très bien imaginer des réseaux dans lesquels les sources peuvent émettre en utilisant une quantité quelconque de ressource du réseau. Le but de la sous-section suivante est de montrer par des exemples simples que de tels réseaux peuvent être instables et contre-performants.

C'est alors pour éviter ce genre de situations qu'a été introduit le contrôle de congestion.

### 1.1.1 Pourquoi le contrôle de congestion est-il nécessaire ?

#### 1.1.1.1 Inefficacité d'un réseau

Je propose un petit exemple, inspiré du tutorial de Le Boudec [21]. Considérons le réseau filaire représenté en figure 1.1 dans lequel deux sources sont en compétition pour émettre. A chaque source  $S_i$  correspond un destinataire  $D_i$  (noté "Dest" sur le schéma). J'appelle "connection" une paire source-destination. Les deux connections du réseau sont alors  $(S_1, D_1)$  et  $(S_2, D_2)$ . Le média utilisé dans ce réseau est le câble. On suppose que les câbles ont des capacités différentes, comme indiqué sur la figure 1.1.

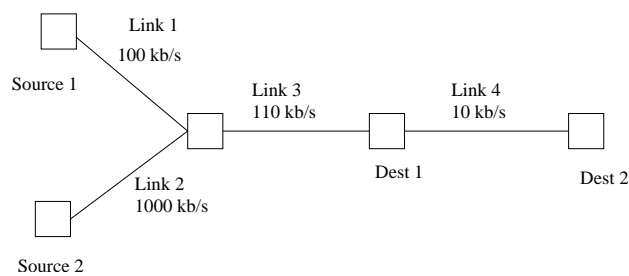


FIG. 1.1 – Petit réseau à 2 connexions.

A l'échelle des paquets, on peut se représenter chaque lien comme étant précédé par une file d'attente dans laquelle ils arrivent. Un ordonnanceur décide alors l'ordre dans lesquels ils vont être envoyés sur le réseau.

De façon imagée, on peut faire une analogie avec les fluides et considérer les données circulant sur un réseau comme un fluide. Les liens sont alors des canalisations et les files d'attentes, des réservoirs. Elle revient à dire dans notre cas que les paquets vont être envoyés sur le lien dans l'ordre d'arrivée dans le buffer, l'ordonnanceur étant la vanne entre le réservoir et le canal.

Une politique classique d'ordonnement est appelée FIFO (First In First Out). Dans la version imagée, la politique FIFO correspond à une vanne située dans le fond du réservoir (en considérant le fluide homogène). Au niveau macroscopique, dans une politique FIFO, si le trafic arrivant au niveau d'un lien est supérieur à sa capacité, le débit sortant de chaque connexion sera proportionnel à son débit d'arrivée.

Dans notre petit réseau (figure 1.1), supposons que chacune des sources émette au débit maximum autorisé par son lien d'accès.  $S1$  va donc utiliser un débit de  $100\text{ kb/s}$  et  $S2$  aura un débit de  $1000\text{ kb/s}$ , ce qui est inutile car le dernier lien traversé par la connexion est le lien 4, d'une capacité de  $10\text{ kb/s}$ .

La capacité du lien 3 est inférieure à la somme des débits des sources donc le lien sera saturé et les allocations des connexions  $(S1, D1)$  et  $(S2, D2)$  seront respectivement  $10\text{ kb/s}$  et  $100\text{ kb/s}$ .

De même, le lien 4 est saturé car sa capacité est inférieure à  $100\text{ kb/s}$  donc l'allocation à la connexion  $(S2, D2)$  sur le lien 4 sera de  $10\text{ kb/s}$ .

Ainsi, les allocations globales sont de  $10\text{ kb/s}$  pour chacune des deux connexions.

On voit donc dans ce petit exemple que  $(S2, D2)$  pénalise inutilement la connexion  $(S1, D1)$ , qui se voit allouée un débit global de  $10\text{ kb/s}$  alors que le réseau lui permettait d'avoir une allocation dix fois plus grande.

On voit dans cet exemple que la source de l'inefficacité de ce réseau est due à la présence sur le réseau de paquets qui n'arriveront jamais à destination car ils seront refusés par des buffers "en cours de route". Ces paquets empêchent d'autres paquets de circuler, qui eux, pourraient être menés à destination.

### 1.1.1.2 Effondrement de congestion

La situation peut même être bien plus catastrophique et mener à un effondrement de congestion.

**Définition 1.1** *L'effondrement de congestion est une situation où une augmentation de la charge du réseau provoque une diminution de la charge utile écoulée.*

Ces situations sont catastrophiques car en général, une source qui se voit attribuer un débit anormalement faible comprend que le réseau est congestionné. Une stratégie classique est alors d'augmenter sa demande afin d'augmenter son "poids" relatif par rapport aux autres et donc de voir son allocation croître. Or si le réseau est en situation d'effondrement, c'est la situation inverse qui se produit, et bien que la source cherche à augmenter son débit, elle augmente la congestion globale du système et voit sa propre allocation diminuer. On entre alors dans une situation en spirale. L'exemple suivant propose un petit réseau pouvant mener à une telle situation.

Considérons un réseau circulaire comme présenté à la figure 1.2. Il est composé d'un nombre de liens  $N$  quelconque ( $N > 2$ ). Il est traversé par  $N$  connexions, chaque connexion  $i$  utilisant les liens  $i$  et  $i + 1 \pmod{N}$ . Supposons que chaque source émette avec un même débit  $\lambda$ ,  $\lambda \geq 0$ . On suppose que la politique de chacun des liens est la même que celle de l'exemple précédent. De plus, tous les liens ont une capacité identique, que je note  $c$ .



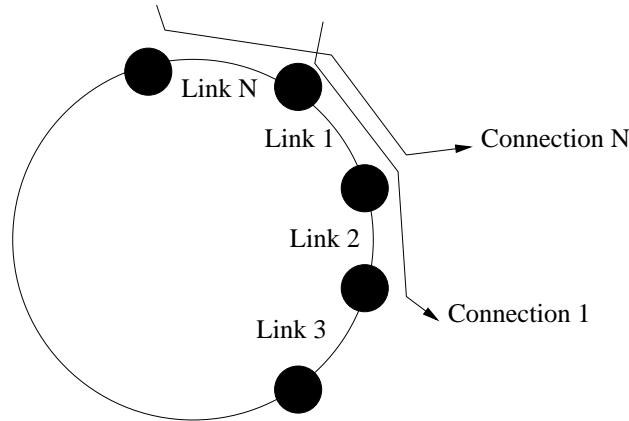


FIG. 1.2 – Réseau cyclique.

Alors, on peut alors montrer [21] que pour des grandes valeurs de  $\lambda$ , l'allocation de la connexion  $i$  au niveau du deuxième lien qu'elle traverse est :

$$x_i^2 = \frac{c^2}{\lambda} + o\left(\frac{1}{\lambda}\right)$$

Ainsi, nous sommes en présence d'une situation paradoxale dans laquelle plus les demandes des connexions augmentent (i.e  $\lambda$  croit) et plus le débit que chaque connexion obtient est faible. Le réseau est alors de moins en moins utilisé et les performances sont dégradées pour tous les utilisateurs. C'est l'effondrement de congestion.

Or, on remarque que si le réseau était bien utilisé, chaque source pourrait se voir allouer la moitié de la capacité de chaque lien.

### 1.1.2 Conclusion

D'autres situations menant à des situations d'effondrement sont présentées dans [38]. Dans les premiers temps de l'Internet, il n'y avait pas de contrôle de congestion. C'est tout logiquement qu'un effondrement a alors été rencontré dans les années 80 ([87]), essentiellement dû aux connexions qui retransmettaient inutilement des paquets qui étaient en transit ou avait été déjà reçus par leurs destinataires.

Les contrôles de congestion ont pour but d'éviter ce genre de scénario catastrophe en limitant aux sources leur taux de transmission en fonction de l'état du réseau. Les solutions classiques sont les protocoles de "bout-en-bout" qui ont une vue globale de la connexion, ou ceux interdisant l'accès au réseau à des paquets dont on sait qu'ils ne pourront être acheminés à destination.

Aujourd'hui, les grands réseaux déployés (reposants sur ATM, TCP) comprennent un mécanisme de contrôle de congestion. Il garantit l'équilibre et l'utilisation efficace des ressources du réseau. Les mécanismes de contrôle de congestion sont différents d'un protocole à l'autre, ce qui induit donc des différences dans les allocations rencontrées. Nous verrons dans la suite du chapitre qu'ainsi par exemple, le protocole ATM alloue les ressources de façon plus égalitaire ou équitable que TCP. Notons que l'introduction de notions d'équité dans les réseaux est prise de plus en plus en compte comme en témoignent les propositions

de directions d'investigations en matière de mécanismes futurs de contrôle de congestion de la RFC 2914 ([38]).

## 1.2 Définitions générales

Nous évoquons ici les définitions générales nécessaires à la compréhension de l'exposé.

Pour un réseau et un ensemble de connexions donnés, il existe une infinité d'allocations possibles. Les définitions qui suivent ont pour but de permettre de les particulariser et les comparer.

### 1.2.1 Ressource

La ou les ressources qui sont partagées entre les utilisateurs peuvent être de natures différentes. Je vais tâcher d'utiliser le terme de "ressource" dans toute cette partie afin de garder la généralité des formulations d'équité. Il suffit ensuite de remplacer le mot "ressource" par "bande passante" et l'on obtient la formulation pour les réseaux terrestres. Dans le cas des réseaux mobiles où la ressource partagée est un canal radio, on affectera aux utilisateurs un "débit (ou taux) de transmission" et une puissance. Enfin, dans les réseaux satellitaires que nous étudierons, ce débit sera discrétisé et on parlera alors d'allocation de porteuses. Nous y reviendrons dans les chapitres consacrés.

### 1.2.2 Connexion

On distingue deux types de connexions : *unicast* et *multicast*. Dans les premières, un émetteur envoie un certain nombre de paquets à un destinataire unique. Dans le second, les récepteurs sont multiples. Les connexions multicast concernent beaucoup de situations multimedia. C'est par exemple le cas des émissions de télévision (lorsque le nombre de récepteurs est grand, on parle alors de "broadcast"). Il s'agit souvent de diffusion de documents audio (vidéo) sur Internet, lorsque chaque utilisateur se connecte à un serveur et peut écouter (voir) un document sans le télécharger sur sa machine.

Dans notre contexte une *connexion* désignera toujours une paire constituée d'une source et d'une destination. Dans le cas du multicast on appellera *session* l'ensemble des connexions impliquées.

### 1.2.3 Allocation faisable

C'est une lapalissade mais il est utile de préciser qu'une allocation est dite *faisable* si elle satisfait les contraintes du problème. Ces dernières sont de types divers selon la nature du réseau considéré. C'est pourquoi elles seront précisées dans chaque chapitre.

D'une façon générale, on distingue :

- Les contraintes du réseau : la capacité totale des liens dans le cas des réseaux filaires, des contraintes sur le plan de fréquence dans les réseaux satellite, etc.
- Les contraintes des terminaux : par exemple, dans les réseaux mobiles, chaque mobile voit sa puissance d'émission limitée, dans les réseaux satellitaires, chaque terminal ne peut utiliser qu'un nombre fini de type de porteuses.

- Les contraintes de l'utilisateur : comme nous le verrons dans la suite, à chaque utilisateur est associé une demande minimale et maximale.

Nous ne nous intéressons dans la suite de la thèse qu'aux allocations faisables.

## 1.2.4 Ordre parmi les allocations

Nous définissons ici des relations d'ordre entre les allocations afin de nous permettre dans la suite de les comparer, en terme d'équité et d'efficacité notamment. Soit  $x$  une allocation. C'est en fait un vecteur qui associe à chaque connexion une quantité de ressource. On suppose qu'il existe une relation d'ordre sur les éléments de  $x$ . On peut définir plusieurs relations d'ordre sur les allocations ([17]) et montrer aisément qu'elles sont réflexives, symétriques et transitives. On note alors :

$$\begin{aligned} x \leq y &\Leftrightarrow \forall i \in 1, \dots, N, x_i \leq y_i \\ x \leq_{lg} y &\Leftrightarrow \text{Soit } x = y \text{ soit } \exists i, i > 0, \forall j \in 1, \dots, i - 1, x_j = y_j \text{ et } x_i < y_i \text{ (c'est l'ordre lexicographique)} \\ x \preceq y &\Leftrightarrow \sigma(x) \leq_{lg} \sigma(y) \text{ avec } \sigma(x) \text{ et } \sigma(y) \text{ les versions ordonnées de } x \text{ et de } y^1. \end{aligned}$$

## 1.2.5 Efficacité

L'efficacité définit l'utilisation des ressources du réseau. On dit ainsi qu'une allocation est efficace au sens de Pareto s'il n'est pas possible d'augmenter la quantité de ressource allouée à une connexion sans diminuer l'allocation d'une autre connexion. Dans le cadre des réseaux, cela revient à dire que chaque connexion traverse au moins un lien qui est saturé.

D'une façon plus formelle, une allocation est optimale au sens de Pareto si elle est maximale au sens de la relation d'ordre  $\leq$  de la sous-section 1.2.4.

En général, si  $N$  est le nombre de connexions, l'ensemble des points optimaux au sens de Pareto forme un sous ensemble de dimension  $N - 1$ . On peut comprendre qu'un opérateur veuille profiter d'une utilisation optimale du réseau. Le problème auquel nous nous intéressons aujourd'hui est de savoir quelle allocation choisir dans cet hyper-espace.

## 1.2.6 Notations

Dans la suite de l'exposé nous utilisons une notation cohérente résumée dans le tableau ci-dessous.

$n$	une connexion
$\mathcal{N}$	l'ensemble des connexions
$x_n$	l'allocation de la connexion $n$ .
$\vec{x}$	une allocation (faisable)
$X$	l'ensemble des allocations (faisables)
$l$	un lien
$L$	l'ensemble des liens

Ces notations étant explicitées, nous présentons dans la section suivante les allocations classiques rencontrées dans les réseaux. Nous présentons pour chacune un exemple simple

<sup>1</sup> $\sigma(x)$  est la version ordonnée de  $x$  si pour tout  $i$ ,  $\sigma(x)_i$  est le  $i^{eme}$  plus petit élément de  $x$

tiré d'une situation de réseau terrestre. Cependant, les critères ne sont bien évidemment pas limités à ces types de réseaux.

## 1.3 Allocations classiques

Nous décrivons dans cette section les algorithmes de partage de ressources dans les réseaux rencontrés fréquemment dans la littérature. Une brève description de ces allocations se trouve dans [78].

Chacun d'eux se présente sous une forme simple et une version pondérée. Cette dernière a été introduite pour permettre aux utilisateurs d'exprimer la valeur relative de leur trafic. Cette valeur est souvent liée à une idée de tarification. L'idée générale des poids est la suivante. Considérons un réseau à  $N$  connexions. Supposons que la connexion  $n$  ait un poids de  $k$ ,  $k \in N, k > 1$ . Nous la remplaçons alors par  $k$  connexions de poids unitaire utilisant les mêmes ressources que la connexion  $n$ . Nous les notons  $n_1, \dots, n_k$ . Alors,  $x_n = \sum_k x_{n_k}$ . Ainsi, une connexion de poids  $k$  est équivalente à  $k$  connexions de poids 1.

En revanche, si l'on remplace la connexion  $n$  par une connexion  $n'$  de poids unitaire, alors nous n'avons pas nécessairement  $x_n = k.x_{n'}$ . Ainsi, il n'existe pas de fonction générale montrant l'influence du poids d'une connexion sur l'allocation : l'utilisation des poids ne peut permettre de garantir l'accès à un certain pourcentage de ressource.

Je vais illustrer les allocations présentées dans la suite sur l'exemple d'un réseau linéaire (Fig 1.3). Dans la représentation adoptée ici, chaque carré représente un lien du réseau, et les connexions sont représentées par des segments de droites traversant ces carrés. Les numéros correspondent à ceux des connexions.

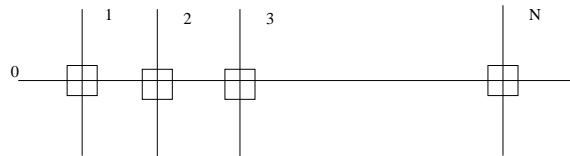


FIG. 1.3 – Réseau linéaire

Le réseau se compose de  $N$  liens (numérotés de 1 à  $n$ ) et sur lequel circulent  $N + 1$  connexions (numérotées de 0 à  $n$ ). On numérote les connexions comme suit :

- chaque connexion  $i$ ,  $i \neq 0$  utilise uniquement le lien  $i$
- la connexion 0 utilise tous les liens.

### 1.3.1 Maximisation du débit total

#### 1.3.1.1 Définition

Ce critère, le plus élémentaire, consiste en la maximisation de la somme des bandes passantes allouées :  $\max \sum_{n \in \mathcal{N}} x_n$ .

Il apparaît sous différents noms dans la littérature et a été très largement déployé. En effet, dans le cas où la tarification se fait proportionnellement au débit de la connexion par exemple, cela revient à maximiser le revenu du réseau.

Cependant, un gros inconvénient de ce critère est qu'il peut mener à des allocations où plusieurs connexions se voient refuser l'accès au réseau, comme le montre l'exemple suivant. C'est pourquoi il n'est pas considéré comme équitable.

### 1.3.1.2 Exemple : le réseau linéaire

Les contraintes du système sont les contraintes de capacité :

$$\forall l \in [1; N], \quad x_0 + x_l \leq c_l \quad (\text{en notant } C_l \text{ la capacité du lien } l)$$

Le revenu global du système s'écrit alors :

$$\sum_{0 \leq n \leq N} x_n \leq x_0 + \sum_{1 \leq n \leq N} (C_n - x_0) = (1 - N)x_0 + \sum_{1 \leq n \leq N} C_n. \quad (1.1)$$

En fait, on voit aisément que pour cet exemple, on peut (en l'absence de contraintes supplémentaires) saturer le lien  $i$  et obtenir :  $\forall l \in [1; N], \quad x_0 + x_l = c_l$ .

Le revenu total est ainsi :

$$(1 - N) * x_0 + \sum_{l=1}^N c_l$$

Ainsi, dès lors que  $N > 1$ , on voit que la maximisation du revenu total implique que  $x_0$  soit minimal. Ainsi  $x_0 = 0$ .

On voit donc dans cet exemple très simple que le critère de maximisation du débit total peut aboutir à des allocations dans lesquelles une ou plusieurs connexions se voient refuser l'accès au réseau.

Une variante courante est illustrée dans la littérature. On associe à chaque utilisateur une fonction d'utilité (nous verrons à la section 1.6 comment cela se justifie dans le cas des réseaux). De façon informelle, la fonction d'utilité représente la satisfaction que l'utilisateur tire de son allocation. Ce peut être par exemple la qualité perçue d'un signal audio ou video. Les algorithmes proposés visent alors la maximisation de la somme des utilités des utilisateurs. Par exemple, [56] propose un tel algorithme.

Un exemple d'extension possible pour l'allocation à des sessions multicast est donné dans [57]. Leur algorithme est "scalable" grâce à l'architecture en arbre qui est utilisée. Ainsi, chaque source est le sommet d'un arbre dont les destinataires sont les feuilles. Cependant, cette architecture en arbre nécessite qu'un algorithme tourne au niveau de chaque source, chaque destination et au niveau de chaque embranchement de l'arbre.

En outre, on peut noter que contrairement à toutes les allocations que nous allons présenter dans la suite, l'optimisation globale est la seule allocation qui n'est pas définie de manière unique, comme on peut le remarquer dans l'exemple de la figure 1.3 ([21]).

A l'opposé de ce critère de maximisation globale, on trouve le critère d'équité max-min, qui vise à donner aux connexions les allocations les plus proches possibles les unes des autres tout en restant optimal au sens de Pareto. L'allocation globale au niveau du système est évidemment inférieure à celle que l'on vient d'étudier. On voit donc dès à présent que la recherche d'une allocation est bien souvent un *compromis entre équité et optimalité*.

## 1.3.2 Équité Max-Min

### 1.3.2.1 Définitions

Ce critère d'équité peut être défini de plusieurs façons équivalentes que je décris brièvement ci-dessous.

La preuve de l'unicité de l'allocation max-min lorsque le nombre de liens et de connexions est fini est disponible dans [21].

#### Définition intuitive

Intuitivement, ce critère est le plus équitable possible dans la mesure où c'est celui qui donne naissance aux allocations dont les valeurs sont les plus proches possibles les unes des autres.

Considérons un système ne comprenant qu'un seul lien saturé. Une définition de l'équité max-min est la suivante. Soit  $l$  le lien congestionné,  $l_c$  sa capacité,  $n_l$  l'ensemble des connexions qui traversent ce lien et  $|n_l|$  son cardinal (c'est-à-dire le nombre de connexions traversant ce lien). Alors l'équité max-min va allouer à chaque connexion de  $n_l$  une bande passante de  $l_c/|n_l|$ . (dans la version pondérée, si l'on note  $w_i$  le poids d'une connexion  $i$ , alors l'allocation au lien congestionné sera de  $c_l w(i) / \sum_{j \in n_l} w(j)$ ).

On peut ensuite généraliser ce résultat au cas d'un nombre quelconque de liens congestionnés. La bande passante allouée à une connexion sera le minimum des allocations possibles à chaque lien. Si une ou plusieurs connexions ne peuvent utiliser la fraction de ressource auquel elles peuvent prétendre, c'est la quantité de ressource restante qui sera allouée de façon égale entre les autres. Pour clarifier cela, je considère l'exemple du petit réseau 1.4.

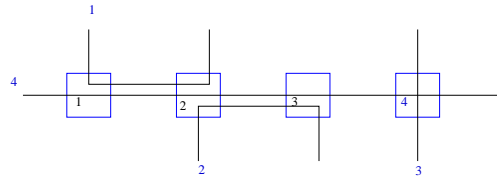


FIG. 1.4 – Petit réseau

Supposons que les capacités des 4 liens soient égales à  $C$ . La connexion 1 peut prétendre à une bande passante de  $C/2$  au niveau du premier, troisième et quatrième liens, et de  $C/3$  au niveau du lien 2. Son allocation sera donc  $\min(C/2, C/3) = C/3$  (car les deux connexions 4 et 2 peuvent utiliser également une bande passante de  $C/3$ ).

Les allocations des autres connexions seront alors  $x_1 = x_2 = C/3$  et  $x_3 = 2C/3$ . On voit bien sur ce petit exemple, que l'allocation est la plus "égale" possible tout en tenant compte des contraintes du système. Ce qui donne un revenu de  $5C/3$ .

L'optimisation globale aurait donné :

Connexion	Allocation
1	$C - x_2$
2	$x_2$
3	$C$
4	$0$

D'où un revenu total de  $2C$ . On a ici un exemple où l'optimisation globale n'est pas unique.

Une caractérisation de l'allocation max-min équitable découle presque immédiatement de cette définition. Commençons par définir le goulot d'étranglement d'une connexion.

**Définition 1.2 (goulot d'étranglement d'une connexion)** *On dit qu'un lien  $l$  est un goulot d'étranglement pour une connexion  $n$  ssi :*

- le lien est saturé
- la connexion  $n$  a l'allocation maximale parmi toutes les connexions utilisant le lien  $l$ .

On a alors la proposition suivante (la preuve est donnée dans [21]) :

**Proposition 1.1 (caractérisation de l'allocation max-min équitable)** *Une allocation faisable est max-min équitable si et seulement si chaque source a un goulot d'étranglement.*

Dans le petit exemple précédent, le lien 2 est goulot d'étranglement des connexions 1, 2 et 4. Le lien 4 est le goulot d'étranglement de la connexion 3.

On peut voir que l'allocation d'optimisation globale n'est pas max-min équitable car la connexion 4 n'a pas de goulot d'étranglement.

On peut enfin remarquer un fait intéressant. Dans ce petit réseau, contrairement au réseau linéaire, il n'est pas possible de saturer les liens. On peut comprendre le critère d'optimalité de Pareto comme étant : une allocation optimale saturera tous les liens qu'il est possible de saturer. Dans notre petit exemple il est évident qu'il n'est pas possible de saturer les liens 1, 2 et 3 simultanément.

### Définition itérative

L'allocation max-min est celle qui est la plus favorable au plus petit bénéficiaire. Je la note  $\lambda$ . C'est un vecteur de taille  $\mathcal{N}$ . J'appelle plus petit bénéficiaire la connexion dont l'allocation est la plus faible. On voit ainsi que dans le petit exemple de la figure 1.4, pour l'allocation max-min les plus petit bénéficiaires (ils sont ex-aequo) sont les connexions 1, 2 et 4. Leur allocation est de  $C/3$ . Dans la version d'optimisation globale, le plus petit bénéficiaire est la connexion 4 avec une allocation de 0.

Cette propriété se réitère. Ainsi, si l'on enlève des vecteurs d'allocations l'élément correspondant, alors le vecteur  $\lambda'$ , de dimension  $\mathcal{N} - 1$  est également max-min dans l'ensemble des allocations de taille  $\mathcal{N} - 1$  (associées au mêmes contraintes évidemment), c'est-à-dire que l'allocation du plus petit bénéficiaire est maximale dans l'ensemble des allocations faisables.

On voit également avec cette méthode de construction que l'allocation obtenue est ainsi la plus équitable possible.

De façon plus formelle, on peut reprendre les relations d'ordre entre les allocations de la sous section 1.2.4. L'allocation max-min est maximale pour la relation d'ordre  $\preceq$ , c'est-à-dire que si  $A$  est l'allocation max-min ordonnée, alors pour toute autre allocation faisable,  $B$ , on a  $B \preceq A$ .

### Définition courante

On peut définir de manière équivalente l'équité max-min comme étant telle que toute augmentation d'allocation pour une connexion  $i$  se fasse au détriment de celle d'une connexion  $j$  qui était déjà défavorisée par rapport à la première connexion. ( $x_j \leq x_i$ ).

De façon plus formelle, si  $x$  est une allocation max-min et  $y$  une autre allocation alors  $y_n > x_n \Rightarrow \exists s, x_s \leq x_n, y_s < x_s$ .

### 1.3.2.2 Exemple du réseau linéaire

Revenons maintenant à l'exemple du réseau linéaire (Fig. 1.3). Comme précédemment, les liens sont saturés et, pour un  $x_0$  donné, on obtient :

$$\forall n \in [0; N], x_n = C_n - x_0.$$

Soit maintenant  $l_{\min}$  l'indice du lien de plus petite capacité. Pour un  $x_0$  donné, c'est donc l'indice du plus petit  $x_l$  pour  $l \in [1; N]$ . L'équité max-min implique alors que  $x_0 = x_{l_{\min}} = C_{l_{\min}} - x_0$ .

Ainsi,  $x_0 = C_{l_{\min}}/2$ .

Le revenu total est alors de  $(1 - \frac{N}{2})C_{l_{\min}} + \sum_{n \in \mathcal{N}} C_n$ .

### 1.3.2.3 Algorithme de construction de l'allocation max-min

L'équité Max-min peut être obtenue par la procédure de remplissage suivante ([14, 21]) :

1. A l'instant 0, toutes les connexions ont une bande passante allouée nulle.
2. On augmente l'allocation de la même façon pour tous (même taux d'augmentation)
3. Quand, à un instant donné, la limite de la capacité est atteinte pour un lien, on gèle l'allocation de bande passante pour toutes les connexions passant par ce lien, mais on continue la procédure de remplissage pour les connexions n'utilisant que des liens non saturés.

On peut aisément se convaincre que cette procédure de remplissage tend vers l'allocation max-min. Cependant, un problème se pose quand à sa mise en pratique. Supposons qu'une nouvelle connexion arrive dans le réseau, il ne suffit pas de pouvoir informer toutes les sources qui utilisent les mêmes liens de réinitialiser leur allocation à 0 car alors elles mêmes seront pénalisées par rapport aux autres connexions qui utilisent leurs liens.

Je dis que deux connexions  $A$  et  $B$  sont *liées directement* et je note  $A \leftrightarrow B$  si elles ont au moins un lien en commun. La relation  $\leftrightarrow$  est évidemment symétrique et réflexive. En revanche, elle n'est pas transitive. Dans l'exemple de la figure 1.5, la connexion 4 est liée uniquement directement à la connexion 1. Je dis maintenant que deux connexions  $A$  et  $B$  sont *liées indirectement* et je note  $A \rightleftharpoons B$  s'il existe une suite de connexions sur le réseau  $A_i, 1 \leq i \leq k$  telle que :

- $A \leftrightarrow A_1$
- $\forall i \in [1, k - 1], A_i \leftrightarrow A_{i+1}$
- $A_k \leftrightarrow B$ .

On peut remarquer qu'en revanche, la relation  $\rightleftharpoons$  est réflexive, symétrique et transitive. On peut alors montrer que lorsqu'une nouvelle connexion  $A$  arrive dans le réseau, il faut recommencer l'algorithme de remplissage de toutes les connexions  $B$  telles que  $A \rightleftharpoons B$



Prenons un petit exemple pour clarifier les choses. Dans la figure 1.5, je suppose que la connexion 4 (en bleu) se joint au réseau. Elle partage avec la connexion 1 un lien en commun, donc je dis qu'elles sont connectées directement, il est évident qu'il va falloir recommencer la procédure de remplissage pour la connexion 1. Mais on voit également que la connexion 1 possède un lien en commun avec la connexion 2. Donc les connexions 4 et 2 sont connectées de façon indirecte. Ainsi, lors de l'arrivée de la connexion 4, les connexions 1 et 2 devront recommencer la procédure de remplissage. On comprend aisément que dans un réseau réel, il n'est pas aisé en général de connaître l'ensemble des connexions liées, ce qui limite malheureusement la portée de cet algorithme.

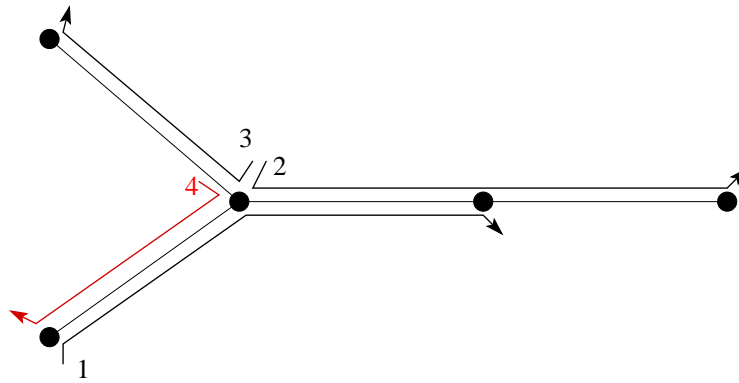


FIG. 1.5 – Un autre petit exemple

En revanche, cette algorithme est facilement adaptable à la version pondérée de l'allocation max-min. On modifie alors l'étape 2 de l'algorithme pour que l'augmentation de chaque connexion soit proportionnelle à son poids.

#### 1.3.2.4 Implémentation et discussion

L'équité max-min a été adoptée par l'ATM forum pour le service Available Bit Rate (ABR) de l'ATM ([1]). Dans les recommandations pour l'ABR, différents critères sont proposés : l'équité max-min, l'équité max-min couplée à une demande minimale, l'équité max-min pondérée. Dans l'équité max-min couplée à une demande minimale, chaque connexion est associée à une demande minimale. L'allocation du surplus de bande passante disponible est alors attribuée de façon max-min équitable. Dans l'ABR, chaque source est informée de la bande passante qui va lui être attribuée dans une phase de négociation avant le début de la transmission.

[104] présente une étude intéressante dans le cas du multicast. En effet, adapter les définition d'équité max-min n'est pas trivial. En multicast, deux scénarios sont possibles. Soit l'on alloue un même débit à tous les receveurs d'une même connexion, soit l'on permet d'avoir des débits différents selon les receveurs, c'est le multicast en couches. Il y a alors autant de débits différents que de couches.

La première version de multicast est évidemment la plus facile à mettre en œuvre, mais la moins performante dans la mesure où un utilisateur disposant de moins de bande passante va pénaliser l'ensemble des receveurs. On augmente ainsi de plus le nombre de liens non saturés. La nouvelle définition de l'équité max-min est alors la même que précédemment

en les étendant du niveau de "connexion" à celui de "session"<sup>2</sup> Nous n'avons plus alors la propriété que chaque connexion possède un goulot d'étranglement. Cette propriété n'est alors plus vraie qu'au niveau d'une session. Les auteurs définissent une mesure d'équité. Leur analyse confirme le fait qu'en présence de multicast sans couche, le niveau d'équité est plus faible et permet de mesurer cet écart.

Dans le cas où chaque connexion à l'intérieur d'une session multicast peut disposer de sa propre allocation, la définition "classique" d'équité peut être utilisée.

Cependant dans le cas du multicast, des résultats surprenant peuvent se produire. En effet, si un receveur de l'une des sessions se retire, l'allocation des autres receveurs de la session considérée comme des autres sessions peut augmenter ou diminuer selon la configuration du réseau comme le montre l'exemple de la figure 3 de [104].

De plus, de gros problèmes se posent dans la pratique. Dans les implémentations acutelles, chaque source transmet à un ensemble de débits différents, appelés couches. Les receveurs peuvent alors se connecter à un certain nombre de couches pour obtenir un débit voulu. Si les receveurs étaient tenus de rester toujours au niveau de la même couche, l'équité ne pourrait être obtenue. A la place, les receveurs changent de couche afin qu'à long terme l'allocation soit max-min équitable. Cependant, ces changements font que les liens ne sont plus bien utilisés à cause des délais de coordination entre la source et les receveurs.

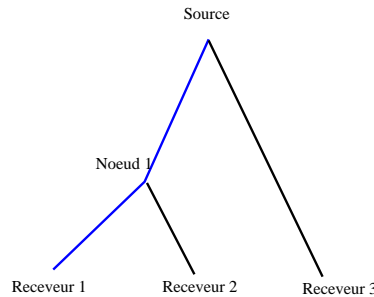


FIG. 1.6 – Exemple de réseau multicast en couche

Le problème survient lorsqu'un utilisateur se déconnecte. Supposons que le receveur 1 se déconnecte. Le problème se pose au niveau du lien entre la source et le nœud 1. En effet, la bande passante qui lui était allouée était  $\max(x_1, x_2)$  (en notant  $x_i$  l'allocation associée au receveur  $i$ ). Notons au passage, que dans le cas unicast, cette allocation aurait été de  $x_1 + x_2$ . Le problème vient donc lorsque l'on avait  $x_1 > x_2$ . Alors, l'allocation du (ou des) lien(s) entre la source et le nœud 1 est trop importante. Le système est sous-optimal, et sur une certaine durée, le taux d'utilisation du lien est supérieur au débit maximum utilisé sur tous les receveurs. C'est ce que les auteurs appellent la redondance. Ils étudient alors l'impact de la redondance sur l'équité. Ils montrent également que l'augmentation du nombre de couches augmente l'équité obtenue. L'article s'intéresse uniquement au cas de l'équité max-min, mais les résultats seraient assez clairement étendables aux autres utilités.

Bien qu'il soit Pareto optimal, il a été reproché à l'équité max-min de favoriser trop

<sup>2</sup>Je rappelle qu'une session multicast regroupe l'ensemble des connexions entre la source et les multiples récepteurs.

les longues connexions et de ne pas utiliser pas de façon assez efficace la bande passante. Pour s'en convaincre, revenons à l'exemple du réseau linéaire. Supposons que tous les liens aient la même bande passante  $C$ . La connexion 0 est très coûteuse car elle traverse tous les liens et pénalise donc beaucoup de "petites" connexions. Or elle reçoit, d'après l'allocation max-min, la moitié de la bande passante disponible, soit autant qu'une "petite" connexion. Une allocation intermédiaire entre l'optimisation globale et l'équité max-min a donc été introduite, c'est l'équité proportionnelle.

### 1.3.3 Equité proportionnelle

#### 1.3.3.1 Définition

L'équité proportionnelle peut être définie de deux façons équivalentes. On trouve la preuve de l'équivalence de ces deux définitions dans [21].

##### Première définition

Ce critère peut être exprimé comme un problème d'optimisation :

$$\max_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \ln(x_n).$$

On peut montrer que dans le cas d'un nombre fini de connexions et de liens, ce critère mène à une solution unique ([21]).

##### Deuxième définition

On peut également considérer l'allocation équitable proportionnellement comme l'unique allocation  $\vec{x}$  telle que, pour toute autre allocation  $\vec{x}'$ , on obtienne :

$$\sum_{n \in \mathcal{N}} \frac{x'_n - x_n}{x_n} \leq 0$$

#### 1.3.3.2 Exemple : le réseau linéaire

Utilisons la première définition de l'équité proportionnelle. Comme dans les cas précédents, on peut saturer les liens et le critère s'écrit (ceci est dû au fait que le critère est Pareto optimal) :

$$\max \left( \ln(x_0) + \sum_{n=1}^N \ln(c_n - x_0) \right).$$

On obtient immédiatement en dérivant la formule :

$$x_0 = 1 / \sum_{n \in \mathcal{N}} \frac{1}{c_n - x_0}.$$

Dans le cas de  $N$  liens identiques, on obtient donc :  $x_0 = \frac{c}{N+1}$ .

On remarque alors un résultat attendu :

$$x_{0_{\text{opt globale}}} < x_{0_{\text{éq prop}}} < x_{0_{\text{max-min}}}.$$

L'équité proportionnelle permet bien d'obtenir une allocation pour  $x_0$  qui soit strictement inférieure aux autres connexions. De plus,  $x_0 \rightarrow_{L \rightarrow \infty} 0$ .

On peut vérifier par le calcul que ni l'allocation max-min équitable, ni l'allocation d'optimisation globale ne sont proportionnellement équitables. En effet, supposons  $n > 1$  (on considère que l'on a au moins 2 liens, et donc 3 connexions). L'allocation max-min ne peut être proportionnellement équitable, car si  $\vec{x}'$  est l'allocation d'optimisation globale alors :

$$\sum_{n \in \mathcal{N}} \frac{x'_n - x_n}{x_n} = \frac{0 - C/2}{C/2} + n * \left( \frac{C - C/2}{C/2} \right) = n - 1 > 0.$$

Quant à elle, l'allocation d'optimisation globale ne peut être proportionnellement équitable car elle refuse l'accès au réseau à la connexion 0.

On remarque alors que le revenu total du réseau vaut, pour l'allocation équitable proportionnelle :  $R_{prop} = (1 + N^2) \frac{C}{N + 1}$ . On remarque que  $R_{\text{opt globale}} \geq R_{\text{éq prop}} \geq R_{\text{max-min}}$ . Ce résultat se généralise : l'allocation proportionnelle équitable offre une meilleure utilisation de la bande passante que l'allocation max-min. Ce compromis entre l'équité et l'optimalité lui a valu son succès.

### 1.3.3.3 Implémentation et discussion

Le concept d'équité proportionnelle a été introduit par Kelly. En offrant davantage de bande passante aux "petites" connexions que l'équité max-min, il donne naissance à une meilleure utilisation de la bande passante.

Il existe également une version pondérée de l'équité proportionnellement. Une allocation est dite équitable proportionnellement pondérée et notée  $\vec{x} \in X^{wpf}$ , si pour toute autre allocation  $\vec{x}^* \in X$ , la somme des changements proportionnels pondérée par la famille  $w_i$  est négative ou nulle ([60, 61] (voir également [17, 77, 85])).

$$\sum_{n \in \mathcal{N}} w_n \frac{x_n^* - x_n}{x_n} \leq 0.$$

Nous avons la définition équivalente de la solution  $\overrightarrow{x^{wpf}}$  :

$$x^{wpf} = \max_{\vec{x} \in X} \sum_{x \in \mathcal{N}} w_n \ln(x_n) \tag{1.2}$$

$$\text{ou } x^{wpf} = \max_{\vec{x} \in X} \prod_{n \in \mathcal{N}} x_n^{w_n}. \tag{1.3}$$

Les mécanismes de contrôle de congestion basés sur des augmentations linéaires et des décroissances multiplicatives (Linear Increase Multiplicative Decrease) donnent lieu à des allocations proportionnellement équitables sous certaines conditions ([61]). Malheureusement, ce résultat n'est pas applicable au cas de toutes les versions de TCP, et en particulier de la variante déployée actuellement de TCP Reno. Nous y reviendrons dans la section 1.5.

### 1.3.4 Minimisation du délai potentiel

#### 1.3.4.1 Définition

La définition de ce critère, introduit dans [78] est :

$$\min_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \frac{1}{x_n}$$

Alors que les critères précédents visaient la maximisation d'une fonction des bandes passantes des connexions, on minimise ici une fonction de leurs inverses, c'est-à-dire une fonction des temps de transfert des connexions, d'où le nom de ce critère d'équité.

#### 1.3.4.2 Application au réseau linéaire

Comme dans les cas précédents, les liens sont saturés et on a :  $\forall n \in [0; N], x_n = C_n - x_0$ . Le critère s'écrit alors :

$$\min \frac{1}{x_0} + \sum_{i \in \mathcal{N}} \frac{1}{c_i - x_0}.$$

On ne peut résoudre analytiquement ce problème sans hypothèses supplémentaires. Supposons alors que tous les liens soient identiques (de même capacité). Il vient alors en dérivant  $\frac{-1}{x_0^2} + \frac{N}{(c - x_0)^2} = 0$ , d'où finalement :

$$x_0 = \frac{c}{\sqrt{N} + 1}.$$

### 1.3.5 Liens entre les équités

#### 1.3.5.1 Définition

Récemment, Mo et Walrand ont montrés ([85]) que tous ces critères pouvaient être écrits sous la forme d'une seule fonction d'optimisation. Considérons l'allocation  $\vec{x}_\alpha$  solution de :

$$\vec{x}_\alpha = \max_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \frac{x_n^\alpha}{1 - \alpha} \quad (1.4)$$

Avec  $\alpha \geq 0, \alpha \neq 1$ . Alors, pour des valeurs particulières de  $\alpha$ ,  $\vec{x}_\alpha$  correspond à l'allocation :

- Pour  $\alpha = 0$  : maximisation du débit global
- Pour  $\alpha \rightarrow 1$  : équité proportionnelle
- Pour  $\alpha = 2$  : potential delay minimisation
- Pour  $\alpha \rightarrow \infty$  : équité max-min

Dans [21], Le Boudec présente une preuve élégante du cas max-min, en utilisant une famille plus générale que celle de notre formulation. Dans [17], les auteurs s'intéressent au cas d'un système où le processus d'arrivée et de départ des connexions est dynamique et en étudient la stabilité et la performance du système. Leurs résultats montrent que cette famille d'équité est efficace dans le sens où les temps de réponse restent finis tant que la demande est inférieure ou égale à la capacité sur chacun des liens.

### 1.3.5.2 Exemple du réseau linéaire

Je m'intéresse à la valeur de  $x_0$  lorsque  $\alpha$  croit de 0 à l'infini. Cette valeur croit continuellement de 0 à  $c/2$ .

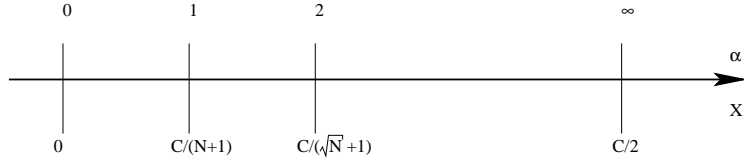


FIG. 1.7 – Allocation de la connexion 0 en fonction de  $\alpha$

On voit bien ici l'influence du paramètre  $\alpha$ . Plus  $\alpha$  est grand, plus l'allocation est équilibrée, au sens où les différences entre les allocations s'évanouissent.

### 1.3.6 Autres définitions d'équité

On trouve quelquefois dans la littérature, une équité comme étant la maximisation d'une fonction  $H$  ([21]) définie par :

$$H(x) = \sum_{n \in \mathcal{N}} u_n(x_n) - \sum_{l \in \mathcal{L}} g_l(f_l)$$

avec :

- $u$  une fonction d'utilité lié à chaque connexion.
- $g$  une fonction de coût lié à l'importance d'un lien donné
- $f$  le flot circulant sur un lien

D'une façon générale  $u$  représente le gain de l'utilisateur et  $g$  est le coût que voit l'administrateur du réseau. Cela permet par exemple de considérer la disparité de qualité des liens.

Dans la littérature,  $u$  est choisi comme étant égal à la fonction identité (maximisation du débit total) ou la fonction logarithme népérien (équité proportionnelle ou proportionnelle pondérée) et  $g$  est une fonction affine par morceaux, de la forme  $\sum_{l \in \mathcal{L}} a_l f_l$  avec  $a_l$  des constantes réelles positives et  $f_l$  représentant l'occupation du lien (la somme des bandes passantes des connexions qui le traversent).

## 1.4 Critère de Nash

Le concept de Nash Bargaining Solution (NBS) est issu de la théorie des jeux coopératifs. Il est défini par un ensemble d'axiomes qui sont intéressants en matière d'équité. Ces axiomes s'appliquent à des utilités associées aux joueurs. C'est tout l'intérêt et la différence entre ce critère et les critères classiques d'équité vus précédemment.

Nous pouvons déjà noter que le concept d'équité de Nash a déjà été utilisé (sous une forme simplifiée) dans des réseaux de topologie générales pour permettre le partage équitable de ressources ([117, 79]). Cependant dans les deux cas, les fonctions d'utilités étaient supposées linéaires. En fait, nous pouvons nous intéresser à n'importe quelle fonction d'utilité concave.

### 1.4.1 Deux mots sur la théorie des jeux

J'introduis ici quelques notions de théorie des jeux. Pour plus d'informations, le lecteur pourra se référer à [75, 94, 102, 111].

On distingue deux branches dans la théorie des jeux, appelées *jeux coopératifs* et *jeux non-coopératifs*. L'idée générale dans les deux cas est qu'un certain nombre de joueurs participent à un jeu (à prendre ici dans le sens large du terme). Chacun peut entreprendre un certain nombre d'actions et l'on s'intéresse aux éventuels équilibres qui sont issus de leurs comportements. Dans les deux cas, on associe des *utilités* aux joueurs, qui représentent l'attrait relatif des différentes actions.

#### 1.4.1.1 Jeux non-coopératifs

Dans de tels jeux, chaque joueur utilise une stratégie pour maximiser son utilité. Ce peut être par exemple une situation où un client peut acheter des biens à deux vendeurs indépendants. Chaque vendeur est libre de fixer ses prix et l'acheteur peut décider à quel vendeur s'adresser afin de maximiser son utilité, ce qui correspondra probablement dans ce cas à minimiser son coût. Chaque vendeur est libre de baisser ses prix afin d'attirer davantage le client. Dans ce genre de situation, les solutions éventuelles obtenues sont des *équilibres de Nash*.

**Définition 1.3 (équilibres de Nash)** *Dans un équilibre de Nash, aucun joueur ne peut augmenter son utilité sans la coopération des autres joueurs. D'une façon plus formelle, soit  $X$  un vecteur de stratégies. Je note  $(Y_i, X^{[i]})$  une perturbation de  $X$  définie de la façon suivante :*

$$(Y_i, X^{[i]}) = (X_1, X_2, \dots, X_{i-1}, Y_i, X_i, \dots, X_N)$$

. Cela revient à dire qu'un seul des joueurs a modifié sa stratégie. Je note  $\widehat{X}$  l'ensemble des vecteurs de stratégies possibles et  $u_i$  l'utilité associée au joueur  $i$ . Alors  $X$  est un équilibre de Nash si et seulement si  $\forall i, \forall y, (y_i, X^{[-i]}) \in \widehat{X}, u_i(y_i, X^{[-i]}) \leq u_i(X)$ .

Il est évident qu'un équilibre de Nash est en général non optimal au sens de Pareto. Un exemple classique de tel équilibre peut s'observer dans le trafic routier. En cas de forte circulation dans une intersection, une optimisation "égoïste", ou non-coopérative, consiste à avancer au maximum une fois le feu passé au vert. Ce faisant, chaque conducteur qui restera bloqué au milieu de l'intersection bloquera les automobilistes circulant dans l'autre sens au prochain changement de feu.

On se rend compte que lorsque tous les conducteurs adoptent ce comportement, le temps global passé au niveau de l'intersection augmente : le système n'est pas Pareto optimal. C'est pourquoi tous les jeunes conducteurs sont informés, lors du passage de leur permis de conduire, qu'il ne faut s'engager dans une intersection que si l'on sait que l'on ne s'y retrouvera pas coincé. Mais qui aujourd'hui applique ce beau principe ? La situation est d'autant plus compliquée que dans le cas de  $N$  voitures, si  $N - 1$  conducteurs adoptent ce comportement "peu courtois", le  $N$ ième utilisateur aura tout intérêt à l'utiliser également de façon à ne pas se voir fortement pénalisé face aux autres.

### 1.4.1.2 Les jeux coopératifs

Les problèmes de cette deuxième famille implique, comme leurs noms l'indiquent, une coopération entre les joueurs. Ces derniers sont alors désignés dans la littérature sous le nom de *bargainer*. Le résultat d'un marchandage porte le nom d'*équité*.

C'est à eux que je m'intéresse dans la suite. Plusieurs critères se trouvent dans la littérature (voir pour s'en convaincre [25] et les références qui y sont incluses). Ils ont d'abord été traités dans le cas de deux joueurs puis étendus pour la plupart à un nombre quelconque de joueurs. Un exemple typique est la négociation autour d'un prix.

Je me penche aujourd'hui sur le critère d'équité le plus célèbre, celui de Nash.

### 1.4.2 Petit voyage dans le temps

Le texte fondateur du NBS est [55]. Dans cet article, John Nash se penche sur le problème de 2 individus, que nous appelons Bill et Jack, qui ont l'occasion de collaborer afin d'obtenir un bénéfice mutuel. Aucune action prise par Jack sans le consentement de Bill ne peut affecter la satisfaction de Bill, et réciproquement.

On idéalise le problème en supposant que les deux individus sont parfaitement rationnels et que chacun peut comparer l'intérêt qu'il porte à chaque objet. Ils ont les mêmes qualités de marchandage et chacun a la connaissance complète des désirs de l'autre.

#### 1.4.2.1 Fonctions d'utilités

Chaque joueur a une fonction d'utilité. Cette fonction contient des valeurs de satisfaction du joueur pour chaque objet du marchandage. Évidemment, cette fonction n'est pas unique. En particulier, si  $u$  est une telle fonction, alors  $au + b$  en est une équivalente (pour  $a \in \mathbb{R}_*^+$  et  $b \in \mathbb{R}$ ). Ainsi,  $u(x) > u(y)$  signifie que l'objet  $x$  est d'avantage désiré que l'objet désigné par  $y$ . Aussi, la fonction d'utilité, telle que désignée par Nash est linéaire : si  $0 \leq p \leq 1$ ,  $u[px + (1 - p)y] = pu(x) + u(y)$ . Ainsi, la satisfaction qu'un joueur tire de la possession de deux objets est la somme des satisfactions qu'il tirerait de la possession de chacun d'entre eux.

La table suivante résume les objets possédés par chaque individu et la satisfaction qu'il lui associe.

objet	propriétaire	utilité pour Bill	utilité pour Jack
Livre	Bill	2	4
Lasso	Bill	2	2
Balle	Bill	2	1
Batte	Bill	2	2
Boite	Bill	4	1
Stylo	Jack	10	1
Jouet	Jack	4	1
Couteau	Jack	6	2
Chapeau	Jack	2	2

Si l'on trace dans le plan  $(U_1, U_2)$  avec  $U_1$  l'utilité de Bill et  $U_2$  celle de Jack l'ensemble des points correspondants à des échanges d'objets entre les deux individus, on obtient le



nuage de points de la figure 1.8. On note cet ensemble de point  $S$ , c'est l'ensemble des "solutions" possibles.

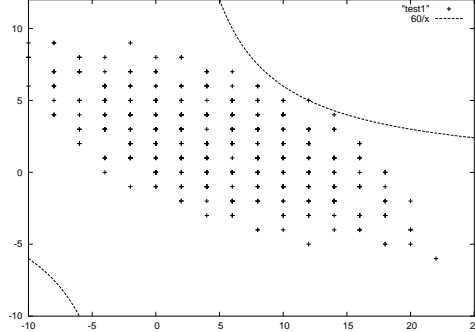


FIG. 1.8 – Utilités de Bill et Jack

En fait, si l'on supposait que dans un monde idéal chaque joueur pouvait obtenir une fraction de chaque objet, alors l'ensemble des paires  $(U_1, U_2)$  possibles serait exactement l'aire bornée par ces points. Parmi tous ces points correspondant à des marchandages possibles, lequel choisir ? La réponse donnée par Nash se présente sous forme d'axiomes.

#### 1.4.2.2 Les axiomes de Nash

Soit  $c(S)$  le point solution. On a la première condition :

*(Cond 1) Si  $\alpha$  est un point de  $S$  tel qu'il existe  $\beta \in S$  avec la propriété  $u_1(\beta) > u_1(\alpha)$  et  $u_2(\beta) > u_2(\alpha)$  alors  $\alpha \neq c(S)$ .*

Cela paraît évident. Si les deux joueurs peuvent augmenter leur utilité tout en restant dans le domaine solution, ils le feront.

Avec cette hypothèse, l'ensemble des points potentiellement solution se restreint à une seule courbe, celle du bord supérieur droit de l'ensemble de solutions.

La deuxième condition énoncée par Nash est la suivante :

*(Cond 2) Si  $S$  est symétrique par rapport à la droite  $u_1 = u_2$ , alors  $c(S)$  est un point de la forme  $(a, a)$ , c'est-à-dire un point de la droite  $u_1 = u_2$*

Cette condition exprime le fait que les deux joueurs ont des qualités de marchandage égales. Ainsi, si l'ensemble est symétrique, il n'y a aucune raison pour que l'un des joueurs voit son utilité plus importante que l'autre à l'issue de la transaction.

Enfin, la troisième condition est :

*(Cond 3) Si l'ensemble  $T$  contient l'ensemble  $S$  et que  $c(T) \in S$  alors  $c(T) = c(S)$ .*

Cette condition est plus difficile à comprendre. Soit  $T$  l'ensemble des marchandages possibles. Si deux individus seraient d'accord pour dire que  $c(T)$  est un marchandage équitable, alors ils seront d'accord sur ce même accord si l'ensemble des marchandages est réduit à un ensemble contenant  $c(t)$ .

#### 1.4.2.3 La résolution

L'idée de Nash est alors d'appliquer des transformations linéaires aux fonctions d'utilités des joueurs (c'est-à-dire d'utiliser des représentations équivalentes) de sorte que le nuage

de points solutions se trouve sous une droite de type  $u_2 = a - u_1$  pour un  $a \in R_+^*$  et que  $(a/2, a/2)$  soit un point solution. On peut donc alors considérer un rectangle dont cette droite est l'un des cotés et qui contient le nuage de point et lui appliquer les hypothèses (Cond 2) et (Cond 3). Le point solution sera alors de coordonnées  $(a/2, a/2)$ . L'idée est représentée dans la figure 1.9.

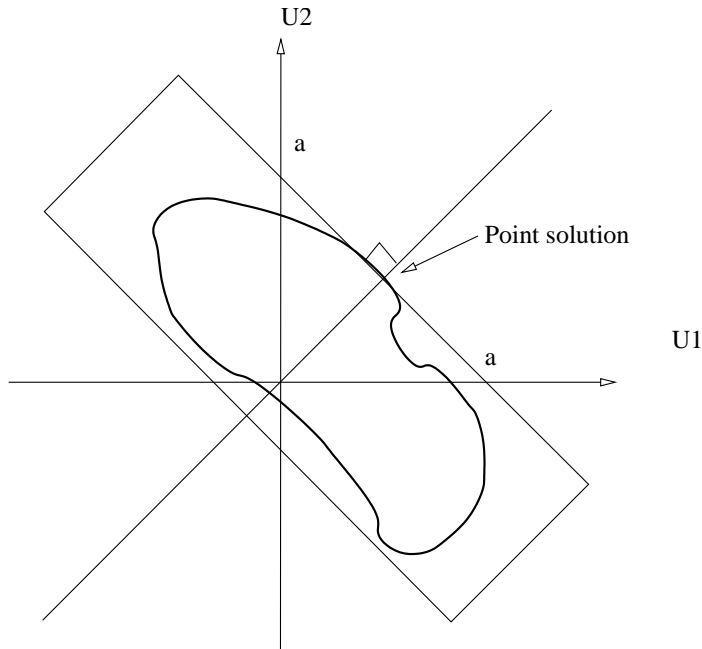


FIG. 1.9 – Normalisation des fonctions d'utilités

Dans l'exemple de Bill et Jack, le nuage de points a l'allure de 1.10.

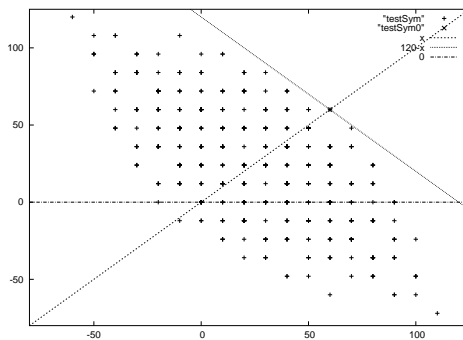


FIG. 1.10 – Utilités de Bill et Jack "normalisées"

Cette approche correspond en fait à maximiser  $u_1 u_2$  dans le problème initial. D'une façon géométrique, il suffit de chercher l'hyperbole tangente à  $S$ , comme représentée dans la figure 1.8. La solution de notre problème est obtenue pour  $u_1 u_2 = 60$  (avec les formulations initiales de  $u$  et  $v$ ) : Bill donne à Jack son livre, son lasso, sa balle et sa batte. En échange Jack lui donne son stylo, son jouet et son couteau.

### 1.4.3 L'évolution

La formulation de Nash a été légèrement modifiée par la suite.

On considère tout d'abord souvent que dans le cas où les deux joueurs ne parviennent à trouver un accord, ils ressortent chacun du jeu avec une certaine utilité, notée  $d_i$ ,  $1 \leq i \leq 2$  (utilité de désaccord).

De plus, les utilités ne sont plus limitées à des fonctions linéaires mais seul l'équilibre l'est. Pour comprendre ce que cela signifie, j'introduis :

- $S$  l'ensemble des états possibles
- $d$  l'utilité obtenue dans le cas où les joueurs ne parviendraient pas à un accord
- $g_i$  la fonction qui, a un couple  $(S, d)$  associe l'utilité du joueur  $i$ .

$$g_i : (S, d) \mapsto u_i.$$

Alors, si je note  $S'$  et  $d'$  les images de  $S$  et  $d$  par une transformation affine  $\phi$ <sup>3</sup>, alors l'image par toute fonction  $g_i$  de  $(S', d')$  sera l'image par la même transformation affine de  $g_i(S, d) : g_i(\phi(S), \phi(d)) = \phi(S, d)$ .

Cette propriété est la quatrième condition de Nash. Associée au trois précédentes, elle permet de caractériser de façon unique une solution du marchandage. De façon équivalente, elle est l'unique solution du problème d'optimisation  $\max_{u_i, i \in \{1, 2\}} \prod_i (u_i - d_i), \forall i \in \{1, 2\}, u_i \geq d_i$ .

En 1999, Muthoo reprend dans le premier chapitre de son livre [86] le concept de Nash. Cet économiste s'interroge sur le bien fondé des axiomes. Il présente en outre deux exemples, celui du contrôle du crime, et l'autre relève du domaine de l'économie.

Muthoo se limite, comme Nash, au cas de 2 joueurs. En fait, tous les résultats qu'il présente sont directement généralisables à  $n$  joueurs.

Le premier axiome représente le fait que la solution est Pareto optimale. Ce n'est pas le cas dans toutes les négociations, en particulier lors des traités de paix, mais cela est parfaitement justifié dans le cadre d'une allocation dans les réseaux. Le deuxième axiome atteste de la symétrie du problème et le quatrième de l'invariance à des représentations de la fonction d'utilité. Le troisième axiome, souvent appelé l'indépendance aux possibilités non pertinentes, est le plus critiqué. En effet, l'idée est que la décision des joueurs n'est pas modifiée si l'on enlève des choix non satisfaisants. Cependant, dans certains processus de négociations le résultat final peut être influencé par la présence de choix qui seront rejetés immédiatement. C'est par exemple le cas pour la négociation du prix d'une vente.

Ainsi, d'une façon générale, Muthoo recommande à chacun de se demander si les critères proposés par Nash sont applicables dans le cas de son marchandage. Il propose également dans le reste de son ouvrage d'autres équilibres de marchandages et discute de l'adéquation de chacun en fonction des situations rencontrées. On peut citer à ce sujet les solutions proposées par Raiffa ou Thomson ([25, 86]).

---

<sup>3</sup>c'est-à-dire que j'introduis deux paires  $(\alpha_1, \beta_1)$  et  $(\alpha_2, \beta_2)$  telles que  $S' = \{(\alpha_i u_i + \beta_i)_{i \in \{1, 2\}} : (u_i)_{1 \leq i \leq 2} \in S\}$  et  $d' = (\alpha_i d_i + \beta_i)_{1 \leq i \leq 2}$ .

### 1.4.4 Reformulation

La formulation que l'on utilise aujourd'hui, la généralisation à  $n$  joueurs et le formalisme mathématique ont été apportés par deux mathématiciens roumains : Anton et Maria Ştefănescu ([34]). Ils ont donné au problème sa forme actuellement utilisée.

Soit  $X \subset \mathbb{R}^n$  un ensemble convexe fermé non vide et  $f_i : X \rightarrow \mathbb{R}, i = 1, \dots, k$  des fonctions bornées supérieurement. Le problème de marchandage est alors un problème d'optimisation multi-objectif.

Pour tout ensemble  $U \subset \mathbb{R}^k$  non vide, convexe et fermé et  $u_0 \in \mathbb{R}^k$ , on note  $U_0$  l'ensemble  $U_0 = \{u \in U | u_0 \leq u\}$ .

Le lien avec le problème de Nash est évident.  $X$  est l'ensemble des stratégies possibles, les fonctions  $f$  sont les utilités associées aux joueurs,  $U$  est l'espace des utilités et  $u_0$  le vecteur d'utilités à partir duquel les joueurs acceptent le marchandage.

La solution est alors donnée sous forme axiomatique dépendant uniquement de  $U$  et de  $u_0$  (définition 1.4). Les auteurs montrent ensuite qu'en considérant l'ensemble particulier  $U = \{u \in \mathbb{R}^k, \exists x \in X, u \leq f(x)\}$  on obtient la formulation équivalente d'optimisation (théorème 1.1).

### 1.4.5 Adaptation pour les réseaux

En 1991, [79] présente un cadre mathématique adaptant le concept du marchandage de Nash aux réseaux de Jackson. Quelques années plus tard, Yaïche, Mazumdar et Rosenberg présentent dans [117] une adaptation des formules de Ştefănescu pour l'allocation de bande passante dans les réseaux. Leur idée est la suivante.

- $\mathcal{N}$  est le nombre de connexions (ou d'utilisateurs),
- $Y$  l'ensemble des vecteurs d'utilités possibles. De façon évidente,  $Y$  est un sous ensemble convexe fermé de  $\mathbb{R}^n$  ( $Y \subset \mathbb{R}^n$ ).
- Chaque utilisateur du réseau possède une fonction d'utilité  $f_i$ .
- On note  $y_i^0$  l'utilité minimale requise par la connexion  $i$  et  $MR_i$  l'allocation associée :

$$y_i^0 = f_i(MR_i).$$

- $\mathcal{G} = \{(Y, y^0), Y \subset \mathbb{R}^n\}$  est la paire constitué de l'ensemble des utilités solutions éventuelles et des utilités minimales. Elle caractérise parfaitement un problème de marchandage.

**Définition 1.4** *Un mapping  $S : \mathcal{G} \rightarrow \mathbb{R}^n$  est un Nash Bargaining Point si :*

1. *Il garantit les **performances minimales**, c'est-à-dire :  $S(Y, y^0) \in Y^0 := \{y \in Y | y \geq y^0\}$ .*
2.  *$S(Y, y^0)$  est **Pareto optimal**.*
3. *Il est **linéairement invariable**. Autrement dit, le NBS est inchangé si les objectifs de performances sont changés d'échelle (de façon affine). Plus précisément, si  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  est définie par  $\forall i, \phi_i(u) = a_i u_i + b_i$ , alors  $S(\phi(y), \phi(y^0)) = \phi(S(Y, y^0))$ .*
4.  *$S$  est **symétrique**, c'est-à-dire qu'il ne dépend pas des noms de variables spécifiques. Ainsi, deux connexions ayant les mêmes mesures de performances et les mêmes utilités auront les mêmes fonctions de performances.*

5. *S n'est pas affecté par un élargissement de domaine si une solution du problème avec un domaine plus large peut être trouvé sur un domaine plus restrictif. Plus précisément, si  $V \subset U$ ,  $(V, u^0) \in \mathcal{G}$  et  $S(U, u^0) \in V$  alors  $S(U, u^0) = S(V, u^0)$ .*

**Définition 1.5** *Le point  $y^* := S(Y, y^0)$  est appelé le Nash Bargaining Point (NBP) et  $f^{-1}(y^*)$  est l'ensemble des Nash Bargaining Solutions (NBS).*

Notons  $X_0 := \{x \in X | f(x) \geq y^0\}$ .

**Théorème 1.1** [117, Thm. 2.1, Thm 2.2] et [34]. *Supposons les fonctions d'utilité  $f_i$  concaves, bornées supérieurement et définies sur  $X$  qui est un sous ensemble convexe et compact de  $\mathbb{R}^n$ . Soit  $J$  l'ensemble des utilisateurs qui peuvent obtenir des performances strictement supérieures à leurs performances initiales, c'est-à-dire  $J = \{j \in \{1, \dots, n\} | \exists x \in X_0, \text{ s.t. } y_j(x) > y_j^0\}$ . Supposons de plus que les fonctions  $\{f_j\}_{j \in J}$  sont injectives. Alors, il existe un unique NBP ainsi qu'un unique NBS  $x$  qui vérifie  $f_j(x) > f_j(x), j \in J$ , et c'est l'unique solution du problème  $P_J$  :*

$$(P_J) \quad \max \prod_{j \in J} (f_j(x) - y_j^0), \quad x \in X_0. \quad (1.5)$$

*De façon équivalente, il est l'unique solution de :*

$$(P'_J) \quad \max \sum_{j \in J} \ln(f_j(x) - y_j^0), \quad x \in X_0. \quad (1.6)$$

#### 1.4.5.1 Remarques

Le NBS correspond à l'équité proportionnelle dans le cas où les fonctions d'utilités sont linéaires et où la demande minimale est égale à 0. Il correspond à l'équité proportionnelle pondérée si les fonctions d'utilité sont de type  $x \mapsto x^k, k \in \mathbb{R}$ .

#### 1.4.6 La résolution de Yaiche, Mazumdar et Rosenberg

Dans la suite de leur article, les auteurs s'intéressent au cas où les fonctions  $f$  sont égales à l'identité et proposent des solutions centralisées et décentralisées pour résoudre le problème.

Dans leur version centralisée, ils adoptent une approche de Lagrangien, que nous généraliserons à la section 3.2. Ils proposent ensuite une implémentation possible du problème centralisé en utilisant son dual.

Leur version décentralisée est plus intéressante. Il est évident que si chaque source utilise un algorithme pour maximiser son utilité, l'équilibre obtenu sera un équilibre de Nash et donc ne pourra être un NBS.

Cependant, les auteurs montrent qu'une modification du critère à maximiser peut permettre d'obtenir le NBS de façon décentralisée. Magique ?

Ce résultat repose en fait sur le concept de pénalités ou taxes. On introduit dans le système  $n$  taxes, notées  $\alpha_n, n \in \mathcal{N}$ . On peut alors montrer, que dans le cas où les fonctions

d'utilités sont linéaires, le NBS est l'unique solution du problème décentralisé :

$$(C) \begin{cases} \max_{x_i} \ln(x_i - MR_i) - \alpha_i x_i \\ x_i \geq MR_i \\ x_i \leq PR_i \end{cases} \quad (R) \begin{cases} \max_x \sum_i \alpha_i x_i \\ x_i \geq MR_i \\ x_i \leq PR_i \\ (Ax)_l \leq C_l \end{cases} \quad (1.7)$$

Le problème d'optimisation (C) tourne à chaque source alors que l'optimisation (R) est effectuée au niveau du réseau. En fait, un résultat plus général avait déjà été établi dans [61] dans le cas où le but était de maximiser la somme des utilités des utilisateurs du réseau. Ainsi, c'est l'introduction des prix  $\alpha_i$  associés aux connexions qui permet de lier les deux parties de l'optimisation. L'intérêt de cette formulation est également que les fonctions d'utilités des joueurs n'ont pas besoin d'être communiquées au réseau.

## 1.5 Algorithmes classiques de partage de bande passante

De nombreux algorithmes de partages de bande passante ont été développés. Nous verrons que quelques uns sont basés sur un calcul explicite des taux à affecter à chaque connexion (cf sous-section 1.5.1), mais que la plupart sont des algorithmes décentralisés (cf sous-section 1.5.2).

### 1.5.1 Calculs explicites des taux

Certains algorithmes reposent sur le calcul explicite des débits à allouer aux différentes connexions. C'est par exemple le cas dans l'ATM. Avant le début de chaque transmission, la source est informée du débit qu'elle va pouvoir utiliser : c'est un réseau avec réservation. Dans le service ABR (Available Bit Rate) de l'ATM, le taux est imposé de façon dynamique à la source.

Dans [28] les auteurs présentent un algorithme qui converge en un nombre fini d'itérations vers une allocation exacte d'équité max-min. L'algorithme est basé sur le fait que les utilisateurs découvrent progressivement le taux  $x_n$  qui leur est affecté par comparaison avec le "débit annoncé" des liens qu'ils traversent. Le débit annoncé  $A_l$  du lien  $l$  est donné par la formule :

$$A_l = \frac{c_l - \sum_{n \in \Gamma_l} x_n}{n_l - g_l}$$

Avec :

- $n_l$  le nombre de connexions utilisant le lien  $l$ ,
- $\Gamma_l$  l'ensemble des connexions dont le goulot d'étranglement est autre que  $l$ ,
- $g_l$  le nombre de connexions de  $\Gamma_l$  qui utilisent le lien  $l$ .

Nous avons précédemment vus qu'une caractérisation d'une allocation max-min était que chaque charge possède un goulot d'étranglement. Ainsi :

$$\begin{cases} x_l < A_l & \text{si } n \in \Gamma_l, \\ x_l = A_l & \text{sinon.} \end{cases}$$

A chaque étape du processus, les utilisateurs mettent à jour un estimateur de leur allocation de bande passante, en positionnant  $x_l$  au taux annoncé de leur connexion. En même temps, les liens découvrent progressivement les membres de  $\Gamma_l$ .

D'autres algorithmes de calculs explicites des taux, étudiés dans le contexte de l'ABR, ont été repris dans [9] et [48]. L'article met en exergue la difficulté de trouver un compromis entre équité, stabilité, robustesse, vitesse de convergence et utilisation des liens. Les algorithmes imposent généralement de grosses charges de calculs sur les nœuds du réseau et nécessitent une implémentation uniforme tout au long du réseau afin d'avoir des résultats optimaux.

### 1.5.2 Algorithmes décentralisés

Le calcul explicite des débits est en général irréalisable dès que la fonction d'allocation que l'on cherche à optimiser devient complexe car la topologie du réseau n'est pas connue par chaque source. A la place sont donc développées des algorithmes décentralisés. Ils sont évidemment bien trop nombreux pour que l'on puisse en faire une description exhaustive.

Ils se composent généralement d'un algorithme au niveau de chaque source, qui réajuste dynamiquement le débit alloué en fonction de la congestion des liens et d'un algorithme au niveau des liens qui met à jour, implicitement ou explicitement une certaine mesure de congestion à chaque lien et en informe, (implicitement ou explicitement) les sources qui utilisent ce lien. Par exemple, dans la version actuelle d'Internet l'algorithme de source est effectué par TCP, et l'algorithme de liens par des protocoles de managements de files d'attente appelés AQM (Active Queue Management) tels que DropTail et Red ([73]).

Les protocoles de source fréquemment utilisés sont basés sur l'utilisation de fenêtres. La taille de la fenêtre représente le nombre maximal de paquets qui sont autorisés à être en transit dans le réseau. Le système de fenêtre peut alors être soit utilisé de bout en bout soit au niveau de chaque lien. Supposons par exemple qu'un réseau utilise un système de fenêtrage de bout en bout. Si la taille de la fenêtre est de  $k$ , et que la source a déjà transmis  $k$  paquets, alors celle-ci devra attendre un accusé de réception de la part de son destinataire indiquant qu'un ou plusieurs paquets sont arrivés à destination afin de pouvoir émettre un nouveau paquet. On voit ainsi que lorsque le réseau devient congestionné, les paquets mettent plus de temps pour arriver à destination, et donc que la fenêtre "glisse" plus lentement. Au niveau macroscopique, cela revient à une diminution du débit de la source.

Notons cependant que pour ces algorithmes, les connexions entre utilisateurs sont considérées comme des fluides parfaits. Les buffers sont assimilés à des réservoirs et les liens à des tuyaux. Cependant dans leur implémentation ces algorithmes doivent prendre en compte la nature discrète du trafic.

Le système des fenêtres était auparavant le seul contrôle de congestion utilisé. Puis, Jacobson a eu l'idée [53] d'adapter dynamiquement la taille de la fenêtre au niveau de congestion. On distingue aujourd'hui les protocoles à taille de fenêtre fixe et variable. Ils peuvent tout deux être utilisés pour obtenir des allocations correspondants à des équités diverses comme nous le voyons dans les sous-sections suivantes.

### 1.5.2.1 Algorithmes à taille de fenêtre fixe

#### Fenêtres de bout en bout.

En général, l'utilisation des fenêtres de contrôle mène à des débits fluctuants, c'est-à-dire que  $x_n$  varie avec le temps à cause des rafales du trafic. Cependant, on suppose dans ces approches que le réseau est équipé d'un mécanisme supplémentaire qui enlève de telles fluctuations et permet ainsi l'établissement d'un régime stationnaire. Suivant la politique implémentée à chaque noeud, on peut alors réaliser différents types d'équité. (Les démonstrations sont disponibles dans [78].)

On peut utiliser des fenêtres de taille fixe lorsque les buffers des liens sont suffisamment grands pour éliminer la possibilité de perte de données. Soit  $n$  une connexion avec une fenêtre de taille  $B_n$  (par exemple exprimée en octets) et  $T_n$  le RTT (Round Trip Time) associé à cette connexion, en excluant tout temps de file d'attente sur la voie de retour.

Dans le cas de *files FIFO* (First In First Out), les taux associés à chaque connexion sont les uniques solutions des problèmes d'optimisation :  $\max \sum_n B_n \ln(x_n) - x_n \cdot T_n$  avec les contraintes de positivité des débits et les contraintes de capacités des liens. On remarque ainsi que dans le cas où les RTT sont négligeables les débits alloués correspondent à *l'allocation équitable proportionnelle* avec des poids donnés par la taille des fenêtres.

Dans le cas des *files LQF* (Longest Queue First) les débits stationnaires sont alors solution de  $\max \sum_n B_n * x_n - \frac{1}{2} \cdot (x_n)^2 T_n$  soumises aux mêmes contraintes que précédemment. On remarque également que, lorsque les RTT sont négligeables, le système tend vers la *maximisation globale* du système avec des poids correspondants à la taille des fenêtres.

Si les noeuds utilisent une politique définie dans [78] sous le nom de *Fair Queueing* le système tend vers un équilibre réalisant une *équité max-min*.

On peut trouver une politique réalisant l'équilibre de *minimisation des délais potentiels* (c'est le cas d'équité avec le paramètre  $\alpha = 2$ ). Cette politique a été proposée par [78]. A chaque noeud, on implémente une file d'attente dans laquelle le taux de service est partagé entre les files proportionnellement à la racine carrée des contenus des buffers correspondants.

Enfin, dans [85] ont été étudiées des politiques possibles pour le cas de l'équité avec un paramètre quelconque.

#### Fenêtres par liens.

[71] propose une méthode basée sur une fenêtre de taille fixe. Elle présente l'originalité d'être utilisée sur chaque lien et non de bout en bout. Ainsi chaque session a une fenêtre différente au niveau de chaque lien. La politique de file d'attente qui est utilisée est celle du round-robin, c'est-à-dire que chaque lien parcourt périodiquement la liste des connexions qui le traversent. Chaque fois qu'une connexion est scrutée, elle a la possibilité d'émettre un paquet. Si elle n'en a pas de prêts quand vient son tour, alors l'ordonnanceur passe à la connexion suivante et ainsi de suite. L'utilisation de fenêtres permet ici d'éviter que des files d'attentes excessives ne se forment au niveau d'un lien. L'article montre que l'utilisation d'une politique de round robin couplée avec un contrôle par fenêtre peut effectivement être utilisé pour obtenir des allocations max-min équitables.



### 1.5.2.2 Fenêtres de taille variables

Cette série d'algorithmes est de loin la plus répandue. Notamment, c'est celle implémentée dans le protocole TCP. L'idée de tels protocoles est de faire varier la taille de la fenêtre de chaque connexion. Ainsi, si la taille de la fenêtre augmente, cela signifie que la source va pouvoir envoyer d'avantage de paquets dans le réseau et donc augmenter son débit. Si l'on suppose qu'il n'y a pas de perte de paquets dans le réseau et que la durée qui sépare l'émission d'un paquet de la réception d'un accusée de réception est constante et vaut RTT, cela signifie que dans un intervalle de temps de 1 RTT, la source pourra émettre  $k$  paquets si la taille de la fenêtre est de  $k$ . La taille de la fenêtre est donc alors proportionnelle au débit de la connexion.

Au moment où une connexion est initialisée, on lui alloue une petite taille de fenêtre. En l'absence de congestion, les sources augmentent alors la taille de leur fenêtre jusqu'à ce que de la congestion apparaisse. Les connexions réduisent alors brutalement leur fenêtre de sorte à faire disparaître la congestion puis le processus reprend.

Ce principe nécessite que la source ait un moyen de savoir quel est l'état de congestion du réseau. Pour cela des indications binaires de congestion sont issues indépendamment par les liens du réseau. En pratique, l'état de congestion du réseau peut être mesurée par le taux d'occupation des buffers, le taux mesuré du débit d'entrée ou une combinaison des deux.

Évidemment, le choix des taux d'augmentation et de décroissance est crucial. Leur but est double : ils doivent être choisis de façon à limiter les amplitudes des oscillations qui pourrait sinon conduire à une utilisation sous optimale du système et doivent assurer une convergence rapide quand la nombre de flot actifs change.

### Famille AIMD

Une famille populaire de ces taux correspond aux systèmes AIMD (Additive Increase Multiplicative Decrease). Chiu et Raj ont étudiés plusieurs familles de paramètres et les ont comparés suivant plusieurs critères, notamment en matière d'efficacité, de rapidité de convergence, de capacité à être distribués (on recherche en effet une politique qui minimise les échanges d'informations sur le réseau autres que les données des connexions). Dans [31], ils concluent la politique AIMD est optimale. Ce raisonnement a été également repris dans [21].

Dans un système à taille de fenêtre variable, le temps est supposé discret. A chaque intervalle de temps, la source reçoit du réseau une information de congestion du réseau, notée  $y(t)$ . C'est une information binaire, telle que :

$$y(t) = \begin{cases} 1 & \text{si le réseau est congestionné} \\ 0 & \text{sinon} \end{cases}$$

A chaque instant  $t$ , je note  $f(t)$  la taille de la fenêtre. On a alors :

$$f(t+1) = u(y(t))f(t) + v(y(t)),$$

avec  $(u, v) : \{0, 1\} \rightarrow R$

Dans un système AIMD, en l'absence de congestion ( $y(t) = 0$ ), l'augmentation est linéaire, donc  $u(0) = 0$  et  $v(0) > 0$ . Dans le cas contraire, la fenêtre est réduite "multiplicativement", c'est-à-dire que  $v(1) = 0$  et  $u(1) \in R^+, u(1) < 1$ .

Cette famille est la plus déployée et la plus largement étudiée dans la littérature aujourd'hui car c'est celle qui est utilisée dans le protocole TCP. Kelly, Maulloo and Tan ont montré dans [61] que dans un réseau déployant une forme spécifique de contrôle de congestion AIMD, les taux alloués tendent à être distribués de façon proportionnellement équitable. Cependant, ce résultat n'est pas applicable à la version actuellement déployée de TCP. Un article très similaire est [114], qui considère des RTT hétérogènes. Les auteurs montrent alors que la fonction d'utilité maximisée par le réseau est :

$$\sum_i \frac{1}{\tau_i} \log \frac{x_i}{v(0) + u(1)x_i}$$

soumis aux contraintes du réseau.  $\tau_i$  est le RTT de la connexion  $i$ .

Une description précise des équilibres obtenus par les différentes versions et implémentations de TCP se trouve dans [73]. Low et Srikant y étudient également l'impact des AQM (Active Queue Management) sur l'équilibre obtenu. Ils montrent en particulier que l'équilibre obtenu par un couple TCP/AQM dépend très largement du protocole TCP utilisé tandis que le rôle de l'AQM est de le stabiliser. Ils prouvent également que l'implémentation déployée de la version Reno de TCP n'est pas équitable proportionnellement, contrairement à sa variante Vegas.

Enfin, dans [33] Crowcroft et Oeschlin proposent une variante de TCP. Leur but est de changer l'agressivité du protocole. Ceci est sous l'hypothèse que la tarification se fait au volume et que donc les utilisateurs ne vont pas chercher à augmenter de façon inconsidérée leur débit par rapport aux autres utilisateurs. Dans la version de Crowcroft et al., un paramètre  $n$  est utilisé pour représenter l'agressivité d'une connexion. Une  $n$ -connexion correspond alors à  $n$  connexions TCP. En fait, les auteurs se basent sur la version Reno de TCP, qui n'est pas proportionnellement équitable. Cependant, cette idée peut être adaptée afin de modifier l'agressivité d'algorithmes décentralisés existants réalisant un partage équitable de ressource pour une certaine valeur de  $\alpha$  en leur version pondérée.

### Algorithmes aléatoires

L'idée de tels algorithmes, proposés dans [78], est d'associer un processus stochastique aux tailles de fenêtres. Dans ces modèles, on suppose que chaque source peut savoir instantanément si une augmentation mènera à la saturation d'un lien. Ces algorithmes sont seulement présentés sur le plan théorique.

Les auteurs utilisent un processus de naissance et de mort, c'est-à-dire une chaîne de Markov avec pour chaque état  $n$  deux possibilités. Soit l'on passe à l'état  $n + 1$  avec la probabilité  $b_n$  (naissance), soit l'on retourne à l'état  $n - 1$  avec la probabilité  $d_n$  (mort) comme illustré en figure 1.11. A chaque instant, la taille de la fenêtre peut ainsi être augmentée ou diminuée d'une unité. Les auteurs montrent alors qu'un choix judicieux des probabilités  $b_n$  et  $d_n$  permet d'obtenir différents équilibres.

Ainsi, si l'on choisit des suites  $b_n$  et  $d_n$  constantes (i.e.  $\forall n, b_n = b, d_n = d$ ) alors, à mesure  $b/d$  croît, la distribution des taux se concentre sur les allocations qui maximisent le débit total  $\sum_N x_n$ .

Lorsque  $b_n = (n + 1)^a$  et  $d_n = (n - 1)^a$  avec un paramètre  $a$  quelconque, alors quand  $a$  devient grand, la distribution stationnaire se concentre sur les allocations qui maximisent

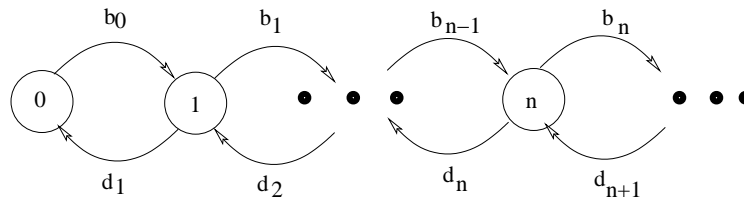


FIG. 1.11 – Processus de naissance et de mort

la somme des logarithmes des taux : c'est l'équité proportionnelle.

Si, pour deux paramètres  $A$  et  $M$ , on choisit  $b_n$  et  $d_n$  de sorte que  $\frac{b_{n-1}}{d_n} = \exp(A^{M-n})$  alors on peut montrer que si  $M$  est suffisamment grand la distribution  $\pi$  converge vers l'allocation max-min quand  $A \rightarrow \infty$ .

Enfin, en choisissant  $b_n$  et  $d_n$  telles que  $\frac{b_{n-1}}{d_n} = \exp\frac{a}{n(n-1)}$  pour  $n > 1$  alors on peut montrer que la distribution tend vers l'équité qui minimise le délai potentiel.

### 1.5.2.3 Indications de congestions

#### Indications binaires classiques.

Les protocoles de fenêtres à taille variables reposent sur le fait que chaque source a une information binaire sur l'état du réseau. A titre d'exemple, dans la version Reno de TCP cette information est celle de la perte de paquet : les sources augmentent ainsi leur fenêtre de congestion jusqu'à l'accident, c'est-à-dire l'apparition de pertes. La détection de perte d'un paquet peut se faire grâce aux accusés de réception. Dans le cas où une source envoie des paquets mais ne reçoit pas d'accusé après un certain laps de temps, elle peut supposer que ce paquet a été perdu.

On peut remarquer que mesurer la congestion d'un réseau par la perte des paquets n'est pas optimal. Des solutions alternatives ont donc été développées.

#### Anticipation des pertes.

Contrairement à la version Reno de TCP, Vegas quand à elle utilise comme critère de congestion le temps passé par les paquets de chaque connexion dans les files d'attente des liens. Le protocole est donc plus performant car l'on attend plus que des paquets soient perdus (c'est-à-dire que les buffers soient pleins) avant de réduire le débit d'une source.

L'idée d'essayer de réagir avant la perte de paquets est aussi présente dans les protocoles tels que l'ECN (Early Congestion Notification). L'idée de l'ECN est de marquer les paquets au niveau des routeurs. Un paquet est marqué s'il traverse un lien congestionné. Ainsi la source peut adapter son débit en conséquence. L'intérêt de cette méthode est alors que la source peut adapter son taux lorsque ses paquets traversent des liens à l'occupation très forte sans attendre la perte des paquets, d'où le nom de "early congestion notification". L'avantage de cette méthode par rapport aux méthodes de congestion utilisés dans l'Internet est que l'on peut obtenir des systèmes "loss-free", c'est-à-dire quasiment sans perte et obtenir ainsi une meilleure utilisation des ressources. L'ECN a été proposé et décrit dans [99, 38].

Deux techniques sont principalement utilisées sur l'ECN pour déterminer comment choisir les paquets à marquer : le Random Exponential Marking (REM, [10]) et le Self-normalized Additive Marking (SAM, [3]). L'ECN est par exemple utilisé dans [66] pour proposer un algorithme visant la maximisation de la somme des utilités des utilisateurs.

### Connaissance du nombre de liens congestionnés.

Dans l'ECN, la source ne connaît pas le nombre de liens congestionné traversé par ses paquets, ce qui en fait la principale différence avec l'algorithme présenté dans [56]. Dans cette article, les auteurs visent la maximisation de la somme des utilités des connexions. En considérant des utilités comme étant le critère général d'équité, leur algorithme peut donc être utilisé pour obtenir une allocation équitable. L'idée est que chaque source re-calcule à chaque itération le débit qu'elle va utiliser en fonction du nombre de liens congestionnés rencontrés.

Le principe est donc proche de celui des algorithmes à base de fenêtre de taille évoluant dans le temps, à la différence que l'on tient compte du nombre de liens congestionnés traversés. On a ainsi une mesure plus fine de l'état du réseau. En revanche, son implémentation nécessite une modification conséquente des éléments du réseau puisque ceux ci doivent être capables de modifier l'entête des paquets qui les traversent de façon à y inclure des informations de congestion (rappelons que dans les protocoles actuels, seul un bit est réservé pour les informations de contrôle de congestion). L'extension au cas du multicast se trouve dans [57].

### 1.5.3 Conclusion

De nombreux algorithmes ont été développés dans le but de maximiser une fonction du réseau. La tendance actuelle est dans la recherche d'algorithmes itératifs où la source adapte son débit (par l'intermédiaire d'une fenêtre de congestion) en fonction de l'état du réseau. L'idée est aujourd'hui d'utiliser des mesures de congestions qui permettent de réagir avant la perte de paquets dans les buffers, afin d'optimiser l'utilisation des ressources du réseau.

Un axe de recherche est l'obtention de protocoles qui convergent vers un certain débit et d'éviter les solutions oscillatoires, telles que la version Reno de TCP (car elle modifie à chaque RTT la taille de sa fenêtre). Un exemple de mécanisme non oscillatoire est TCP Vegas. Un autre axe est la modification des éléments de réseaux actuels, tels que les routeurs afin que ceux ci soient en mesure de fournir aux sources une indication plus précise de l'état du réseau. Enfin, on voit aujourd'hui un intérêt certain pour l'obtention de critères d'équités de plus en plus variés, et non plus restreints, comme il y a encore une quinzaine d'années, à l'équité max-min.

## 1.6 Les fonctions d'utilité

Revenons au concept de Nash. Son intérêt, par rapport aux critères utilisés aujourd'hui pour l'allocation de ressource dans les réseaux, est qu'il permet de prendre en compte les

différences d'intérêt des différents joueurs pour la ressource que l'on doit partager.

Il s'avère qu'en effet l'idée que la bande passante ou le débit n'ait pas la même valeur pour les utilisateurs d'un réseau est tout à fait justifiée, notamment dans le cas des applications temps réels.

Ainsi, si par exemple, dans ATM, l'idée est de partager la bande passante entre les connexions de la façon la plus égalitaire possible compte tenu des capacités du réseau, on comprend bien qu'une quantité  $\lambda$  de bande passante n'a pas la même valeur pour un utilisateur téléphonant par l'intermédiaire d'Internet et un autre lisant son courrier électronique.

On trouve alors quelquefois implémentées des "classes de services". Elles sont pour l'instant souvent liées à des notions de tarification. Ainsi un utilisateur pourrait payer pour un niveau de service parmi  $n$ . En général, l'idée est d'alors de modifier les priorités implémentées aux niveau des buffers pour donner une priorité aux paquets marqués comme étant d'une certaine classe. On peut noter qu'il n'y a pas (pour l'instant) de classes de services dans l'Internet. Cependant la situation pourrait changer dans les évolutions futures. Ainsi, dans l'architecture de Internet2 ([108]) les auteurs proposent un service, "QBone Premium Service" offrant une minimisation des pertes, des délais et de la gigue<sup>4</sup>. On remarque que ce service est principalement intéressant pour les utilisateurs d'applications multimedia et temps-reels.

Les fonctions d'utilité sont représentatives de l'importance de l'accès à la bande passante et notamment de l'impact d'une variation d'allocation dans l'impact de la qualité perçue. La discussion qui suit est d'ordre qualitative et directement issue de [105]. Pour des résultats quantitatifs, sur les communications audio par exemple, on pourra se référer à [54, 26].

Nous montrons dans la suite de cette partie les allures des fonctions d'utilités correspondants aux différents types d'applications.

### 1.6.1 Les fonctions ON/OFF

Elles sont caractéristiques des applications ayant des contraintes temps réel dures. En effet, ces applications ont besoin d'une certaine quantité de bande passante et tout surplus leur est inutile. Leur fonction à alors l'allure représentée à la figure 1.12.

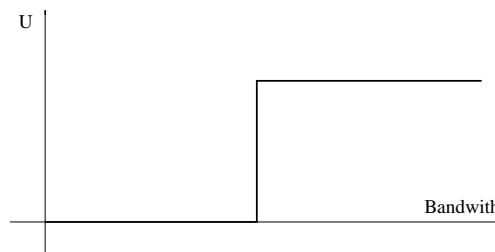


FIG. 1.12 – Fonction ON/OFF

<sup>4</sup>La gigue caractérise les variations dans les délais de transmission. C'est notamment un paramètre critique pour les applications temps réels car il est vital que les données parviennent de façon régulière à destination.

### 1.6.2 Applications élastiques

A l'opposé des applications temps réel se trouvent les applications sans contraintes de débit, comme les transferts de fichiers où le courrier électronique. Ces applications ne requièrent pas de bande passante minimale et leur utilité est croissante avec la bande passante. L'allure de la fonction est donc donnée la figure 1.13.

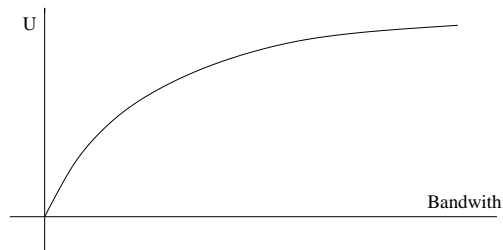


FIG. 1.13 – Application élastiques

### 1.6.3 Applications à débit variable

Il s'agit des applications ayant des contraintes souples de temps réel. Les exemples classiques sont la voix ou la vidéo sur IP. Dans de telles applications, le taux de compression des données est calculé en fonction de la quantité de ressource disponible.

Cependant, un taux maximum de compression existe, au-delà duquel la qualité du signal n'est plus acceptable pour l'utilisateur. Ce taux correspond donc à une valeur minimale de débit dont à besoin la connexion, on le note  $MR$ . De la même façon, il n'est pas nécessaire d'allouer une bande passante supérieure à un certain seuil car alors le gain en qualité ne serait plus perceptible par l'homme. On note ce maximum  $PR$ . A titre d'exemple, on considère pour la voix des débits compris entre 16 et 40 kb/s. Un utilisateur auquel on allouerait un débit de 200 kb/s n'aurait pas une impression de meilleure qualité que si son débit était de moitié.

Nous pouvons remarquer que ces fonctions sont dérivables en tout point de leur ensemble de définition.

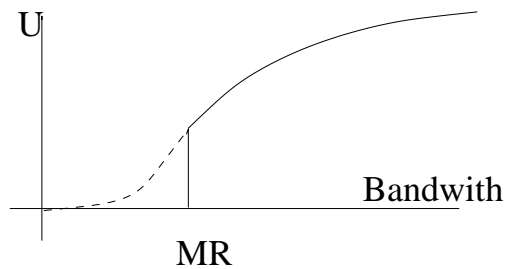


FIG. 1.14 – Application à débit variables

### 1.6.4 Conclusion

On peut associer aux applications des fonctions d'utilités. Les utilités des applications élastiques et à débit variable sont toutes deux concaves et croissantes. Au-delà d'une valeur minimale  $MR$ , elles ont la même allure (dans le second cas,  $MR = 0$ ). Comme la solution NBS est invariante par translation, on peut dire, sans perte de généralité que l'on peut traduire les fonctions d'utilités, dans le cas de l'équité de Nash, de sorte que  $f(MR) = 0$ .

Nous proposons dans notre étude d'appliquer la famille de critère d'équités présentée précédemment (Section 1.3) aux fonctions d'utilités des applications, plutôt qu'aux débits directement, comme faits dans les travaux précédents. Le critère d'équité que nous proposons s'énonce

$$\max_{n \in \mathcal{N}} \frac{[f_n(x_n)]^{1-\alpha}}{1-\alpha} \quad (1.8)$$

sous les contraintes du système

Comme expliqué par Nash, un équilibre atteint à l'issue de la compétition entre les connexions ne doit pas dépendre de la représentation des fonctions d'utilités choisie. C'est pourquoi nous proposons de considérer la représentation de l'utilité de chacun des joueurs qui est telle que  $f_n(MR_n) = 0$ .

On remarque qu'au delà de  $MR$ , la fonction est concave et croissante avec la bande passante. Nous proposons alors de l'approcher par une parabole avec certains paramètres, dépendant des applications.

#### Aproximation parabolique.

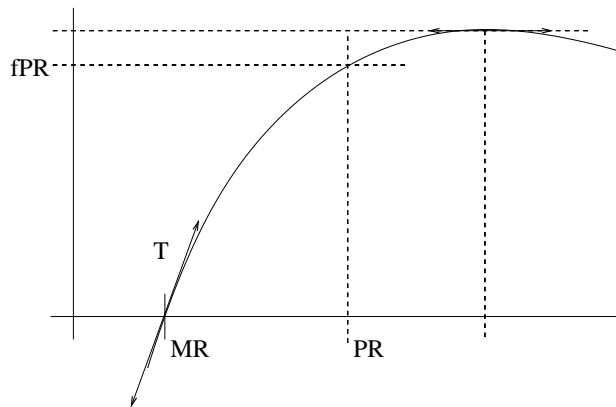


FIG. 1.15 – Fonction d'utilité quadratique

La forme générale d'une telle courbe est :

$$f_i(x_i) = c_i - a_i(x_i - b_i)^2. \quad (1.9)$$

On note alors :

- $MR_i$  le minimum demandé par la connexion

- $PR_i$  la bande passante maximale dont l'application a besoin
- $T_i$  la valeur de la tangente au point  $(MR_i, 0)$
- $fPR_i$  la valeur de l'utilité en  $PR_i$

Alors, la parabole peut être parfaitement définie soit par les valeurs de  $a_i, b_i$  et de  $c_i$  soit par les équations  $f_i(MR_i) = 0$ ,  $f_i(PR_i) = fPR$  et  $f'_i(MR_i) = T_i$ .

On peut noter cependant que, comme  $PR_i$  est situé dans la partie croissante de la courbe, alors :

$$\frac{1}{2}T_i(PR_i - MR_i) \leq fPR_i \leq T_i(PR_i - MR_i).$$

On peut alors définir la concavité de l'utilité,  $\beta$  comme étant :

$$fPR_i = T_i \cdot \beta_i \cdot (PR_i - MR_i).$$

Nous choisirons dans la suite comme paramètres de l'application :

$$\boxed{MR_i, PR_i, T_i \text{ et } \beta_i}$$

Et nous avons les équations suivantes :

$$a_i = T_i \frac{1 - \beta_i}{PR_i - MR_i}, \quad b_i = \frac{PR_i - (2\beta_i - 1)MR_i}{2(1 - \beta_i)}, \quad c_i = \frac{T_i PR_i - MR_i}{4(1 - \beta_i)} \quad (1.10)$$

On peut remarquer que  $1/2 \leq \beta < 1$  et que plus  $\beta$  est petit, plus la courbe est concave. La limite,  $\beta = 1$  est le cas linéaire.

## 1.7 Comparaison et propriétés des critères d'équité

### 1.7.1 Introduction

Le but de notre travail est d'appliquer les critères classiques d'équités dans les réseaux aux fonctions d'utilités des applications afin de proposer de nouveaux schémas d'allocations de ressources.

On se rend compte que lorsque l'on applique cette philosophie à l'équité proportionnelle on obtient le concept d'équité de Nash, issu de la théorie des jeux et introduit historiquement non pas comme sous forme d'un problème d'optimisation mais comme une série d'axiomes jugés intéressants pour définir l'équité.

On sait que le paramètre  $\alpha$ , utilisé dans l'expression générale de l'équité définit le compromis fait entre le désir d'équité et de rentabilité de l'opérateur de réseau. En appliquant ces critères aux fonctions d'utilité, l'intérêt du paramètre  $\alpha$  reste inchangé, même s'il n'est pas possible de quantifier son influence de façon analytique, sauf évidemment pour des réseaux très simples.

L'allocation des critères d'équités directement aux débits des connexions revient en fait à utiliser notre version généralisée des concepts d'équités à des utilités linéaires. L'intérêt d'introduire des fonctions quadratiques est, entre autre, de pouvoir étudier de façon simple l'influence de quelques paramètres sur les équilibres obtenus. En particulier, nous nous étudions ici l'influence de la concavité dans l'allocation équitable.



### 1.7.2 Impact de la concavité de la fonction d'utilité

Nous étudions ici l'impact de la concavité sur l'allocation NBS. Soient  $f$  et  $g$  deux fonctions différentiables sur le même intervalle  $[MR, PR]$ . Ces deux fonctions sont strictement positives sur  $(MR, PR]$ .

**Définition 1.6** *On dit que  $f$  est plus concave que  $g$  si, pour tout  $x \in (MR, PR]$ , la dérivée relative de  $f$  est plus petite que celle de  $g$ , c'est-à-dire que  $f'(x)/f(x) \leq g'(x)/g(x)$ .*

*Si  $f$  ou  $g$  n'est pas différentiable en  $x$ , on peut étendre la définition en considérant alors les supergradients de  $f$  et  $g$  : si  $\hat{f}(x)$  est le plus grand supergradient de  $f$  en  $x$  et  $\hat{g}(x)$  est le plus petit de  $g$  en  $x$  alors la condition devient :  $\hat{f}(x)/f(x) \leq \hat{g}(x)/g(x)$ .*

**Définition 1.7** *On dit qu'une allocation  $\bar{x}^1$  est plus équitable que  $\bar{x}^2$  dans le sens de NBS si  $\prod_{i=1}^n f_i(x_i^1) \geq \prod_{i=1}^n f_i(x_i^2)$ .*

Dans l'exemple suivant, on considère deux connexions avec des fonctions d'utilité  $f$  et  $g$  concurrentes sur un même lien de capacité  $c$ . Si nous avions ignoré les utilités des connexions, nous leur aurions donné une bande passante identique (selon par exemple la définition originelle de l'équité proportionnelle), qui est  $c/2$ .

**Proposition 1.2** *Dans le cas de deux connexions concurrentes sur un même lien, on augmente l'équité au sens du NBS en transférant de la bande passante de la connexion avec la fonction d'utilité la plus concave vers l'autre (en supposant bien sûr que cela ne viole pas les contraintes de demandes minimales et maximales des connexions).*

**Démonstration :** Nous avons en effet :

$$g(x + \epsilon)f(x - \epsilon) = g(x)f(x) \left( 1 + \epsilon \left[ \frac{g'(x)}{g(x)} - \frac{f'(x)}{f(x)} \right] + o(\epsilon) \right)$$

où  $o(\epsilon)$  est une fonction qui tend vers zéro quand elle est divisée par  $\epsilon$  et que  $\epsilon$  tend vers zéro. Nous concluons qu'il existe un  $\epsilon_0$  tel que pour tout  $\epsilon < \epsilon_0$ ,  $g(x + \epsilon)f(x - \epsilon) > g(x)f(x)$ . Ainsi, nous améliorons strictement l'équité en transférant une quantité  $\epsilon_0$  vers la connexion ayant l'utilité la moins concave.

En augmentant encore le montant transféré, nous devrions finalement atteindre un maximum local (puisque notre fonction est continue sur un interval compact). Ceci serait un maximum global car (1.6) est un problème de maximisation d'une fonction concave sur un ensemble convexe. Nous concluons que l'allocation équitable à la propriété que plus de bande passante est allouée à la fonction la moins concave. ■

**Exemple 1.1** *Soit deux connexions de bande passante  $x_1$  et  $x_2$  se partageant un même lien. Leurs fonctions d'utilités respectives sont  $f$  et  $g$  comme représentées sur la figure 1.16.*

$$\begin{cases} f(x) = 3x\mathbb{1}_{\{0 \leq x \leq 1\}} + (2+x)\mathbb{1}_{\{x > 1\}} \\ g(x) = 2x \text{ pour } x \geq 0. \end{cases}$$

Alors

$$f'(x)/f(x) = \begin{cases} x^{-1} & \text{pour } x \in [0, 1), \\ (2+x)^{-1} & \text{pour } x \geq 1 \end{cases}$$

alors que  $g'(x)/g(x) = x^{-1}$  partout.

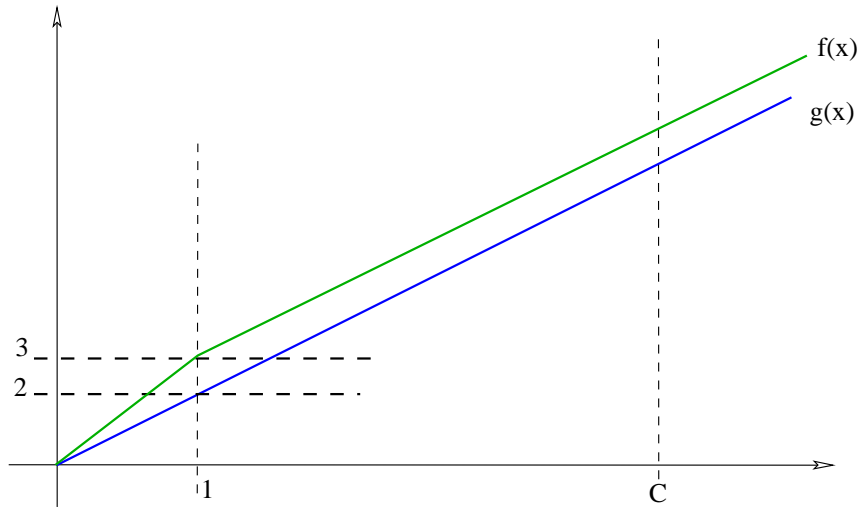


FIG. 1.16 – Fonctions d'utilités des 2 connexions concurrentes.

(En  $x = 1$ ,  $f$  n'est pas différentiable mais ses supergradients à ce point sont l'ensemble  $[1/3, 1]$ ). Ainsi  $f$  est plus concave que  $g$ . Nous supposons que pour les deux connexions  $PR > c$ .

Le NBS est les arguments de :

$$\zeta(x) = \max_x f(x)g(c - x) = \max \left( \max_{x \in [0,1]} h(x), \max_{x > 1} k(x) \right)$$

avec  $\begin{cases} h(x) = 6x(c - x) \\ k(x) = 2(x + 2)(c - x). \end{cases}$

Le NBS est représenté à la figure 1.17. On peut distinguer 3 régions :

- (i)  $c < 2$ , où  $\zeta(c) = 3c^2/2$  et le NBS vaut  $(x_1^*, x_2^*) = (c/2, c/2)$ .
- (ii)  $2 \leq c < 4$ , où  $\zeta(c) = 6(c - 1)$  et  $(x_1^*, x_2^*) = (1, c - 1)$ .
- (iii)  $c \geq 4$ , où  $\zeta(c) = 2(1 + c/2)^2$  et  $(x_1^*, x_2^*) = (c/2 - 1, c/2 + 1)$ .

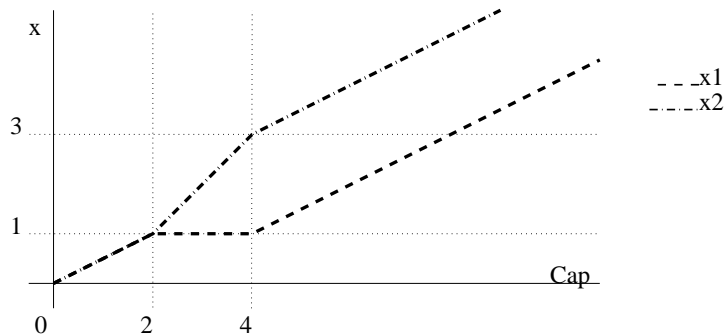


FIG. 1.17 – NBS de deux connexions se partageant un lien.

Nous voyons dans cet exemple qu'en fait la fonction la moins concave reçoit au moins autant de bande passante que l'autre, et que la différence augmente avec la capacité. Il est impressionnant de remarquer qu'il y a une région dans laquelle une augmentation de la capacité ne bénéficie qu'à l'une des connexions. Cet exemple illustre la puissance de l'approche

NBS : l'équité proportionnelle originelle, où même l'équité proportionnelle pondérée affecterait une proportion de la capacité à chaque connexion serait indépendante des capacités. En revanche, les concepts d'équité relatives aux équités allouent la bande passante de façon dynamique : la proportion allouée à chaque connexion est une fonction de la capacité.

## 1.8 La programmation semi-définie positive (SDP)

La SDP (Semi-Definite Programming) est un programme mathématique qui est une extension de la programmation linéaire. Il est polynomial et tractable en pratique. Il existe des programmes SDP dans le domaine public<sup>5</sup>. SDP résoud un problème de minimisation d'une combinaison linéaire de variables (données par le produit scalaire d'un vecteur  $L$  et du vecteur de variables) sous la contrainte de semi-définie positivité d'une matrice générale symétrique  $P$  dont les entrées sont soit des variables, soit des constantes. La matrice est diagonale par blocs. Les entrées du problème sont donc en fait l'ensemble des blocs constitutifs de la matrice et le vecteur objectif. Pour plus de détails sur la SDP, se référer à [22].

Dans cette section, nous verrons comment passer de la programmation linéaire à la programmation SDP, quelques rappels sur les matrices positives et symétrique et enfin quelques exemples classiques de résolutions par SDP.

### 1.8.1 De la programmation linéaire à la programmation Semi-Définie Positive

On présente la SDP comme étant une extension de la programmation entière. Mais qu'est-ce que la programmation entière ? En quoi la SDP lui est-elle liée ? C'est ce à quoi cette sous-section désire apporter des réponses.

Des explications plus exhaustives se trouvent dans la littérature. Seuls les principaux résultats sont reportés ici, principalement inspirés des travaux de Nemirovski ([89]).

#### 1.8.1.1 La programmation linéaire

##### Définition

La programmation linéaire est née en 1948, de George Dantzig. Un programme linéaire (LP) est un programme d'optimisation de la forme :

$$\min\{c^T x \mid Ax \geq b\}$$

où

$$\left\{ \begin{array}{l} x \text{ est le vecteur inconnu} \\ c \text{ est un vecteur de coefficients (donné)} \\ A \text{ est une matrice de contraintes } m \times n \text{ et} \\ b \text{ est le vecteur de contraintes} \end{array} \right.$$

---

<sup>5</sup>voir par exemple <http://www.cs.nyu.edu/cs/faculty/overton/sdppack/sdppack.html>

(LP) est appelé :

- $$\left\{ \begin{array}{l} \mathbf{Faisable} \text{ si l'ensemble } \{x \mid Ax - b \geq 0\} \text{ est non vide.} \\ \mathbf{Borné en minimum} \text{ si il est soit infaisable, soit si la fonction objectif } c^T x \text{ est bor-} \\ \text{née inférieurement.} \\ \mathbf{Solvable} \text{ si il est faisable, borné en minimum et si la valeur optimale est atteinte.} \end{array} \right.$$

En général, un système faisable et borné en minimum n'est pas toujours solvable. Par exemple la fonction inverse (c'est-à-dire  $f : x \rightarrow 1/x$ ) sur l'ensemble  $\{x \mid x \geq 1\}$  est faisable et bornée par zéro mais le système n'est pas solvable.

Une propriété intéressante des problèmes (LP) est qu'**un système faisable et borné est toujours solvable**.

### Problème dual

Le problème dual de (LP) est :

$$\max_y \{b^T y \mid A^T y = c, y \geq 0\}$$

Il est intéressant de noter que le problème dual d'un programme linéaire est encore un programme linéaire. On a le théorème de faible dualité :

*La valeur optimale du problème dual est inférieure ou égale à la solution optimale du problème primal.*

### **Théorème 1.2 (Dualité en Programmation Linéaire)**

1. *La dualité est symétrique : le problème dual est équivalent au primal.*
2. *Pour chaque solution faisable du dual, la valeur de la fonction objective est inférieure ou égale à la valeur de la fonction objectif du primal appliquée en tout point faisable du primal*
3. *Les 5 propriétés suivantes sont équivalentes :*
  - *Le primal est faisable et borné (inférieurement)*
  - *Le dual est faisable et borné (supérieurement)*
  - *Le primal est solvable*
  - *Le dual est solvable*
  - *Le primal et le dual sont faisables*

*De plus, si ces conditions sont vérifiées, alors les solutions optimales du primal et du dual sont égales.*

### **Théorème 1.3** (Condition nécessaire et suffisante d'optimalité en programmation linéaire)

*Une paire  $(x, y)$  de solutions primale et duale sont les solutions optimales de leurs problèmes respectifs si et seulement si :*

$$y_i [Ax - b]_i = 0, i = 1, \dots, m \tag{1.11}$$

*ou, de façon équivalente*

$$c^T - b^T y = 0 \tag{1.12}$$

### 1.8.1.2 Programmation conique

On peut étendre la programmation linéaire à la programmation conique. Pour comprendre ce que cela signifie, il faut tout d'abord définir un cône pointé convexe. Soit  $E$  un espace Euclidien de dimension finie.

**Définition 1.8 (Cônes pointés convexes)** *Un cône pointé convexe  $\mathcal{K}$  est un sous-ensemble de  $E$  qui vérifie les conditions :*

- $\mathcal{K}$  est non vide et stable pour l'addition :  $a, a' \in \mathcal{K} \Rightarrow a + a' \in \mathcal{K}$
- $\mathcal{K}$  est un ensemble conique :  $\forall a \in \mathcal{K}, \forall \lambda \geq 0, \lambda a \in \mathcal{K}$ ,
- $\mathcal{K}$  est pointé :  $a \in \mathcal{K}$  et  $-a \in \mathcal{K} \Rightarrow a = 0$ .

Chaque cône pointé convexe induit une relation d'ordre dans  $E$ . Cette relation est partielle. On la note  $\geq_{\mathcal{K}}$  :

$$a \geq_{\mathcal{K}} b \Leftrightarrow a - b \geq_{\mathcal{K}} 0 \Leftrightarrow a - b \in \mathcal{K}$$

**Définition 1.9 (Programmation conique)** *Soit  $\mathcal{K}$  un cône de  $\mathbf{E}$ . (convexe, pointé et d'intérieur non vide). Soit un vecteur  $c \in \mathbb{R}^n$  (vecteur objectif) et un mapping linéaire :*

$$\begin{aligned} A : x &\mapsto Ax \\ \mathbb{R}^n &\rightarrow \mathbf{E}. \end{aligned}$$

et  $b \in E$ . Le problème d'optimisation suivant :

$$\min_x \{c^T x \mid Ax \geq_{\mathcal{K}} b\}$$

est un problème conique.

**Définition 1.10 (Cône dual)** *Soit  $\mathbf{E}$  un espace euclidien de dimension finie, muni d'un produit scalaire  $\langle \cdot, \cdot \rangle$  et soit  $\mathcal{K} \subset \mathbf{E}$ . Soit  $\mathcal{K}^*$  le dual de  $\mathcal{K}$  défini par :*

$$\mathcal{K}^* = \{\lambda \in \mathbf{E} : \langle \lambda, a \rangle \geq 0, \forall a \in \mathcal{K}\}.$$

C'est un cône convexe fermé.

- Si  $\text{int}(\mathcal{K}) \neq 0$ , alors  $\mathcal{K}^*$  est pointé.
- Si  $\mathcal{K}$  est un cône convexe pointé fermé alors  $\text{int}(\mathcal{K}) \neq 0$ .
- Si  $\mathcal{K}$  est un cône convexe fermé alors  $\mathcal{K}^*$  également et le cône dual de  $\mathcal{K}^*$  est  $\mathcal{K}$  lui-même :  $(\mathcal{K}^*)^* = \mathcal{K}$ .

**Théorème 1.4 (théorème de dualité conique)** *Considérons un problème conique :  $c^* = \min_x \{c^T x \mid Ax \geq_{\mathcal{K}} b\}$  et son problème dual associé :  $b^* = \max_{\lambda} \{\langle b, \lambda \rangle \mid A^* \lambda = c, \lambda \geq_{\mathcal{K}} 0\}$ . On a le théorème suivant :*

1. La dualité est symétrique : le problème dual est un problème conique et le dual du dual est le primal.
2. Le trou de dualité  $c^T x - \langle b, \lambda \rangle$  est positif ou nul pour toute paire faisable  $(x, \lambda)$ .
3. (a) Si le primal est borné et strictement faisable (c'est-à-dire qu'il existe  $x$  tel que  $Ax >_{\mathcal{K}} b$ ) alors le dual est solvable et les valeurs optimales des deux problèmes sont égales ( $c^* = b^*$ ).

(b) Si le dual est borné et strictement faisable (c'est-à-dire qu'il existe  $\lambda \succ_K 0$  tel que  $A^*\lambda = c$ ) alors le primal est solvable et  $c^* = b^*$ .

4. Supposons que au moins un des problèmes est borné et strictement faisable. Alors une paire faisable  $(x, \lambda)$  est une paire de solutions optimales à son problème respectif

(a) Si et seulement si  $\langle b, \lambda \rangle = c^T x$  (pas de trous de dualité)

(b) ou, si et seulement si  $\langle \lambda, Ax - b \rangle = 0$ .

**Corollaire 1** Supposons que les deux problèmes (primal et dual) soient strictement faisables. Alors les deux problèmes sont solvables, les valeurs optimales sont égales et chacune des deux conditions (4a et 4b) est une condition nécessaire et suffisante pour l'optimalité d'une paire faisable primal-dual.  $\square$

Plusieurs familles de problèmes coniques sont intéressantes, sur le plan de la théorie et des applications. Les cônes invoqués dans ces problèmes sont suffisamment simples pour que l'on puisse décrire explicitement le cône dual. La simplicité relative des cônes sous-jacents permet également de développer des méthodes calculatoires pour les problèmes coniques correspondants. À part la programmation linéaire, deux autres problèmes coniques généraux sont d'extrême importance : les cônes quadratiques et la programmation semi-définie.

### 1.8.1.3 Programmation conique-quadratique

La programmation conique quadratique est obtenue en considérant  $\mathcal{K}$  comme produit direct de cônes dit de Lorentz ou de "crème glacée" (ice cream cone)

$$L^m = \{x = \{x_1, \dots, x_m\} \in \mathbb{R}^m \mid x_m \geq \sqrt{x_1^2 + \dots + x_{m-1}^2}\}, m \geq 2$$

Et

$$K = \left\{ y = \begin{pmatrix} y[1] \\ y[2] \\ \dots \\ y[k] \end{pmatrix} \mid y[i] \in L^m, i = 1, \dots, k \right\}.$$

Des formulations équivalentes se trouvent dans la littérature. L'intérêt de ce cône est qu'il est son propre dual :  $K_* = K$ .

**Remarque 1** Bien souvent, les problèmes d'optimisations se présentent sous la forme :

$$\min_x \{f(x) \mid x \in X\}.$$

Si  $f$  n'est pas linéaire, on peut la remplacer par un ensemble de contraintes et le problème devient :

$$\min_x \{t : (t, x) \in \widehat{X} \equiv \{(x, t) : x \in X, t \geq f(x)\}\}.$$

La question devient alors "est-ce que le système de contraintes est représentable par des contraintes coniques-quadratiques ?".  $\square$

Un intérêt de cette famille de cônes, est que l'on peut construire une sorte de dictionnaire des ensembles de contraintes qui peuvent être représentées par des cônes quadratiques.

On trouve dans [89] et dans [88] une méthode pour construire un tel dictionnaire. Par exemple, une fonction constante, affine, la norme Euclidienne, une branche d'hyperbole, les formes quadratiques, les fonctions puissances sont représentables par des contraintes quadratiques coniques. De plus, certaines opérations préservent la "représentabilité" des fonctions : la prise de maximum, la somme avec des poids positifs ou nuls, la somme directe, etc.

#### 1.8.1.4 La programmation Semi-Définie Positive

Soit  $E = S^m$  l'ensemble des matrices de taille  $m \times m$  symétriques. On note  $\mathcal{K} = S^m_+$  le cône des matrices semi-définies positives.

Une matrice symétrique de taille  $m \times m$  est dite semi-définie si  $\forall x \in R^m, x^T A x \geq 0$ . De façon équivalente toutes ses valeurs propres sont positives ou nulles.

Un problème SDP est :

$$\min_x \{c^T x \mid \mathcal{A}x - B \succeq 0\}$$

Dans cette formulation  $x$  est un vecteur,  $\mathcal{A}$  est un mapping linéaire de  $R^n$  dans  $S^m$  et  $B$  une matrice symétrique.

En fait, une façon naturelle de caractériser un mapping linéaire  $\mathcal{A} : R^n \rightarrow S^m$  est par la donnée d'une suite  $A_i$  de  $n$  matrices de  $S^m$  telles que :

$$\forall x = (x_1, \dots, x_n)^T \in R^n, \mathcal{A}x = \sum_{j=1}^n x_j A_j$$

D'où la nouvelle expression SDP :

$$\min_x \{c^T x \mid \sum_{j=1}^n x_j A_j - B \succeq 0\}$$

Le problème dual associé est alors :

$$\max_Y \{tr(BY) \mid Y \succeq 0, tr(A_j Y) = c_j, j = 1, \dots, n\}$$

Ceci est la formulation utilisée par Nemirovski ([89]) et par Laurent et Rendl ([68]).

#### Formulation équivalente.

Une autre formulation équivalente se trouve dans la littérature. Ainsi, on trouve dans [46] :

$$\min c^T vec(X) \quad \text{tel que} \quad \begin{cases} Avec(X) = b \\ X \succeq 0. \end{cases}$$

Avec  $X \in S_n^+$ ,  $c \in R^{n^2}$ ,  $b \in R^m$  et  $A \in M_{m, n^2}$ .<sup>(6)</sup>

---

<sup>6</sup>Notation : si  $X$  est la matrice  $X = (x_{i,j}), 1 \leq i, j \leq n$  alors  $vec(X)$  est le vecteur  $(x_{1,1}, \dots, x_{1,n}, x_{2,n}, \dots, x_{n,n})$ .

On remarque que, en notant  $\mathbf{A}$  le mapping de  $S_n$  dans  $R^m$  associé et défini par <sup>7</sup> :

$$\mathbf{A}X = \begin{bmatrix} \langle A_1, X \rangle \\ \langle A_2, X \rangle \\ \dots \\ \langle A_n, X \rangle \end{bmatrix}.$$

Et  $C \in M_n$  la matrice associée à  $c$  obtenue par :

$$C = \begin{pmatrix} c_1 & c_{1+n} & \dots & c_{1+n(n-1)} \\ c_2 & c_{2+n} & \dots & c_{2+n(n-1)} \\ \vdots & \vdots & & \vdots \\ c_n & c_{2n} & \dots & c_{n^2} \end{pmatrix} \quad (1.13)$$

Alors la formulation SDP devient :

$$\left\{ \begin{array}{l} \min \langle C, X \rangle \\ \text{tel que } \begin{cases} \mathbf{A}X = b \\ X \succeq 0 \end{cases} \end{array} \right. \quad \text{et le dual s'écrit alors : } \left\{ \begin{array}{l} \max \langle b, y \rangle \\ \text{tel que } \begin{cases} \mathbf{A}^T y + Z = C \\ y \in R^m, Z \succeq 0 \end{cases} \end{array} \right.$$

Nous utilisons dans la suite cette dernière formulation.

### Propriété des problèmes SDP

Le problème primal et le problème dual sont des problèmes SDP.

**Définition 1.11** *Un point  $X$  est strictement faisable pour le primal si il est faisable et satisfait  $X \succ 0$*

*Une paire  $(y, Z)$  est strictement faisable pour le dual si elle est faisable et satisfait  $Z \succ 0$ .*

**Théorème 1.5 (dualité forte)** *Supposons qu'il existe une solution strictement faisable  $(\hat{y}, \hat{Z})$  pour le dual. Soit :*

$$p^* = \inf \{ \langle C, X \rangle : \mathbf{A}X = b, X \succeq 0 \} \text{ et } d^* = \sup \{ \min \langle b, y \rangle, \mathbf{A}^T y + Z = C, Z \succeq 0 \}$$

*Alors  $p^* = d^*$  et si  $p^*$  est une valeur finie, alors il peut être atteint pour un  $X \in \{X \succeq 0 : \mathbf{A}X = b\}$ .*

**Corollaire 1.1** *Soit  $p^*$  et  $d^*$  définis comme dans le théorème précédent, alors :*

- *Si le primal est strictement faisable et  $p^*$  est fini alors  $p^* = d^*$  et cette valeur est atteinte par le dual.*
- *Si le dual est strictement faisable et  $d^*$  est fini alors  $p^* = d^*$  et cette valeur est atteinte par le primal.*
- *Si le primal et le dual sont tous deux strictement faisables, alors  $p^* = d^*$  et sont atteints par les deux problèmes.*

---

<sup>7</sup>Le produit scalaire considéré est le suivant :  $\langle A, B \rangle = \text{tr}(B^T A)$ . On peut alors remarquer qu'en fait  $\mathbf{A} = \mathcal{A}^T$



Des exemples montrant que toutes les hypothèses du théorème sont nécessaires pour assurer que le système est solvable se trouve dans [46].

**Exemple 1.2** *Considérons par exemple l'exemple suivant (le problème primal se trouve à gauche, le dual à droite) :*

$$\begin{array}{ll} \min x_{11} & \max 2y_1 \\ \text{tel que } \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0 & \text{tel que } \begin{bmatrix} 1 & -y_1 \\ -y_1 & 0 \end{bmatrix} \succeq 0 \end{array}$$

*Le problème primal à une solution strictement faisable ( $x_{11} = 2, x_{22} = 2$ ) mais on voit que le dual n'a pas de solution strictement faisable. En effet, sa seule solution est  $y_1 = 0$ .*

*La semi-définie positivité de la matrice implique  $x_{11} \geq 0, x_{22} \geq 0, x_{11}x_{22} - 1 \geq 0$ . D'où  $x_{11} \geq 1/x_{22}$ . Par le théorème de dualité, on sait que 0 est la valeur optimale du primal, mais elle ne peut être atteinte. On peut remarquer en effet que la limite inférieure de  $x_{11}$  tend vers 0 quand  $x_{22} \rightarrow +\infty$ .*

Dans [100], Ramana présente une autre expression de la dualité pour SDP qui d'obtenir le théorème de dualité forte sans avoir besoin de la stricte faisabilité (comme c'était le cas en programmation linéaire). Malheureusement, l'expression du dual qu'il obtient a  $\mathcal{O}(mn^2)$  variables, ce qui est réditatoire sur le plan calculatoire.

On peut remarquer que toutes les fonctions "représentables" en programmation conique quadratique le sont également en programmation semi-définie positive. En fait, l'ensemble des fonctions représentable en programmation SDP est sensiblement plus important que celles de la programmation CQ (conique quadratique). En particulier, la programmation SDP permet de manipuler les valeurs propres. Ainsi, la plus grande valeur propre d'une matrice, considérée comme une fonction d'une matrice symétrique  $m \times m$  est représentable en SDP. De la même façon, la somme des valeurs propres d'une matrice symétrique est représentable en SDP. D'autres exemples de fonctions représentables par la SDP se trouvent dans [89, 88].

Les applications de la SDP sont nombreuses. La SDP peut être utilisée pour résoudre des problèmes en temps polynomial ou pour fournir des approximations de solutions de façon tractable. Pour ce faire, on substitue un ensemble de contraintes par un ensemble plus large, qui lui, peut être résolu efficacement. C'est le principe de la relaxation. [89] propose quelques exemples tirés de la mécanique, l'électricité et la théorie des graphes.

[68] propose de nombreux exemples dans lesquels la relaxation SDP apporte de très bons résultats.

## 1.8.2 Petits rappels d'algèbre linéaire.

### Formes quadratiques

**Définition 1.12 (Formes bilinéaires)** *Une fonction  $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  est dite bilinéaire si :*

$$- \forall x, y, z \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \phi(x + \lambda y, z) = \phi(x, z) + \lambda \phi(y, z)$$

$$- \forall x, y, z \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \phi(x, y + \lambda z) = \phi(x, y) + \lambda \phi(x, z)$$

Si l'on note  $e_i$  le  $i^{\text{eme}}$  vecteur de la base canonique de  $\mathbb{R}^n$ , et écrivons :

$$A := [\phi(e_i, e_j)]_{1 \leq i, j \leq n}$$

nous avons alors la relation suivante :

$$\forall x, y \in \mathbb{R}^n, \phi(x, y) = x^t A y$$

**Définition 1.13 (Formes quadratiques)** Une fonction  $q : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  est quadratique si il existe une fonction bilinéaire  $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  telle que :

$$\forall x \in \mathbb{R}^n, q(x) = \phi(x, x).$$

**Proposition 1.3** Une fonction  $q : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  est quadratique si et seulement si :

- $\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, q(\lambda x) = \lambda^2 q(x)$ .
- $(x, y) \rightarrow \frac{1}{2}[q(x + y) - q(x) - q(y)]$  est bilinéaire.

Ainsi, une fonction quadratique peut aussi être écrite comme :  $q(x) = x^t A x$ . Alors si l'on note  $x$  le vecteur  $x = (x_1, \dots, x_n)^t$ , alors  $q(x)$  est un polynome homogène de degré 2 qui peut être noté :

$$q(x) = \sum_{i=1}^n \alpha_i x_i^2 + \sum_{1 \leq i < j \leq n} \alpha_{i,j} x_i x_j.$$

Alors  $A$  la matrice associée à  $q$  dans la base euclidienne a l'allure suivante :

$$A = \begin{pmatrix} \alpha_1 & & & & \\ & \dots & & & \\ & & \dots & & \\ & \frac{1}{2}\alpha_{i,j} & & \dots & \\ & & & & \alpha_n \end{pmatrix}.$$

**Définition 1.14** On dit qu'une fonction quadratique  $q, q : \mathbb{R}^n \rightarrow \mathbb{R}$  est positive si :

$$\forall x \in \mathbb{R}^n, q(x) \geq 0.$$

et l'on note  $q \succeq 0$  (et par extension  $A \succeq 0$  si  $A$  est la matrice symétrique représentant  $q$  dans une base).

Alors, nous écrivons  $q_1 \succeq q_2$  si  $q_1 - q_2 \succeq 0$ . Nous notons  $U_{psd}$  l'ensemble des matrices carrées  $n \times n$  sur  $\mathbb{R}$ .

**Définition 1.15** On dit qu'une fonction quadratique positive  $q : \mathbb{R}^n \rightarrow \mathbb{R}$  est définie (et on écrit  $q \succ 0$ ) est :

$$\forall x \in \mathbb{R}^n, q(x) = 0 \Rightarrow x = 0.$$

**Définition 1.16** Une fonction quadratique semi-définie est soit positive soit négative

## Matrices semi-définies positives

Des résultats plus complets d'algèbres linéaires se trouvent dans le premier chapitre du livre de Helmberg [46].

**Définition 1.17 (Espace vectoriel de matrices)** Soit  $M_{m,n}$  l'ensemble des matrices réelles  $m \times n$ . Il peut être interprété comme un espace de vecteur de  $R^{n+m}$ . Dans cet espace vectoriel, le produit scalaire entre deux éléments  $A$  et  $B$  de  $M_{m,n}$  est :  $\langle A, B \rangle = \text{tr}(B^T A) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{i,j}$ . En notant "tr" la trace d'une matrice, c'est-à-dire la somme de ses éléments diagonaux.

**Proposition 1.4 (Matrices symétriques réelles)** Une matrice symétrique a toutes ses valeurs propres<sup>8</sup> réelles et il existe une matrice orthonormée<sup>9</sup>  $P$  qui diagonalise  $A$ , c'est-à-dire telle que  $P^T A P = \Lambda_A$  (en notant  $\Lambda_A$  la matrice diagonale dont les éléments diagonaux sont les valeurs propres de  $A$ ).

**Définition 1.18** On dit que  $A \in S_n$  est semi-définie positive ( $A \in S_n^+, A \succeq 0$ ) si  $\forall x \in R^n, x^T A x \geq 0$ .

On dit que  $A \in S_n$  est définie positive ( $A \in S_n^{++}, A > 0$ ) si  $\forall x \in R^n, x^T A x > 0$ .

**Théorème 1.6 (Caractérisation des matrices définies positives)** Soit  $A$  une matrice symétrique ( $A \in S_n$ ). Les propositions suivantes sont équivalentes :

- $A$  est définie positive
- $\lambda_i(A) > 0$  pour  $i = 1, \dots, n$
- $\exists C \in M_n$  de rang  $n$  telle que  $A = C^T C$ .
- Pour n'importe quelle suite d'ensemble "nested"<sup>10</sup> de sous-matrices principales de  $A$ ,  $A_i \in S_i, i = 1, \dots, n$ , on a  $\forall i, \det(A_i) > 0$ .

**Théorème 1.7 (Caractérisation des matrices semi-définies positives)** Soit  $A$  une matrice symétrique ( $A \in S_n$ ). Les propositions suivantes sont équivalentes :

- $A$  est semi-définie positive
- $\lambda_i(A) \geq 0$  pour  $i = 1, \dots, n$
- $\exists C \in M_n$  telle que  $A = C^T C$ . Pour une telle matrice  $C$ ,  $\text{rang}(C) = \text{rang}(A)$ .

**Théorème 1.8 (Matrices symétriques de taille 2)** Soit  $a, b$  et  $c$  des nombres réels positifs. Alors la matrice

$$M = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix}$$

est semi-définie positive si et seulement si  $m_{1,1} m_{2,2} \geq M_{1,2}^2$ .

**Théorème 1.9 (Matrices symétriques de taille 1)** Une matrice de taille 1 est un scalaire. Elle est donc définie positive si et seulement si son unique élément est positif ou nul.

$$M = (m_{1,1}) \succeq 0 \Leftrightarrow m_{1,1} \geq 0$$

<sup>8</sup>Les valeurs propres sont les solutions du polynôme caractéristique de  $A$  : c'est  $\det(A - \lambda I)$ .

<sup>9</sup>c'est-à-dire dont les vecteurs colonnes sont 2 à 2 orthogonaux et de norme unitaire.

<sup>10</sup>Une suite "nested" est caractérisée par une suite de sous-ensembles d'indice  $J_i, 1 \leq i \leq n$  telle que  $J_1 \subset J_2 \subset \dots \subset J_n = \{1, 2, \dots, n\}$  avec  $\forall i, |J_i| = i$  (en notant  $|\cdot|$  le cardinal de l'ensemble)

**Théorème 1.10** Une matrice diagonale par bloc est semi-définie positive si et seulement si tous ses blocs sont semi-définis positifs.

### 1.8.3 Exemples classiques

Les exemples classiques de problèmes dans lesquels la SDP a proposé des solutions intéressantes sont souvent tirés de la théorie des graphes.

Un graphe est un ensemble  $(V, E)$  où  $V$  est un ensemble de noeuds et  $E$  un ensemble de liens. Souvent, on associe les liens à des valeurs, appelées poids.

Une coupe consiste à partitionner l'ensemble des noeuds en 2 ensembles  $V_1$  et  $V_2$ . Le poids de la coupe est la somme des poids des liens reliant un des noeud de l'ensemble  $V_1$  à un noeud de  $V_2$ . Le problème MAXCUT consiste à trouver un coupe de poids maximal. Au contraire du problème de la coupe minimale, qui est un problème LP, le problème de la coupe maximale est un problème NP-complet.

Notons  $a_{i,j}$  le poids d'un lien reliant le noeud  $i$  au noeud  $j$ . Une coupe étant un partitionnement des noeuds, peut être représentée par un vecteur  $x$  contenant des 1 et des  $-1$  selon que le lien  $i$  est dans  $V_1$  ou dans  $V_2$ . On montre aisément que le poids d'une coupe est :

$$\frac{1}{2} \left[ \frac{1}{2} \sum_{i,j=1}^n a_{i,j} - \frac{1}{2} \sum_{i,j=1}^n a_{i,j} x_i x_j \right] = \frac{1}{4} \sum_{i,j=1}^n a_{i,j} (1 - x_i x_j)$$

Après quelques manipulations on voit qu'une borne supérieure du problème de max-cut est donné par la résolution du problème SDP :

$$\begin{aligned} & \max \frac{1}{4} \sum_{i,j=1}^n a_{i,j} (1 - x_i x_j) \\ & \text{sous la contrainte } \begin{cases} X = [X_{ij}]_{i,j=1}^n = X^T \succeq 0 \\ X_{ii} = 1, i = 1, \dots, n \end{cases} \end{aligned}$$

Goemans et Williamson ont montrés en 1995 que si  $OPT$  est la valeur optimale du problème de MAXCUT et  $SDPOPT$  la valeur optimale donnée par cette relaxation SDP, alors :

$$OPT \leq SDP \leq \alpha OPT, \alpha = 1.138 \dots$$

D'autres relaxations SDP peuvent être utilisées pour donner des approximations du problème de Max-Cut ([46], [68]). L'une d'elle utilise notamment l'expression de la diagonale (c'est-à-dire le vecteur composé des éléments diagonaux d'une matrice). Elle s'écrit  $\max \langle C, X \rangle$  telle que  $\text{diag}(X) = e, X \succeq 0$ . Cette formulation matricielle montre la puissance de SDP.

#### Autres travaux sur la SDP

On peut citer les travaux faits sur la SDP de Anderson, Engbreetsen et Hastad ([8]) sur les équations linéaires modulo  $p$ .

La programmation SDP peut aussi être utilisée pour faire de la programmation entière et offrir une bonne approximation de problèmes de type

$$\max_x \{c^T x \mid Ax \leq b, x \in \{0, 1\}^n\}.$$

L'idée repose sur des opérations sur la diagonale. Un exemple classique est celui du sac à dos entier. De nombreux exemples détaillés se trouvent dans [46] et [68].

Des articles proposent la SDP pour offrir des relaxations intéressantes de problèmes classique NP-complets [80, 65, 8].

Enfin, les méthodes algorithmiques de résolution de la SDP sont variées et ont données naissance à plusieurs solveurs, dont certains, comme `csdp` sont publiques. Les méthodes qu'ils utilisent reposent souvent sur des méthodes de point intérieur ou de coupe par des plan. Elles sont passées en revue dans [46].

## 1.9 Conclusion

Nous avons présenté dans cette partie une famille de critères d'équité, paramétrés par un réel  $\alpha$  permettant à un opérateur de réseau de choisir quel équilibre choisir entre équité et rentabilité. Nous avons présenté les algorithmes classiques utilisés jusqu'à présent et avons introduit, en nous basant sur les modèles de marchandages issus de la théorie des jeux coopératifs, une nouvelle définition de cette famille de critère d'équité qui prend en compte la nature des besoins des applications temps-réels utilisées dans les réseaux (équation 1.8). Plus précisément, à chaque connexion est associée une fonction d'utilité représentant la qualité perçue par l'utilisateur. Nous avons proposé d'approcher les utilités par des fonctions quadratiques définies par 3 paramètres. Nous avons alors pu observer numériquement leur impact sur les allocations équitables.

Cette approche permet d'ouvrir de nouvelles voies de recherche. Tout d'abord, il serait intéressant de mieux comprendre comment l'équité peut être adaptée pour être utilisée par les sessions multicast, et quantifier, notamment avec une mesure d'équité associé à chaque critère d'équité, la qualité de la session multicast.

Une autre direction intéressante pourrait être celle de la tarification. Nous avons vu que l'on pouvait obtenir un équilibre de NBS en introduisant dans le réseau des taxes. C'est un exemple typique dans lequel un ensemble d'optimisations locales mènent a une solution globale. Cette propriété reste valable même si la fonction d'utilité d'un utilisateur est non connue du réseau et différente pour chaque connexion. La seule contrainte sur cette fonction est que l'utilité est croissante, concave et différentiable ([61]).

Nous nous intéressons dans cette étude à des algorithmes performants de résolution centralisés. Les chapitres qui suivent viendront donc proposer, pour chaque type de réseau, terrestres, mobiles et satellite, des solutions adaptées. Pour les deux premiers type, notre solution sera basée sur une programmation Semi-Définie Positive.

Un intérêt des solutions centralisées est leur application dans des problèmes de dimensionnement ou de compréhension d'interaction entre protocoles. Par exemple, on craint à l'heure actuelle de voir arriver des protocoles plus agressifs que TCP. [38] met en garde contre le fait que de tels protocoles, vendus pour être des "fast TCP" par exemple ne voient le jour. Une résolution centralisée pourrait alors permettre de mieux comprendre comment ces sources influeraient sur les autres connexions TCP.

**Première partie**  
**Les réseaux terrestres**



# *Introduction*

Nous nous plaçons ici dans le cadre des réseaux terrestres. Un tel réseau est constitué d'un ensemble de liens. Il s'agit la plupart du temps des paires de fils de cuivre torsadées ou de fibre optique. Quel que soit le media utilisé, il est, entre autre, caractérisé par sa bande passante. C'est cette ressource que nous allons chercher à allouer équitablement entre les connexions.

Une des caractéristiques de ce type de réseau est que le chemin suivi par les paquets n'est en général pas unique, c'est ce que l'on appelle le routage. Nous prendrons en compte dans cette partie deux stratégies de routage possibles.

Les réseaux terrestres sont ceux où le problème d'optimisation est le plus simple à formuler pour deux raisons. D'une part, une seule ressource est partagée (la bande passante). De plus, comme nous le verrons dans la suite de cette partie, le système de contraintes peut s'écrire sous forme d'une suite d'inégalités.

Dans le prochain chapitre, nous donnons une formulation possible du problème d'allocation équitable. Le deuxième chapitre a pour but de donner des méthodes analytiques de résolution pour des systèmes élémentaires. Enfin nous proposons dans un troisième chapitre une méthode de résolution centralisée basée sur la programmation Semi-Définie Positive et montrons quelques applications numériques.





# Chapitre 2

## Introduction aux réseaux terrestres.

Dans ce chapitre, nous appliquons les critères d'équités vus dans le premier chapitre aux réseaux terrestres. Dans la première section, nous présentons les notations et contraintes du modèle, tandis que dans le second, nous explicitons le problème d'optimisation.

### 2.1 Système de contraintes

Dans le cadre des réseaux terrestres, nous nous intéressons à deux types de contraintes : celles liées aux demandes des utilisateurs (sous-section 2.1.2) et celles liées au système lui-même et au type de routage (sous-section 2.1.3). Nous commençons par introduire, dans la sous-section suivante les notations relatives à notre modèle.

#### 2.1.1 Notations

Je rappelle que nous nous intéressons ici au cas d'un réseau composé d'un ensemble  $L$  de liens. Chaque lien  $l \in L$  est caractérisé par sa bande passante associée. Je note également  $C$  le vecteur de taille  $|L|$  tel que pour tout lien  $l$ ,  $c_l$  représente la capacité du lien.

Sur ce réseau circulent un ensemble  $\mathcal{N}$  de connexions. Une connexion désigne une communication entre deux points. Enfin, on suppose que chaque connexion est associée à une fonction d'utilité  $y$  et une demande minimale et maximale ( $MR_n$  et  $PR_n$  respectivement).

Ces notations sont résumées dans le tableau suivant :

$ ens $	le cardinal de l'ensemble $ens$ ,
$n$	une connexion,
$\mathcal{N}$	l'ensemble des connexions,
$x_n$	la bande passante allouée à la connexion $n$ ,
$x$	le vecteur des allocations (de taille $ \mathcal{N} $ , le nombre de connexions du réseau),
$MR_n$	demande minimale en bande passante de la connexion $n$ ,
$PR_n$	demande maximale en bande passante de la connexion $n$ ,
$l$	un lien,
$L$	l'ensemble des liens du réseau,
$c_l$	la capacité du lien $l$ ,
$C$	le vecteur de capacité des liens.

Je rappelle que notre problème sera alors de chercher la bande passante du réseau entre les différentes connexions tout en respectant un ensemble de contraintes. Je les passe en revue dans les sous-sections suivantes.

### 2.1.2 Demandes des utilisateurs

Les contraintes les plus évidentes sont les contraintes liées aux demandes des utilisateurs. On suppose que chaque connexion  $n$  a une demande minimale  $MR_n$  et une demande maximale  $PR_n$ . Ces contraintes s'écrivent donc :

$$\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n.$$

Elles constituent alors un système de  $|\mathcal{N}|$  inégalités linéaires indépendantes. Le deuxième type de contraintes dépend du type de routage choisi.

### 2.1.3 Types de routage

Une des particularités de ces types de réseaux est que le chemin suivi par une connexion pour aller d'une source à une destination n'est en général pas unique. Le choix d'une telle route constitue ce que l'on appelle *le routage*.

Il y a alors 2 types de routages possibles (correspondants donc à 2 types de contraintes) :

- Soit la route de chaque connexion est prédéfinie (et contenue dans une table, appelée matrice de routage) : c'est le *routage fixe*.
- Soit l'on suppose que le trafic de chaque connexion peut être scindé. Chaque portion de trafic d'une connexion peut alors suivre une route qui lui est propre : c'est le *routage fractionnaire*.

#### 2.1.3.1 Le routage fixe

Dans le cas du routage fixé, on définit une matrice qui spécifie les liens que vont traverser les paquets de chaque connexion, c'est la *matrice de routage*. Elle est de taille  $|L| \times |\mathcal{N}|$ . Si  $A$  est une telle matrice, alors elle est telle que :

$$A_{n,l} = \begin{cases} 1 & \text{si la connexion } n \text{ utilise le chemin } l, \\ 0 & \text{sinon.} \end{cases}$$

Dans certains cas, il se peut que l'on ne souhaite pas saturer les liens du réseau mais conserver une marge, appelée *capacité résiduelle*. Par exemple un opérateur peut souhaiter n'avoir un taux d'occupation de ses liens que de 90% afin de permettre à un trafic supplémentaire éventuel de circuler.

Ainsi, on peut supposer que l'on ne s'autorise à utiliser que la quantité  $(1 - \kappa_l)c_l$  sur chaque lien  $l$ . Soit  $\kappa$  le vecteur de taille  $|L|$  des quantités résiduelles.

La contrainte de capacité s'écrit alors :

$$Ax \leq (1 - \kappa)C.$$

### 2.1.3.2 Le routage fractionnaire

Dans le cas du routage fractionnaire, chaque paquet émis par une source peut suivre un chemin qui lui est propre. On considère alors le réseau comme un ensemble de noeuds reliés entre eux par des liens. Le but est alors de savoir, pour chaque paire de noeuds du réseau  $(u, v)$  le flux associé à chaque connexion, c'est-à-dire la fraction de la bande passante de la connexion qui utilisera le lien reliant  $u$  à  $v$ .

#### Notations :

Si un lien relie les noeuds  $u$  et  $v$ , on note  $\phi_n^{u,v}$  le flux de la connexion  $n$  de  $u$  vers  $v$  et  $\phi_n^{v,u}$  de  $v$  vers  $u$ .

Je note  $N(u)$  l'ensemble des noeuds connectés au noeud  $u$ .

Pour une connexion  $n$ ,  $s(n)$  (respectivement  $d(n)$ ) désigne sa source (respectivement sa destination).

Nous avons six types de contraintes pour chaque connexion  $n$  :

1. La somme des flux sortant de la source  $s(n)$  est égal à  $x_n$  :

$$x_n = \sum_{v \in N(s(n))} \phi_n^{(s(n),v)}.$$

2. La somme des flux arrivant à destination  $d(n)$  est égal à  $x_n$  :

$$x_n = \sum_{u \in N(d(n))} \phi_n^{(u,d(n))}.$$

3. A chacun des autres noeuds, on a la loi de conservation, où loi de Kirchhoff, c'est-à-dire que la somme des flux entrants est égal à la somme des flux sortants :

$$\forall u, u \neq s(n), u \neq d(n), \sum_{v \in N(u)} \phi_n^{(v,u)} = \sum_{v \in N(u)} \phi_n^{(u,v)}.$$

4. Aucun flux n'arrive au noeud source :

$$\forall u \in N(s(n)), \phi_n^{(u,s(n))} = 0.$$

5. Aucun flux ne sort du noeud  $d(n)$  la destination :

$$\forall v \in N(d(n)), \phi_n^{(d(n),v)} = 0.$$

6. Tous les flux sont positifs ou nuls :

$$\forall u, v, \phi_n^{u,v} \geq 0.$$

On peut alors résumer ces équations. Posons  $r_n^{s(n)} = -r_n^{d(n)} = 1$ , et  $r_n^u = 0$  pour  $u$  différent de  $s(n)$  et  $d(n)$ . Soit également  $V$  le nombre de noeuds du réseau. Les contraintes liées au routage fractionnaire s'écrivent :

$$\forall u, v \in [1, V], n \in \mathcal{N}, \begin{cases} x_n r_n^u = \sum_{w \in N(u)} \phi_n^{(u,w)} - \sum_{w \in N(u)} \phi_n^{(w,u)}, \\ \phi_n^{u,v} \geq 0, \\ 0 = \sum_{u \in N(s)} \phi_n^{(u,s(n))} \\ 0 = \sum_{u \in D(n)} \phi_n^{(d(n),u)}. \end{cases}$$

On note alors  $\phi_n^l, l \in L$ , le flot de la connexion  $n$  sur le lien  $l$  (c'est-à-dire la quantité de bande passante qu'elle utilise). Si le lien  $l$  lie les nœuds  $u$  et  $v$  :

$$\phi_n^l = |\phi_n^{(u,v)} - \phi_n^{(v,u)}|$$

Si l'on s'autorise à utiliser la quantité  $(1 - \kappa_l)C_l$ , sur chaque lien  $l$ , alors les contraintes de capacité s'écrivent :

$$\forall l \in L, \quad \sum_{n \in \mathcal{N}} \phi_n^l \leq (1 - \kappa_l)C_l.$$

D'où finalement :

$$\forall n \in [1, N], \quad \left\{ \begin{array}{l} \forall u \in [1, V], \quad x_n r_n^u = \sum_{w \in N(u)} \phi_n^{(u,w)} - \sum_{w \in N(u)} \phi_n^{(w,u)}, \\ \forall v, v \in [1, V], \quad \phi_n^{u,v} \geq 0, \\ 0 = \sum_{u \in N(s)} \phi_n^{(u,s(n))} \\ 0 = \sum_{u \in N(d)} \phi_n^{(d(n),u)}. \\ \forall l \in L, \quad \phi_n^l = |\phi_n^{(u,v)} - \phi_n^{(v,u)}| \\ \forall l \in L, \quad \sum_{n \in \mathcal{N}} \phi_n^l \leq (1 - \kappa_l)C_l. \end{array} \right.$$

## 2.2 Types d'optimisations

Nous considérons ici 2 types d'optimisations appelées *connexion-aware* et *network-aware*. L'idée dans le premier cas est d'optimiser l'allocation des connexions et dans la seconde de préserver au mieux le réseau.

### 2.2.1 L'optimisation "connexion-aware"

On associe à chaque connexion une fonction d'utilité dépendant de l'application de l'utilisateur, comme expliqué dans la section 1.6. L'idée de cette optimisation est alors d'allouer équitablement l'excédent de bande passante (une fois les demandes minimales assouvies) entre les connexions.

Le routage utilisé peut ici être fixe ou fractionnaire et le vecteur  $\kappa$  est une donnée du problème. Le problème d'optimisation est alors de résoudre :

$$\max \sum_{n \in \mathcal{N}} \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$$

avec les contraintes du système vus à la section précédente.

### 2.2.2 L'optimisation "network-aware"

Dans ce cas, on applique les concepts d'équité à la quantité de bande passante restant libre dans les liens. En effet, il peut être judicieux de maximiser la bande passante résiduelle afin de l'utiliser par exemple pour d'autres types de trafic de best-effort, ou pour constituer

un ensemble de ressources libres pour d'éventuelles autres arrivées de connexions (ceci peut être particulièrement intéressant si, en pratique, une fois un routage pour une connexion déterminé, on ne peut dynamiquement le modifier).

Ainsi, ici le vecteur des allocations aux connexion  $x$  est une donnée du problème et le vecteur d'inconnues est  $\kappa$ . Le problème d'optimisation s'écrit alors :

$$\max \sum_{l \in L} \frac{f_l(\kappa_l)^{1-\alpha}}{1-\alpha}.$$

Et les contraintes sont celles exhibées dans la section précédente.

On remarque que, le vecteur d'allocation étant donné, le problème d'optimisation n'a d'intérêt que lorsque l'on se trouve dans le cas d'un routage fractionnaire. En effet, si le routage est fixé ainsi que les bandes passantes associées aux connexions, alors la quantité disponible dans chaque lien est définie et on ne peut pas l'optimiser.

### 2.2.3 Résumé et conclusion

Alors qu'on avait a priori 4 schémas d'optimisations possibles. En fait, on a vu qu'il n'y a aucun intérêt, dans un système à routage fixé, de proposer une optimisation network-aware. Nous allons donc dans la suite étudier les 3 autres schémas résumés dans le tableau 2.1.

	connexion-aware	network-aware
routage fixé	oui	oui
routage fractionnaire	oui	non

FIG. 2.1 – Les trois schémas d'optimisation.

On rappelle que le débit de chaque connexion est la quantité que l'on alloue équitablement dans le cas de l'optimisation "connexion-aware". Dans le cas d'optimisation network-aware, ce débit est fixé à la valeur minimale requise par les utilisateurs (c'est-à-dire  $MR_i$ ) et on s'intéresse alors au partage équitable des bandes passantes résiduelles dans les liens  $l$ , qui sont notés  $\kappa_l$ .

Dans les deux cas, on note  $f_i$  les fonctions d'utilité associées aux quantités que l'on veut allouer équitablement. Notre problème est donc de résoudre le problème d'optimisation suivant :

$$\left\{ \begin{array}{l} \max \sum_{n \in \mathcal{N}} \frac{f_n(x_n)^{1-\alpha}}{1-\alpha} \quad \text{optimisation connexion-aware} \\ \max \sum_{l \in L} \frac{f_l(\kappa_l)^{1-\alpha}}{1-\alpha} \quad \text{optimisation network-aware} \end{array} \right. \quad (2.1)$$

## 2.3 Conclusion

Nous avons explicité dans ce chapitre 3 schémas d'optimisation que nous nous proposons d'étudier. Nous faisons dans toute la suite de notre étude des réseaux terrestres l'hypothèse suivante :

**Hypothèse 2.1** *Le réseau à une capacité suffisante pour satisfaire toutes les demandes minimales.*

Les demandes représentent les valeurs minimales en bande passante des utilisateurs dans le cas de l'optimisation "connexion-aware", ou les valeurs minimales de bandes passante résiduelles dans l'optimisation "network-aware".

Nous présentons dans le chapitre suivant quelques exemples de solutions analytiques dans des situations rendues simples, soit par l'architecture du réseau soit par la forme de l'optimisation choisie.

# Chapitre 3

## Méthodes analytiques

Ce chapitre se compose de deux sections indépendantes. Dans la première, nous proposons de résoudre le problème d'allocation équitable de la bande passante dans les réseaux terrestres par calculs explicites des taux. Cependant cette approche n'est possible que dans des cas relativement simples, de par la topologie du réseau et la forme des fonctions d'utilités choisies. Nous proposons alors dans la seconde section une méthode basée sur une approche Lagrangienne et dérivée des travaux de [117]. L'algorithme est itératif et permet d'obtenir une résolution décentralisée du problème d'équité dans des réseaux de topologie quelconque et avec des fonctions d'utilité générales quadratiques, mais uniquement dans le cas du NBS (c'est-à-dire le cas où  $\alpha = 1$ ).

### 3.1 Exemples de calculs explicites

On peut obtenir une formulation explicite des allocations équitables dans le cas de réseaux de petites tailles (c'est-à-dire de l'ordre de quelques liens) ou de quelques réseaux généraux (comme le réseau cyclique ou le réseau linéaires, déjà vus précédemment et représentés dans les figures 1.2 et 1.3 respectivement).

Nous nous penchons dans cette section sur le réseau linéaire et sa généralisation, le réseau grille.

#### 3.1.1 Les allocations équitables dans le réseau linéaire

On considère un réseau linéaire comme représenté à la figure 3.1. On rappelle qu'un tel réseau se compose de  $N$  nœuds, numérotés de 1 à  $N$  et de  $N + 1$  connexions.

Les paires source-destinations des connexions sont résumées dans le tableau suivant :

Connexion	Nœud source - destination
0	1 - $n$
$i, i \neq 0$	$i$ - $i + 1$

Le réseau se compose de  $N$  liens, chaque lien  $l$  joignant les nœuds  $l$  à  $l + 1$ .

Ainsi, la connexion  $n$  (avec  $n \neq 0$ ) utilise nécessairement, de par la topologie du réseau, au moins une fois le lien liant les nœuds  $n$  à  $n + 1$ . La connexion 0 doit traverser tous les liens du réseau.



On voit en fait que ces contraintes sont suffisantes. Ainsi il existe pour ce réseau un routage évident dans lequel chaque connexion  $n, n \neq 0$  utilise uniquement le lien  $n$  et dans lequel la connexion 0 utilise tous les liens du réseau une et une seule fois.

Tous les autres routages possibles sont alors sous optimaux au niveau du réseau car ils utilisent deux fois le même lien pour au moins une connexion. On ne s'intéresse donc pas, en général, à des possibles routages fractionnaires dans les réseaux linéaires.

On s'intéresse dans la suite de cette sous-section à l'optimisation connexion-aware, avec des contraintes de routage fixé.

### 3.1.1.1 Le réseau et les hypothèses

On considère le réseau représenté à la figure 3.1. On rappelle que dans cette représentation les liens physiques sont schématisés par des rectangles, les connexions logiques par des segments entre ces rectangles. Ainsi, on peut voir sur l'exemple un réseau de 3 liens physiques et 4 connexions.

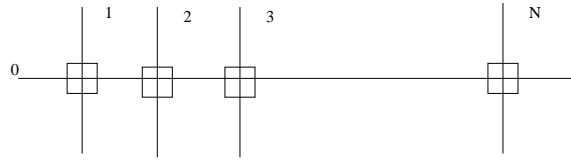


FIG. 3.1 – Un réseau linéaire

Les différents critères d'équités s'obtiennent par la maximisation de :

$$\begin{cases} \frac{1}{1-\alpha} \sum_{n=0}^{\mathcal{N}} (f_n(x_n))^{1-\alpha} & \text{si } \alpha \neq 1, \\ \prod_{n=0}^{\mathcal{N}} f_n(x_n) & \text{sinon.} \end{cases} \quad (3.1)$$

Le système de contraintes s'écrit :

$$\begin{cases} \forall n \in [0; \mathcal{N}], MR_n \leq x_n \leq PR_n & \text{(contraintes des utilisateurs),} \\ \forall n \in [1; \mathcal{N}], x_0 + x_n \leq c_n & \text{(contraintes de capacité).} \end{cases} \quad (3.2)$$

On voit ainsi que les  $x_n, n \neq 0$  sont 2 à 2 indépendantes, et que l'on peut donc saturer tous les liens :

$$\forall n \in [1; \mathcal{N}], x_0 + x_n = c_n.$$

Les fonctions d'utilités sont les fonctions paraboliques définies dans la sous-section 1.6.4. Nous utilisons ici la formulation utilisant les paramètres  $a, b$  et  $c$ .

Nous faisons alors deux hypothèses :

**Hypothèse 3.1** *Les capacités des liens sont identiques et valent  $c$ .*

**Hypothèse 3.2** *Les fonctions d'utilité des connexions  $n, n \neq 0$  sont les mêmes. Ainsi  $\forall n \in 1..\mathcal{N}, a_n = a_1, b_n = b_1, c_n = c_1$ .*

Ces deux hypothèses nous permettent d'affirmer qu'alors, au point d'équilibre, toutes les connexions  $n$ ,  $n \neq 0$  se verront attribuer la même bande passante. Ainsi, si l'on note  $x$  la bande passante affectée à la connexion 0, le terme à maximiser devient celui de l'équation 3.3 sous les contraintes définies dans le système 3.2.

$$\begin{cases} g(x) = \frac{1}{1-\alpha} [(f_0(x))^{1-\alpha} + N(f_1(c-x))^{1-\alpha}] & \text{si } \alpha \neq 1 \\ h(x) = f_0(x)(f_1(c-x))^N & \text{sinon} \end{cases} \quad (3.3)$$

**Remarque 3.1** *Le problème n'a de sens que si :*

$$\begin{cases} MR_0 + MR_1 \leq C & \text{sinon le système n'est pas faisable} \\ PR_0 + PR_1 > C & \text{sinon on peut affecter les demandes maximales} \end{cases} \quad (3.4)$$

### 3.1.1.2 Solution du problème linéaire

En différenciant 3.3, on obtient :

$$\begin{cases} a_0(x-b_0)f_1(c-x)^\alpha = Na_1(c-x-b_1)(f_0(x))^\alpha & \text{si } \alpha \neq 1 \\ a_0(x-b_0)(f_1(c-x))^N = Na_1(c-x-b_1)(f_0(x))(f_1(c-x))^{N-1} & \text{sinon} \end{cases}$$

En fait, on remarque que, pour  $\alpha = 1$ , si  $f_1(c-x) = 0$  alors le terme à maximiser (c'est-à-dire  $h(x)$ ) devient nul. Ainsi,  $f_1(c-x) \neq 0$  pour  $\alpha = 1$  et la bande passante  $x$  associée à la connexion 0 à l'équilibre vérifie pour toute valeur de  $\alpha$  :

$$a_0(x-b_0)(f_1(c-x))^\alpha = Na_1(c-x-b_1)(f_0(x))^\alpha \quad (3.5)$$

### 3.1.1.3 Limites possibles

Nous nous intéressons ici aux limites éventuelles  $x_{lim}$  de la bande passante lorsque  $\mathcal{N}$  tend vers l'infini. On peut montrer que :

**Lemme 3.1** *Quand  $\mathcal{N}$  tend vers l'infini, les seules limites possibles de l'allocation de la connexion 0 sont :*

$$MR_0 \text{ et } c - b_1.$$

**Démonstration :** Reprenons l'équation 3.3. Comme  $x$  est borné (car  $MR_0 \leq x \leq PR_0$ ), le membre de gauche de l'équation est borné également. Alors, la limite  $x_{lim}$ , si elle existe, vérifie :

$$a_1(c-x-b_1) \cdot f_0(x) = 0 \quad (3.6)$$

Cette équation admet clairement trois solutions. Nous cherchons alors à savoir si elles sont possibles, c'est-à-dire si elles se trouvent dans l'intervalle  $[MR_0, PR_0]$ . La première solution est  $x_{lim} = c - b_1$ . D'après 3.2 on a  $PR_0 + PR_1 > C$ . Or comme,  $b_1 > PR_1$  (d'après 1.9), alors on a bien  $C - b_1 < PR_0$ . Cette solution n'est donc acceptable que si  $C - b_1 \geq MR_0$ .

Les deux autres solutions sont les valeurs de  $x$  qui annulent  $f_0$ . L'une est  $MR_0$  (par définition) et l'autre est strictement supérieure à  $PR_0$ , ce qui est une solution non faisable.

Les deux solutions possibles pour la limite de l'allocation de la connexion 0 sont donc  $MR_0$  et  $C - b_1$ . ■

**Proposition 3.1** *Quand  $\mathcal{N}$  tend vers l'infini, la limite de l'allocation de la connexion 0 vaut :*

$$\min(MR_0, c - b_1).$$

**Démonstration :** Nous montrons tout d'abord que  $g(C - b_1) \geq g(MR_0)$  et  $h(C - b_1) \geq h(MR_0)$ .

Supposons que  $\alpha \neq 1$ , on a :

$$\begin{cases} g(MR_0) = \frac{L}{1-\alpha} f_1(C - MR_0)^{1-\alpha} \\ g(C - b_1) = \frac{1}{1-\alpha} [f_0(C - b_1)^{1-\alpha} + L(f_1(b_1))^{1-\alpha}] \end{cases}$$

Par définition  $b_1$  est l'abscisse de l'apogée de  $f_1$ . Ainsi  $f_1(b_1) \geq f_1(C - MR_0)$ . De plus, si  $C - b_1 \geq MR_0$  alors  $f_0(C - b_1) \geq 0$ . Ainsi, on peut aisément vérifier que, quel que soit le signe de  $1 - \alpha$  on a :

Si  $C - b_1 \geq MR_0$  alors  $g(C - b_1) \geq g(MR_0)$ .

De même, si  $\alpha = 1$ ,

$$\begin{cases} h(MR_0) = 0 \text{ et} \\ h(C - b_1) = f_0(C - b_1)(f_1(b_1))^{\mathcal{N}} > 0. \end{cases}$$

Ainsi, quelle que soit la valeur de  $\alpha$  :

**Si  $C - b_1 \geq MR_0$  alors  $g(C - b_1) \geq g(MR_0)$  et  $h(C - b_1) \geq h(MR_0)$ .**

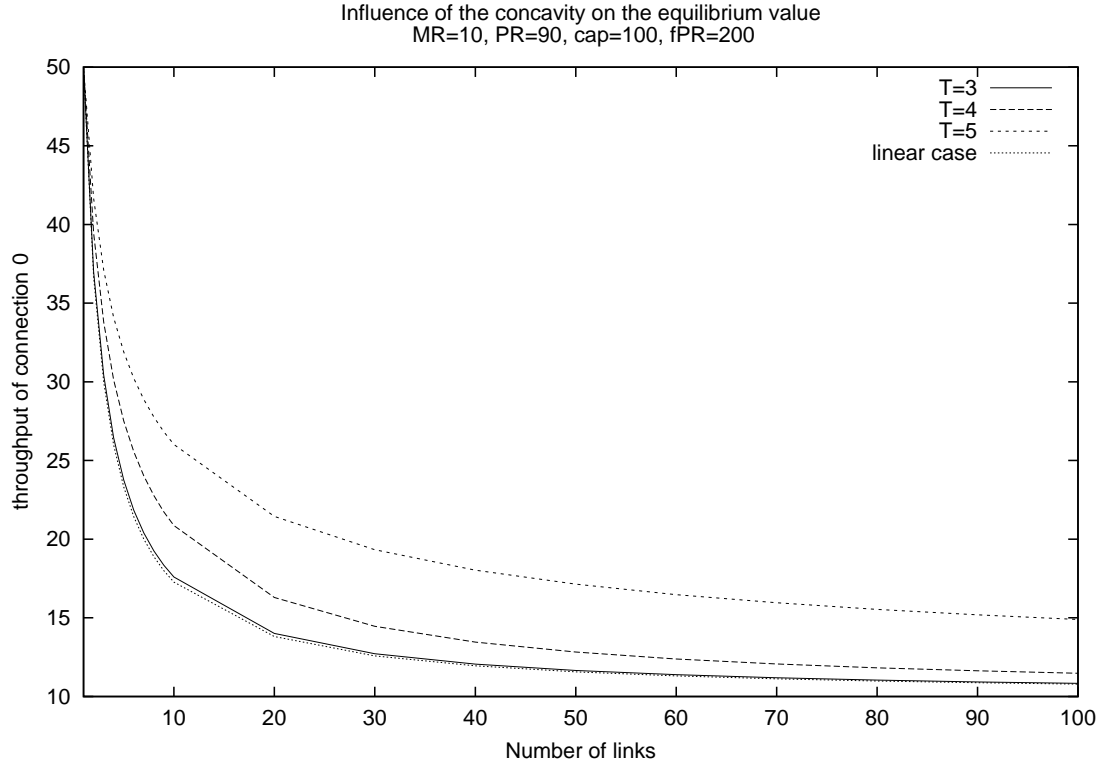
En fait on peut montrer que cette valeur est effectivement la limite de l'allocation. Supposons que  $\alpha = 1$ .

Si  $c - b_1 > MR_0$ , alors on sait que  $f_1(c - Id)$  est maximale en  $c - b_1$  (par définition de  $b_1$ ) et que  $f_0(c - b_1) > 0$  (car  $C - b_1 > MR_0$ ), donc quand  $\lim_{l \rightarrow \infty} \max f_0(x)(f_1(C - x))^l = c - b_1$ .

Considérons maintenant le cas où  $c - b_1 \leq MR_0$ .  $f_1(C - Id)$  est une parabole dont l'apogée se trouve en  $C - b_1 \leq MR_0$  et qui est nulle en  $MR_0$ . Ainsi la courbe est décroissante sur  $[b_1, \infty]$ , et donc également sur  $[c - MR_0, c - PR_0]$  (car  $c - b_1 > MR_0$ ). Donc son maximum se trouve en  $c - MR_0$ . De plus,  $f_0$  est positive sur l'intervalle  $[MR_0, PR_0]$ . Donc finalement  $x_{lim} = MR_0$ .

Les cas  $1 - \alpha > 0$  et  $1 - \alpha < 0$  se traitent de façon similaire. ■

Nous montrons dans la figure 3.1.1.3 comment le système converge vers la solution  $x_{lim}$  lorsque  $N$  tend vers l'infini dans le cas du NBS ( $\alpha = 1$ ). Dans cet exemple, on a  $c - b_1 \leq MR_0$ .



### 3.1.2 Analyse asymptotique

Nous reprenons dans cette sous-section l'analyse de la limite de  $x$  pour les grandes valeurs de  $L$  de façon plus précise.

**Proposition 3.2** *Si  $C - b_1 \leq MR_0$  alors  $x$  vérifie :*

$$x - MR_0 \sim Z \text{ avec } Z^\alpha \sim_{L \in \mathbb{L}} \frac{1}{L} \left( \frac{1}{2(b_0 - MR_0) \cdot a_0} \right)^{\alpha-1} \frac{f_1(c - MR_0)^\alpha}{f_1'(c - MR_0)}. \quad (3.7)$$

*Si  $C - b_1 \geq MR_0$  alors  $x$  vérifie :*

$$x = c - b_1 + Z + o(1/L) \text{ avec } Z = \frac{1}{L} \frac{c_1^\alpha f_0'(c - b_1)}{2a_1 f_0(c - b_1)^\alpha}. \quad (3.8)$$

**Démonstration :** Pour un  $L$  fixé et une valeur de  $\alpha$  donnée, je note  $x(L)$  l'allocation de la connexion 0. Reprenons alors l'équation 3.5.

Premier cas :  $c - b_1 \leq MR_0$

Quand  $L \rightarrow \infty$  le membre de gauche tend vers la constante :

$$\lim_{L \rightarrow \infty} a_0(x - b_0)(f_1(C - x))^\alpha = a_0(MR_0 - b_0)(f_1(C - MR_0)^\alpha).$$

(ceci est non nul car  $MR_0 - b_0 \neq 0$  et  $(f_1(C - MR_0)^\alpha) \neq 0$ )

Nous examinons maintenant le membre de droite de 3.5. Il peut être écrit :

$$\begin{aligned} La_1(\text{cap} - x - b_1)(f_0(x))^\alpha &= La_1(C - b_1 - MR_0 - z)(c_0 - a_0(MR_0 + z - b_0)^2)^\alpha \\ &= La_1(C - b_1 - MR_0 - z)(-2a_0(MR_0 - b_0)z - a_0z^2)^\alpha \\ &\sim_{L \rightarrow \infty} La_0^\alpha a_1 2^\alpha (c - b_1 - MR_0)(b_0 - MR_0)^\alpha z^\alpha \end{aligned}$$

ce qui conduit à l'équation (3.7) en substituant par les expressions appropriées.

### Deuxième cas : $MR_0 \leq C - b_1$

Supposons que  $MR_0 > C - b_1$ . Quand  $L \rightarrow \infty$  le membre de gauche tend vers la constante :

$$\lim_{L \rightarrow \infty} a_0(x - b_0)(f_1(C - x))^\alpha = a_0(C - b_1 - b_0)c_1^\alpha.$$

Le membre de droite est :

$$-La_1z \left( c_0 - a_0(C - b_1 - b_0 + z)^2 \right)$$

d'où le résultat de l'équation 3.8. ■

**Remarque 3.2** *On peut noter que le résultat ne dépend pas de  $T_0$  ni de  $T_1$  (invariant par multiplication par une constante).*

**Remarque 3.3** *On peut reprendre l'équation (3.8) afin d'obtenir les asymptotes correspondants aux cas limites (et dans le cas particulier où  $\alpha = 1$ ) :*

- Quand  $\beta_1 \rightarrow 1$  on a :  $Z = \frac{c - MR_0 - MR_1}{L}$  (cas linéaire, ce résultat a déjà été obtenu dans [117]).
- Quand  $\beta_1 \rightarrow 1/2$  on obtient :

$$Z = \frac{c - MR_0 - MR_1}{2L} \left[ 1 - \frac{PR_1 - MR_1}{\text{cap} - MR_0 - PR_1} \right].$$

#### 3.1.2.1 Conclusion

Nous avons étudié dans cette partie un exemple réseau linéaire avec 2 types de connexions. Dans le cas où  $\alpha = 1$  (NBS), on voit (équation 3.5) que le problème s'écrit comme une équation du troisième degré et peut donc être explicitement résolu. Dans le cas où  $\alpha = 0$  (optimisation globale), le polynôme à résoudre est de degré 1. Dans tous les autres cas, seules des résolutions numériques peuvent être utilisées. Néanmoins, on peut, quelquesoit la valeur de  $\alpha$ , expliciter les allocations limites et donner un équivalent de ces allocations lorsque le nombre de liens devient grand.

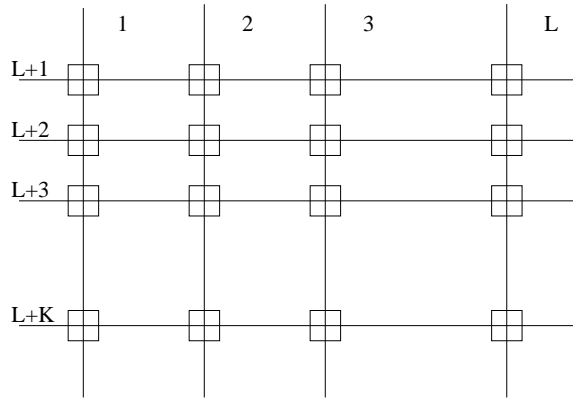


FIG. 3.2 – Un réseau grille.

### 3.1.3 Réseau grille

Ce réseau est la généralisation naturelle du réseau linéaire. Nous nous intéressons ici encore au cas du routage fixe, routage rendu évident par le choix des connexions.

Le réseau consiste en  $K \times L$  liens,  $K$  connexions horizontales et  $L$  verticales comme montrées dans la Figure 3.2.

Nous supposons que toutes les connexions horizontales (respectivement verticales) ont la même fonction d'utilité  $f_h$  (respectivement  $f_v$ ). Nous pouvons alors conclure facilement que toutes les connexions horizontales (respectivement verticales) auront la même bande passante  $x$  (respectivement  $x_v = c - x$ ).

Comme pour l'exemple précédent, nous supposons que pour chaque  $i \in \{1, \dots, L\}$ ,  $j \in \{1, \dots, K\}$ ,  $MR_i + MR_{L+j} < C_{i,L+j}$  et  $PR_i + PR_{L+j} > C_{i,L+j}$ .

Nous voulons alors maximiser dans le cas du NBS :

$$\prod_{i \in [0:L]} f_i(x_i) = (f_h(x))^K * (f_v(cap - x))^L. \quad (3.9)$$

**Proposition 3.3** Dans le réseau grille, si  $C - b_1 < MR_0$ ,  $x$  vérifie :  $x - MR_h \sim Z$  avec :

$$Z = \frac{K f_v(c - MR_0)}{L f'_v(c - MR_0)}.$$

Si non,  $x = c - b_1 + Z + o(K/L)$  avec

$$Z = \frac{K c_v^\alpha f'_h(c - b_v)}{L 2a_v f'_h(c - b_v)}.$$

Un cas particulier est  $L = K$  et  $f_h = f_v = Id$ . Dans ce cas, nous obtenons alors  $x = cap/2$ .

**Démonstration :** Ceci est similaire à maximiser :  $f_h(x) * (f_v(cap - x))^{L/K}$ . Dans ce cas, ce problème est équivalent au cas linéaire en substituant  $L/K$  à  $N$ . Le second résultat est évident. ■

## 3.2 Méthode Lagrangienne

La méthode de Lagrange a été proposée par [117] pour obtenir le NBS dans le cas particulier de fonctions d'utilités linéaires. Elle présente l'avantage de pouvoir être implémentée de façon distribuée. Nous généralisons cette approche au cas de l'utilité quadratique, pour laquelle le cas linéaire peut être retrouvé en prenant  $\beta \rightarrow 1$ . On peut remarquer cependant une limitation importante de la méthode du Lagrangien : il n'est pas possible de la généraliser pour les différentes valeurs de  $\alpha$ .

### 3.2.1 Multiplicateurs de Lagrange

Nous utilisons maintenant les conditions de Kuhn-Tucker pour obtenir une caractérisation alternative du NBS en terme des multiplicateurs de Lagrange correspondants.

**Proposition 3.4** *Sous l'hypothèse que  $\forall l \in \{1..L\}, \sum_{n \in \mathcal{N}} a_{ln} MR_n < C_l$  (ce qui revient à dire qu'aucun lien n'est saturé<sup>1</sup>), le NBS est caractérisé par :*

*$\forall l \in L, \exists \mu_l \geq 0$  tels que  $\forall n \in \{1..\mathcal{N}\}$ , nous avons*

$$x_n = \min \left( PR_n, MR_n + \left( \sum_{l=1}^L \mu_l a_{l,n} \right)^{-1} + \frac{1}{2} \cdot \frac{PR_n - MR_n}{1 - \beta_n} \times \left[ 1 - \sqrt{1 + \frac{4 \left( \frac{1 - \beta_n}{PR_n - MR_n} \right)^2}{\left( \sum_{l=1}^L \mu_l a_{l,n} \right)^2}} \right] \right).$$

**Démonstration :** Sous l'hypothèse  $\sum a_{ln} MR_n < C_l$ , l'ensemble  $A$  des solutions possibles est non-vidé, convexe et compact. Les contraintes sont linéaires en  $x_n$  et  $f(x) = \sum_{n \in \mathcal{N}} \ln f_n(x_n)$  est de classe  $C^1$ .<sup>2</sup> Par conséquent, les conditions de Kuhn-Tucker du premier ordre sont nécessaires et suffisantes pour garantir l'optimalité. Le Lagrangien associé au problème (2.1) est :

$$\mathcal{L}(x, \lambda, \delta, \mu) = -f(x) + \sum_{n \in \mathcal{N}} \lambda_n (MR_n - x_n) + \sum_{n=1 \in \mathcal{N}} \delta_n (x_n - PR_n) + \sum_{l \in L} \mu_l ((Ax)_l - C_l).$$

Pour  $l \in L$ ,  $\lambda_l \geq 0$  et  $\delta_n \geq 0$  sont les multiplicateurs de Lagrange associés aux contraintes des utilisateurs  $x_n \geq MR_n$  et  $x_n \leq PR_n$  respectivement.  $\mu_l \geq 0, l = 1, \dots, L$  sont les multiplicateurs de Lagrange associées aux contraintes de capacité.

Les conditions d'optimalité du premier ordre s'écrivent :

$$\begin{cases} \forall n \in \mathcal{N}, & 0 = (\lambda_n - \delta_n - \sum_{l \in L} \mu_l A_{l,n}) + \frac{f'_n(x_n)}{f_n(x_n)} \\ \forall n \in \mathcal{N}, & (x_n - MR_n) \lambda_n = 0, \\ \forall n \in \mathcal{N}, & (x_n - PR_n) \delta_n = 0, \\ \forall l, l \in L, & ((Ax)_l - C_l) \mu_l = 0 \end{cases}$$

<sup>1</sup>Cette hypothèse n'est pas limitatrice. Si un lien est saturé en considérant les demandes minimales des connexions, alors il ne reste qu'à réécrire le système en supprimant ce lien et les connexions qui le traversent. La bande passante que ces connexions utilisaient devient alors une nouvelle contrainte de capacité sur les autres liens du réseau.

<sup>2</sup>C'est-à-dire continue, dérivable et sa dérivée est également continue.

On note que  $\sum a_{ln}MR_n < C_l$  implique que  $\forall n, \lambda_n = 0$  comme dans [117].  
Enfin, ou bien l'on a  $x_n = PR_n$  ou bien  $\delta_n = 0$ , ce qui mène à la conclusion. ■

On peut noter que :

**Remarque 3.4** Quand  $\beta \rightarrow 1$  nous obtenons la solution de [117] correspondant à l'utilité linéaire :

$$x_i = \min \left( PR_i, MR_i + \left[ \sum_{l=1}^L \mu_l a_{l,i} \right]^{-1} \right).$$

$\mu_l$  représente le coût associé au lien  $l$  du réseau. C'est le coût marginal d'une unité de débit associé à une connexion quelconque traversant le lien  $l$ .

### 3.2.2 Problème dual

Une fois que nous avons explicité le NBS en termes de multiplicateurs de Lagrange, nous pouvons utiliser le problème dual associé afin de construire un algorithme itératif. Rappelons que  $x$  est le vecteur à optimiser,  $\mu$  est le vecteur de multiplicateurs de Lagrange et  $X$  l'ensemble des vecteurs d'allocations possibles.

Notre problème est de résoudre :

$$\min_x f(x)$$

sous les contraintes du système. Le problème dual s'écrit

$$\begin{aligned} & \max_{\mu \in \mathbb{R}_+^L} d(\mu) \\ & \text{avec } d(\mu) = \min_{x \in X} \mathcal{L}(x, \mu) = \mathcal{L}(\bar{x}_i, \mu) \end{aligned} \quad (3.10)$$

si nous notons  $\bar{x}_i$  la valeur optimale. Le vecteur  $\bar{x} = \overline{x_1, x_2 \dots x_n}$  est le NBS.

On peut utiliser le résultat obtenu dans le primal. Ainsi, pour un vecteur  $\mu$  donné, on pose pour chaque connexion  $n$  :

$$x_n(\mu) = g_n \left( \sum_{l \in L} \mu_l \cdot a_{l,n} \right). \quad (3.11)$$

Avec :

$$g_n(p) = \begin{cases} PR_n & \text{si } p \leq \frac{2\beta_n - 1}{\beta_n} \frac{1}{PR_n - MR_n}, \\ MR_n + \frac{1}{p} + \frac{1}{2} \frac{PR_n - MR_n}{1 - \beta_n} \left[ 1 - \sqrt{1 + \frac{4}{p^2} \cdot \left( \frac{1 - \beta_n}{PR_n - MR_n} \right)^2} \right] & \text{sinon.} \end{cases}$$

Nous obtenons pour chaque  $\mu \in \mathbb{R}^L$  :

$$d(\mu) = \sum_{n=1}^N -\ln(f_n(g_n(\sum_{l \in [1..L]} \mu_l a_{l,n}))) + \sum_{l \in L} \mu_l \sum_{n \in \mathcal{N}} a_{l,n} g_n \left( \sum_{l \in L} \mu_l a_{l,n} \right) - \sum_{l \in L} C_l \mu_l. \quad (3.12)$$



Comme dans [117], il n'y a pas de trou de dualité.

L'idée est alors de considérer une taille de pas judicieusement choisie et de poser à chaque itération :

$$\mu_l^{(k+1)} = \max \left( 0, \mu_l^{(k)} + \gamma \cdot (x(\mu^{(k)}) - C_l) \right).$$

On peut alors montrer de façon similaire à [117] que  $\lim_{k \rightarrow \infty} x(\mu^{(k)}) = \bar{x}$ .

### Cas linéaire

On voit que dans le cas linéaire la dérivée partielle de  $d$  peut s'écrire en fonction de  $x$  et des capacités  $C_l$ . On obtient ainsi ([117]) que :

$$\frac{\partial d}{\partial \mu_l} = \sum_{n=1}^N \ln(f_n(g_n(\sum_{l \in [1..L]} \mu_l a_{l,n}))) - C_l. \quad (3.13)$$

L'algorithme se compose alors naturellement en deux parties. D'un coté un *algorithme de liens* dans lequel chaque lien met à jour la valeur de son coût en fonction des bandes passantes des connexions qui le traversent et de sa capacité. De l'autre coté un *algorithme de nœuds* dans lequel chaque source met à jour son débit en fonction des coûts des liens qu'elle traverse et des paramètres de sa fonction d'utilité.

### Cas quadratique

La dérivée partielle de  $d$  ne peut plus alors s'écrire uniquement en fonction des valeurs de  $x$  et de  $C_l$ . Alors malheureusement, on peut toujours utiliser un algorithme itératif en deux parties mais dans ce cas les valeurs des paramètres des fonctions d'utilité des différents joueurs doivent être communiquées aux réseaux, ce qui limite fortement la portée de cette méthode.

## 3.3 Conclusion

Nous nous sommes intéressés dans ce chapitre, à des résolutions analytiques du problème de l'allocation équitable dans les réseaux terrestres. Dans une première section, nous avons vus des systèmes simples dans lesquels des solutions exactes ou des solutions asymptotiques pouvaient être explicitées. Nous avons vu dans une seconde partie comment les multiplicateurs de Lagrange pouvaient être utilisés afin de proposer une solution décentralisée du problème du NBS dans le cas uniquement de l'équité proportionnelle et de fonctions d'utilité linéaires et de réseaux quelconques. Enfin, dans le cas de fonctions d'utilités quadratique, on peut encore utiliser une méthode de relaxation lagrangienne, mais les paramètres des fonctions d'utilités de chaque connexion doit alors être transmise au réseau, ce qui limite grandement l'intérêt de cette méthode.

L'utilisation du problème dual du Lagrangien permet d'obtenir un algorithme itératif décentralisé. Au contraire, la nouvelle méthode de résolution basée sur la SDP que nous proposons dans le chapitre suivant propose une solution centralisée, ce qui peut être intéressant pour résoudre des problèmes de dimensionnement de réseau et de compréhension de l'interaction entre plusieurs protocoles utilisant des équités différentes. En outre, cette méthode n'est pas limitée au cas du NBS ni aux réseaux de petites tailles ou de topologies particulières et peut être également utilisée dans le cas de l'optimisation network-aware.



# Chapitre 4

## Méthode SDP

Comme nous l'avons vu dans la section 1.8, la Programmation Semi-Definie Positive est un algorithme de résolution général de problèmes d'optimisations non-linéaires. Nous montrons dans cette partie comment elle peut être utilisée pour résoudre le problème d'allocation équitable dans les réseaux terrestres.

Ce chapitre est construit de la façon suivante : dans la première section, nous décrivons la philosophie générale de l'approche SDP et son adaptation à notre problème. Dans une deuxième section, nous explicitons les particularités des matrices associées aux différentes valeurs possibles de  $\alpha$ . Nous détaillons alors dans la Section 4 un exemple simple de réseau et le problème SDP associé. Nous présentons enfin dans la dernière section quelques résultats numériques issus de deux réseaux différents dans le cas de l'optimisation connexion-aware et network-aware.

### 4.1 Philosophie générale de l'approche SDP

L'idée de base utilisée ici est de transformer notre problème original de maximisation de *fonction* en un problème de minimisation d'une *variable* sous contrainte de la positivité d'une matrice générale. Nous utilisons la positivité (psd) de cette matrice pour :

- Assurer les contraintes du système (de capacité, des utilisateurs et éventuellement celles liées au routage fractionnaire),
- Remplacer la fonction objectif (de la formule 2.1) par une seule variable.

On sait qu'une matrice par bloc<sup>1</sup> est positive si et seulement tous ces blocs sont positifs.

Nous utilisons cette propriété pour construire notre matrice SDP. Nous verrons que les contraintes de notre problème peuvent être exprimées par des matrices de taille unitaire (c'est-à-dire des scalaires) et que notre fonction objectif peut être exprimée par une suite de matrice de taille  $2 \times 2$ .

---

<sup>1</sup>Une matrice par bloc est de la forme

$$A = \begin{pmatrix} \boxed{A_1} & & 0 \\ & \boxed{A_2} & \\ 0 & & \dots \\ & & & \boxed{A_n} \end{pmatrix} \quad (4.1)$$

Il suffira alors de mettre bout à bout les blocs obtenus pour obtenir notre matrice finale. On peut noter que l'ordre des blocs dans la matrice n'est pas important.

Les entrées que nous donnerons alors au solveur SDP seront la matrice SDP ainsi construite et un vecteur de variables. Dans notre cas ce vecteur se composera de "0" à toutes les entrées et d'un "-1" à l'entrée correspondante à la variable finale que nous voulons optimiser.

Il est à noter que le nombre de variables ne peut être connu a priori. En effet, la création de la matrice et des blocs induit la création de plusieurs variables intermédiaires que nous précisons lors des sections consacrées.

Dans les deux sous-sections suivantes nous voyons quels sont les blocs exprimant les contraintes du système et celles permettant d'intégrer les fonctions d'utilité des utilisateurs.

### 4.1.1 Contraintes linéaires

Comme nous l'avons vu dans la section 2.1.3, que ce soit dans le cas du routage fixe ou du routage fractionnaire, les contraintes du système sont linéaires. On a vu dans la section 1.8 que de telles contraintes pouvaient être exprimées par des matrices de taille 1 (matrices scalaires).

J'explique dans cette sous-section les contraintes relative au système de routage fixe (le cas du routage fractionnaire est similaire).

#### Contraintes de capacité.

Elles sont données par la positivité de  $L$  matrices, correspondants aux  $L$  liens. En effet, on rappelle que ces contraintes s'écrivent  $\forall l \in L, (Ax)_l \leq C_l$ .

Il nous suffit alors d'introduire  $L$  matrices scalaires  $capa^l, l \in L$  dont l'unique élément  $capa_{1,1}^l$  vaut  $C_l - (Ax)_l$  pour tout  $l \in L$ .

#### Contraintes des utilisateurs.

Elles sont données par :  $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$ . On introduit ainsi  $2 \times \mathcal{N}$  matrices scalaires  $util$  telles que

$$util_{1,1}^i = \begin{cases} PR_i - x_i & \text{si } i \leq \mathcal{N} \\ x_i - MR_i & \text{si } \mathcal{N} < i \leq 2\mathcal{N} \end{cases} \quad (4.2)$$

### 4.1.2 Relations entre les variables

Notre but dans cette section est d'introduire des variables intermédiaires permettant de prendre en compte les contraintes liées aux fonctions d'utilités.

On rappelle tout d'abord que :

**Proposition 4.1** Soient  $a, b$ , et  $c$  des réels positifs. Alors

$$\begin{pmatrix} a & c \\ c & b \end{pmatrix} \succeq 0 \text{ si et seulement si } ab \geq c^2.$$

Les variables initiales de notre problème sont les bandes passantes allouées aux connexions  $x_n, n \in \mathcal{N}$ . On introduit les variables  $w_n, n \in \mathcal{N}$  par la proposition suivante.

**Proposition 4.2** *Soit  $x_n$  un réel positif et  $f_n$  sa fonction d'utilité. On introduit :*

$$F_n = \begin{pmatrix} -\frac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix}.$$

Alors  $F_n \succeq 0 \Leftrightarrow w_n \leq f_n(x_n)$ .

**Démonstration :** Il s'agit la d'une simple application de la proposition 4.1. ■

Ainsi, grâce aux matrices  $F_n, n \in \mathcal{N}$ , on remplace les variables initiales  $x_n, n \in \mathcal{N}$  par les variables  $w_n, n \in \mathcal{N}$ . Au lieu de maximiser une fonction des  $x_n$ , on va alors maximiser une fonction des  $w_n$ . On voit déjà, dans ce petit exemple comment l'on a remplacé une fonction d'utilité par une contrainte de semi-définie positivité d'une matrice.

On peut alors prouver facilement que  $\frac{1}{1-\alpha} \sum w_n^{1-\alpha} \leq \frac{1}{1-\alpha} \sum f_n(x_n)^{1-\alpha}$  et ceci indépendamment de la valeur de  $\alpha$  (il suffit de distinguer 3 cas selon la valeur de  $\alpha$  :  $\alpha < 1$ ,  $\alpha = 1$  et  $\alpha > 1$ ).

Supposons maintenant que l'on dispose de  $\mathcal{N}$  variables  $y_n$  telles que  $\forall n \in \mathcal{N}, y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha}$  (pour  $\alpha$  fixé). On peut alors utiliser le lemme suivant :

**Lemme 4.1** *Soit  $z$  une variable réelle et  $y_n, 1 \leq n \leq \mathcal{N}$  des réels positifs tels que  $\forall n, y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha}$ . La positivité de la matrice scalaire  $FIN$  assure que  $z \leq \sum \frac{w_n^{1-\alpha}}{1-\alpha}$ .*

$$FIN_{1,1} = \left( \sum_{1 \leq k \leq \mathcal{N}} y_k - y \right)$$

Ainsi, la maximisation de  $z$  conduit bien a la résolution du problème car :

$$z \leq \sum_n y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha} \leq \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$$

De plus, les contraintes du système ont été incorporées dans la matrice SDP par les matrices scalaires précisées dans la sous-section précédente.

Nous montrons dans la section suivante comment construire des matrices dont la contrainte de positivité assurera que (on omet l'indice "n") :

$$y \leq w^{1-\alpha} \quad \text{si } 1 - \alpha > 0, \quad y \geq w^{1-\alpha} \quad \text{sinon.}$$

## 4.2 Les différentes valeurs de $\alpha$

Nous présentons dans cette section les matrices dont la positivité permet de lier les variables  $y$  aux variables  $w$ . La première sous-section s'intéresse aux cas où  $\alpha \neq 1$  tandis que la seconde traite le cas, plus complexe, du NBS ( $\alpha = 1$ ).

### 4.2.1 Calculs d'équité pour $\alpha \neq 1$

#### Cas $\alpha = 2$ .

Notre but est d'exhiber une matrice dont la positivité assure la contrainte :  $\frac{1}{w} \leq y$ . Nous avons pour cela la proposition suivante.

**Proposition 4.3 (Cas  $\alpha = 2$ , équité harmonique)** *La positivité de la matrice  $H$  assure que  $y \leq 1/w$ .*

$$H = \begin{pmatrix} w & 1 \\ 1 & y \end{pmatrix} \quad (4.3)$$

On utilise une idée de Nemirovski pour offrir une méthode de résolution de notre problème avec une précision aussi bonne que voulue pour toutes les valeurs de  $\alpha > 0$  avec  $\alpha \neq 1, 2$ .

#### Cas $0 < \alpha < 1$ .

On a  $1 - \alpha > 0$ , donc il suffit d'exhiber une (ou plusieurs) matrice(s) dont la contrainte de psd permet d'obtenir :

$$y \leq w^{1-\alpha}.$$

**Proposition 4.4 (Cas  $0 < \alpha < 1$ )** *On a  $1 - \alpha < 1$ . Alors,*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \dots, 2^p - 1\}, |(1 - \alpha) - k/2^p| \leq \varepsilon.$$

*Soit  $w, y \in \mathbb{R}^+$ . Il est possible, en utilisant des contraintes de positivité de borner  $y$  et  $w$  par la relation  $y \leq w^{k/2^p}$ .*

**Démonstration :** Soit  $c_1, \dots, c_p$  une série d'entiers binaires (c'est-à-dire qui valent 0 ou 1) tels que  $k = \sum_{i=1}^p c_i 2^{i-1}$ . On note  $y_0 = 1$  et on soumet  $y_1, \dots, y_p$  aux contraintes suivantes :

$$\left\{ \begin{array}{l} \begin{pmatrix} y_{i-1} & y_i \\ y_i & w \end{pmatrix} \text{ if } c_i = 1 \\ \begin{pmatrix} y_{i-1} & y_i \\ y_i & 1 \end{pmatrix} \text{ if } c_i = 0 \end{array} \right. \quad (4.4)$$

Alors, de façon évidente,  $y_i^2 \leq y_{i-1} w^{c_i}$ , et  $y_p \leq w^{k/2^p}$ . D'où le résultat en posant  $y_p = y$ . ■

On peut donc résoudre le problème avec une précision aussi bonne que souhaitée.

#### Cas $\alpha > 1$ .

Nous présentons ici des solutions simples pour  $\alpha > 1$ . Comme  $1 - \alpha < 0$ , il s'agit de trouver une (ou plusieurs) matrices dont les contraintes assurent que :

$$y \geq w^{1-\alpha}.$$

**Proposition 4.5 (Cas  $1 < \alpha < 2$ )** On a  $0 < -(1 - \alpha) < 1$ . Alors,

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \dots, 2^p - 1\}, |-(1 - \alpha) - k/2^p| \leq \varepsilon.$$

Soit  $w, y \in \mathbb{R}^+$ . Il est possible, en utilisant des contraintes SDP de borner  $w$  et  $y$  par la relation  $w \geq y^{-k/2^p}$ .

**Démonstration :** Soit  $c$  une variable intermédiaire. En utilisant la proposition 4.4, on peut fixer  $c \leq y^\beta$ /Nous pouvons alors écrire :

$$\begin{pmatrix} y & 1 \\ 1 & w \end{pmatrix} \succeq 0$$

ce qui donne  $yw \geq 1$ . Ainsi  $yw^\beta \geq 1$ . ■

La proposition suivante couvre le cas  $\alpha \in [2; +\infty[$ .

**Proposition 4.6 (Cas  $2 < \alpha$ )** On a  $0 < \frac{1}{1-\alpha} < 1$ . Alors, :

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \dots, 2^p - 1\}, \left| \frac{1}{1-\alpha} - k/2^p \right| \leq \varepsilon.$$

Soit  $y, w \in \mathbb{R}^+$ . Il est possible, en utilisant des contraintes SDP de borner  $w$  et  $y$  par la relation  $y \geq w^{2^p/k}$ .

**Démonstration :** De la même façon on obtient  $wy^\beta \geq 1$ . ■

## 4.2.2 Calcul du NBS (cas $\alpha = 1$ )

Nous proposons dans cette sous-section une méthode de résolution du NBS (cas  $\alpha = 1$ ). La méthode utilisée ici est légèrement différente que précédemment, car l'on ne fait plus intervenir les variables intermédiaires  $w$ . Nous exhibons directement une suite de matrice dont la contrainte SDP assure que :

$$z \leq \prod_n b_n.$$

Le résultat repose sur la proposition suivante :

**Proposition 4.7** Soit  $z$  et  $y_1, \dots, y_n$  des nombres réels positifs. Il est possible, en utilisant des contraintes SDP de borner ces nombres par la relation :

$$z^{2^{\lceil \log_2(n) \rceil}} \leq \prod_{i=1}^N y_i.$$



**Démonstration :** Soit  $p$  le plus petit entier tel que  $2^p \geq n$ . On construit une famille de variables réelles positives  $z_{i2^k+1,(i+1)2^k}$  avec  $1 \leq k \leq p$  et  $i \in \{0, \dots, 2^{p-k} - 1\}$  satisfaisant les contraintes exprimée par les  $2^p - 1$  matrices suivantes :

$$\begin{pmatrix} z_{2i2^{k-1}+1,(2i+1)2^{k-1}} & z_{i2^k+1,(i+1)2^k} \\ z_{i2^k+1,(i+1)2^k} & z_{(2i+1)2^{k-1}+1,(2i+2)2^{k-1}} \end{pmatrix} \succeq 0, \quad (4.5)$$

où l'on note  $z_{jj} = y_j$  pour  $j \in \{1, \dots, n\}$  et  $z = z_{1,2^p}$ . Nous obtenons alors des contraintes de la forme  $z_{1,2}^2 \leq y_1 y_2$ ,  $z_{3,4}^2 \leq y_3 y_4$ ,  $z_{1,4}^2 \leq z_{1,2} z_{3,4}$  et finalement  $z_{1,2^p}^{2^p} \leq \prod_i y_i$ . ■

Ainsi, pour résoudre le problème du NBS par la SDP, nous sommes amenés à introduire des connexions fictives de sorte que le nombre total de connexions soit de la forme  $2^p$  afin de pouvoir utiliser la proposition précédente. On peut alors créer autant de liens fictifs, que ces connexions utiliseront, afin de ne pas modifier l'équilibre des autres sources. Il est important également de borner la bande passante attribuée à ces connexions afin que la valeur associée par la SDP n'augmente pas sans contrôle, ce qui provoquerait une erreur. Comme ces connexions fictives utilisent toutes leur lien propre, on peut en déduire que quel que soit leur borne supérieure  $PR_i$  associée, on aura :  $x_i = PR_i$ .

Enfin, les connexions et les liens fictifs ajoutés augmentent certes la taille de la matrice allouée mais cela ne semble pas poser de problème du point de numérique comme le suggère la remarque suivante.

**Remarque 4.1** Dans le cas de  $N$  connexions avec  $L$  liens, nous affirmons que nous aurons au plus  $6n - 7$  variables,  $4N - 5$  blocks de taille 2, et  $4N + L - 4$  blocks de taille 1.

**Démonstration :** Soit  $p$  le plus petit entier naturel tel que  $2^p \geq N$  et  $q = 2^p$ . Alors, on peut facilement montrer que notre méthode mènera à :  $q + \sum_{i=0}^p q2^{-i} = 3 \times 2^p - 1$  variables,  $q + \sum_{i=1}^p q2^{-i} = 2^{p+1} - 1$  blocks de taille 2 et  $L + 2 \cdot 2^p = 2^{p+1} + L$  blocks de taille 1. Dans le pire cas, nous aurons  $N = 2^{p-1} + 1$ , d'où le résultat. ■

### 4.3 Un exemple simple de calcul du NBS dans le cas du routage fixé

Nous illustrons dans cette section les résultats obtenus précédemment par un exemple de calcul du NBS dans un cas simple de réseau.

On considère un réseau de  $L = 4$  liens et  $\mathcal{N} = 3$  connexions. Les chemins suivis par les connexions sont définies par la matrice de routage suivante :

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \quad (4.6)$$

On rappelle que l'élément  $a_{i,j}$  vaut 1 si et seulement si la connexion  $j$  utilise le lien  $i$ . Dans notre programme SDP, nous ajoutons artificiellement des connexions afin que le nombre total de connexions ait la forme  $2^p$  avec  $p \in \mathbb{N}$  comme expliqué dans la sous-section précédente. Nous supposons que cette connexion supplémentaire utilise son propre lien et ne modifie donc pas la valeur du NBS de notre problème. Un réseau possible associé à ce problème se trouve en figure 4.1

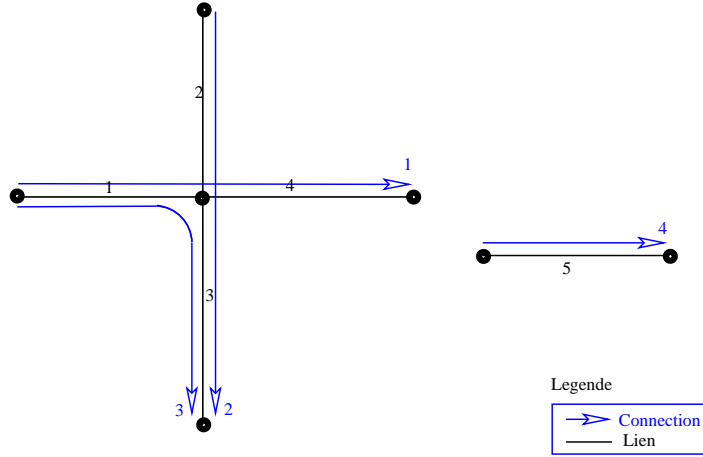


FIG. 4.1 – Exemple simple de calcul du NBS.

### Etape 1

Les quatre premiers blocs de la matrice lient les variables  $x_n$  avec leur utilité  $w_n$  :

$$\text{MAT}_{1,n} = \begin{pmatrix} -\frac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix} \quad (4.7)$$

### Etape 2

Les matrices suivantes lient les variables  $w_i$  ensemble pour obtenir une seule variable que l'algorithme SDP va maximiser (par la proposition 4.7).

$$\text{MAT}_{2,1} = \begin{pmatrix} w_1 & y_{1,2} \\ y_{1,2} & w_2 \end{pmatrix}, \text{MAT}_{2,2} = \begin{pmatrix} w_3 & y_{3,4} \\ y_{3,4} & w_4 \end{pmatrix}, \text{MAT}_{2,3} = \begin{pmatrix} y_{1,2} & y_{1,4} \\ y_{1,4} & y_{3,4} \end{pmatrix}.$$

On voit ici clairement que la positivité de ces matrices implique que  $(y_{1,4})^4 \leq (y_{1,2})^2 \cdot (y_{3,4})^2 \leq w_1 \cdot w_2 \cdot w_3 \cdot w_4$ . Alors, la maximisation de la variable  $y_{1,4}$  permettra d'obtenir l'optimisation voulue de  $\prod_n (c_n - a_n(x_n - b_n)^2)$ .

### Etape 3

Nous incorporons les contraintes (linéaires) du problème. Les contraintes  $(Ax)_l \leq C_l$  se traduisent par la déclaration de  $L$  matrices qui sont dans notre exemple :

$$\text{MAT}_{3,1} = (C_1 - (x_1 + x_3)), \quad \text{MAT}_{3,2} = (C_2 - x_2), \quad \text{MAT}_{3,3} = (C_3 - (x_2 + x_3))$$

et  $\text{MAT}_{3,4} = (C_4 - x_1)$ .

Les contraintes  $x_i \leq PR_i$  et  $x_i \geq MR_i$  sont reflétées par 8 matrices de taille une qui sont :

$$\text{MAT}_{3,4+n} = (PR_n - x_n), \quad \text{et} \quad \text{MAT}_{3,8+n} = (x_n - MR_n), \quad 1 \leq n \leq 4.$$

On peut noter que les valeurs  $PR_4$  et  $MR_4$ , correspondantes à la connexion artificielle ne sont pas importantes car la connexion est indépendante des autres. Cependant, il est important de borner  $x_4$  afin que l'allocation de la connexion virtuelle n'augmente pas indéfiniment ce qui provoquerait une erreur. La solution SDP sera toujours  $x_4 = PR_4$ .

### Conclusion

Les entrées que nous devons donner à l'algorithme SDP sont la matrice que nous obtenons par concaténation des blocs que nous venons de décrire ainsi que le vecteur  $L$ . Comme nous voulons maximiser  $y_{1,4}$ , le vecteur devra avoir un coefficient  $-1$  à la ligne correspondante et 0 aux autres :  $L = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1)$ . On peut remarquer que nous avons dans cet exemple 11 variables :

- 4 variables d'allocation  $x_n$
- 4 variables d'utilités  $w_n$
- 3 variables intermédiaires  $y_{1,2}$ ,  $y_{3,4}$  et  $y_{1,4}$  nécessaire pour le calcul du NBS.

Nous présentons dans la section suivante quelques résultats numériques obtenus par le solveur `csdp` fonctionnant sous Matlab. Les matrices ont été obtenues de la façon que nous venons d'expliquer.

## 4.4 Expériences

Nous avons implémenté l'approche SDP en utilisant un programme Matlab sur une machine SUN ULTRA 1 pour obtenir les allocations équitables. Nous avons tout d'abord testé notre programme sur le réseau linéaire (pour lequel nous avons les expressions analytiques, cf Section 3.1.1). Les résultats coïncidèrent parfaitement.

Nous avons ensuite considéré deux réseaux plus complexes que nous décrivons ci-dessous. Le temps de calcul (y compris la partie affichage) dans les deux cas fut inférieur à la minute.

Dans le cas de l'optimisation connection-aware, les paramètres sont identiques dans les deux réseaux. C'est ainsi que tous les liens sont supposés avoir la même capacité  $C = 100$  (alors que le programme permet de manipuler des capacités différentes sans augmenter la complexité). Pour chaque réseau, nous présentons une figure avec l'ensemble des liens et la quantité de bande passante attribuée à chaque connexion. Toutes les connexions avaient la même utilité quadratique 1.9 avec les paramètres  $MR = 10$ ,  $PR = 80$ ,  $a = 1/490$ ,  $b = 745$  et  $c = 1102.5$ . Les paramètres de bande passante et les affectations sont données en pourcentage de la capacité totale du lien.

Dans le cas de l'optimisation network-aware, les tailles des liens et les demandes des connexions sont différentes les unes des autres et récapitulées dans un tableau.

Nous fournissons en annexe, un programme (écrit en langage C) permettant de construire les matrices SDP dans le cas de l'optimisation connection-aware avec du routage fixé (Annexe A).

Enfin, on peut citer l'article de Pióro et al. [97] qui propose plusieurs formulations de contraintes de problèmes dans l'optique de dimensionnement de réseaux. En fait, les auteurs ne considèrent dans leur étude que l'équité proportionnelle et l'équité max-min, mais leurs formulations sont directement généralisables à un critère d'équité quelconque. Ils se placent toujours à routage fixe. L'intérêt de leur article est de proposer différents scénarios, comme par exemple le cas où les liens ont des coûts pour l'opérateur du réseau, et que le système possède une contrainte supplémentaire sur le coût maximal du réseau. Les divers scénarios de l'article sont en fait autant de systèmes de contraintes que l'on peut utiliser lors d'expérimentations dans le but de dimensionner un réseau avant de le déployer. Ces contraintes étant linéaires, on peut directement les inclure à notre méthode SDP.

#### 4.4.1 Calcul du NBS dans un petit réseau à routage fixé

Considérons le premier réseau représenté en figure 4.2. Il comprend  $L = 10$  liens et  $N = 11$  connexions. Le routage est défini dans la matrice  $A$  ci dessous.

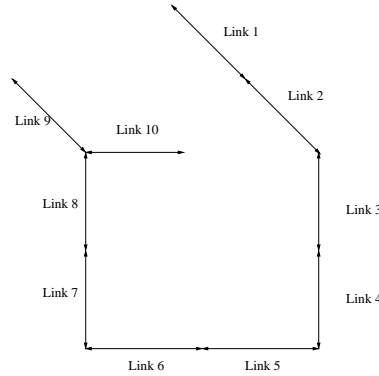


FIG. 4.2 – Petit réseau à routage fixé.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

La solution fournie par la SDP est donnée en figure 4.3. Dans la figure nous représentons chacune des connexions avec un niveau de gris correspondant à l'allocation obtenue.

On peut noter que la formulation SDP demanda l'ajout de 36 variables intermédiaires, et que la matrice SDP était de taille 104 (31 matrices de taille 2, et 42 de taille 1).

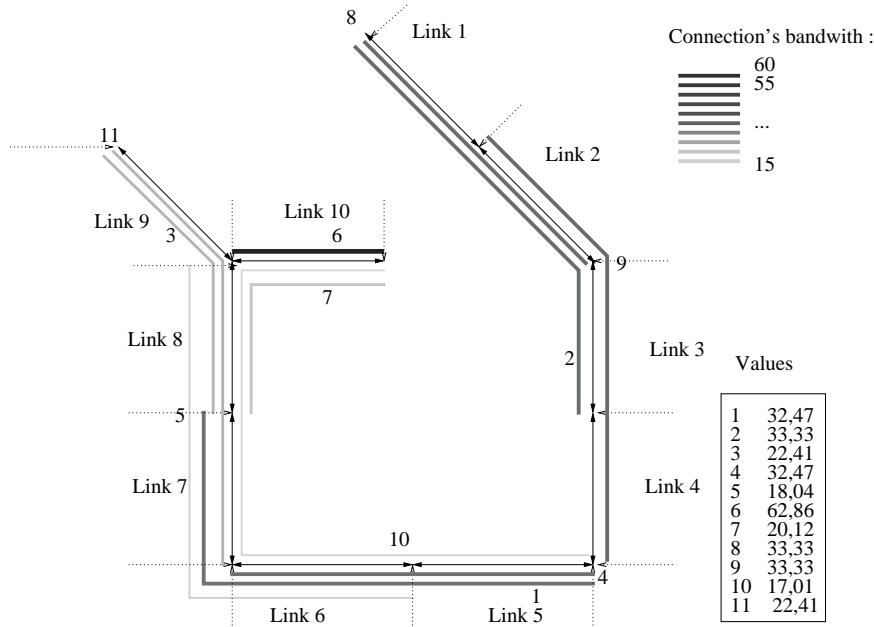


FIG. 4.3 – Premier réseau

On peut remarquer dans ce réseau que tous les liens ne peuvent être saturés bien que la solution obtenue soit effectivement optimale au sens de Pareto. C'est le cas notamment des liens 1, 3 et 4. De plus, on peut distinguer 2 sous-systèmes indépendants :

- Le système formé par les liens 1 à 4 et utilisé uniquement par les connexions 2, 8 et 9,
- Le système formé par les liens 5 à 11, dans lequel les autres connexions se partagent les ressources.

Notons que dans le premier sous-système, seul un lien est saturé. Ce lien étant utilisé par chacune des 3 connexions, elles obtiennent chacune un tiers de sa capacité totale (car elles ont, rappelons le, la même fonction d'utilité).

Le second sous-système est plus complexe, et l'on ne peut pas déterminer, sans l'aide de l'ordinateur, les différentes allocations des connexions. Cependant, les résultats numériques correspondent à notre intuition puisque la connexion 6 (celle qui n'utilise qu'un seul lien) obtient l'allocation la plus grande (presque 63% de la capacité du lien), alors que la connexion 10 (celle traversant le maximum de liens) obtient la plus faible allocation (17% de la capacité d'un lien). Dans le cas de l'optimisation max-min, ces différences auraient été moins importantes.

Enfin, rappelons que l'élément critique dans l'allocation des connexions est la donnée des liens saturés et non le nombre de liens total traversé par une connexion. Ainsi, sur notre exemple, cela explique que les connexions 11 et 3 se voient allouer la même quantité de bande passante bien que la connexion 3 traverse davantage de liens. En fait, en y regardant de plus près, on peut constater que le lien supplémentaire quelle traverse (le lien 7) n'est pas saturé, d'où le résultat. Le même raisonnement s'applique aux connexions 1 et 4.

### 4.4.2 Le réseau COST

Nous avons ensuite considéré le réseau expérimental COST décrit dans [2] et représenté dans la figure 4.4. Il se compose de 11 noeuds, représentant 11 grandes villes Européennes.

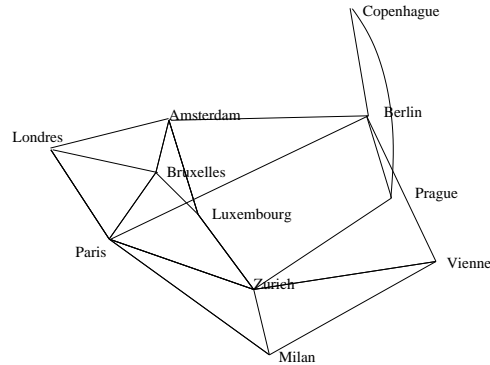


FIG. 4.4 – Réseau COST : liens.

Dans les deux paragraphes suivants, deux schémas d’optimisations sont présentés. Le premier correspond à l’optimisation connection-aware avec routage fixé, le second à l’optimisation network-aware avec routage fractionnaire. Dans les deux cas, les allocations en bande passante sont présentées, tout comme dans l’exemple précédent, en pourcentage de la capacité totale du lien.

#### 4.4.2.1 Le NBS dans le cas de l’optimisation connection-aware avec routage fixe.

Nous avons tout d’abord considéré le cas de l’optimisation connection-aware avec routage fixé pour  $\alpha = 1$ . Nous rappelons que dans cette simulation, tous les liens sont supposés avoir la même capacité et toutes les connexions ont la même fonction d’utilité quadratique. Les paramètres des fonctions d’utilité sont  $MR = 10$ ,  $PR = 80$ ,  $T = 3$  et  $f(PR) = 200$  et la capacité des liens est  $c = 100$ .

Nous avons considéré dans la simulation les 30 connexions ayant les plus grandes demandes selon les prévisions<sup>2</sup>. La solution obtenue dans le cas du NBS (c’est-à-dire le cas  $\alpha = 1$ ) est représentée en figure 4.6 et les résultats sont résumés dans le tableau en figure 4.5. Le routage a été choisi de façon arbitraire, de sorte à minimiser le nombre de liens traversé par chaque connexion.

La résolution a nécessité l’ajout de 65 variables intermédiaires, et la matrice SDP était de taille 215 (63 matrices de taille 2, et 89 matrices de taille 1).

Comme dans l’exemple précédent, les connexions ayant les allocations les plus grandes sont représentées en traits plus sombres. On remarque encore que les connexions utilisant le moins de liens (saturés) sont celles obtenant la plus grande allocation. On remarque même que quatre connexions (Copenhague-Prague, Copenhague-Berlin, Londres-Bruxelles et Paris-Berlin) sont seules à utiliser un lien. Ainsi, elles se voient allouer leur maximum demandé, c’est-à-dire 80% de la capacité du lien. Ces quatre liens ne sont donc pas saturés.

<sup>2</sup>C’est-à-dire que nous n’avons pas inclus les connexions dont les demandes, selon les expérimentations de 1993 étaient inférieures à 2.5 Gb/s

Connection	Bandwith	Connection	Bandwith	Connection	Bandwith
London-Paris	33.93	Zurich-Milano	71.58	Milano-Viena-Berlin	37.00
London-Bruxelles	80.00	Copenhague-Berlin	80.00	Milano-Paris-Bruxelles	27.93
London-Amsterdam	76.27	Copenhague-Prague	80.00	Berlin-Amsterdam-Bruxelles	22.04
Amsterdam-Berlin	27.11	Berlin-Prague	50.00	Paris-Bruxelles-Amsterdam	28.42
Amsterdam-Bruxelles	49.54	Berlin-Viena	63.00	Paris-Zurich-Viena	25.48
Bruxelles-Paris	43.66	Milano-Viena	63.00	London-Paris-Milano	24.74
Paris-Berlin	80.00	Berlin-Amsterdam-Luxembourg	27.11	London-Paris-Zurich	21.87
Paris-Zurich	33.19	Zurich-Prague-Berlin	50.00	London-Amsterdam-Berlin	23.73
Paris-Milano	47.34	Zurich-Luxembourg-Amsterdam	35.79	Viena-Zurich-Paris-London	19.46
Zurich-Viena	55.06	Zurich-Luxembourg-Bruxelles	35.79	Milano-Zurich-Luxembourg-Amsterdam	28.42

FIG. 4.5 – Allocation de bande passante dans les réseaux COST.

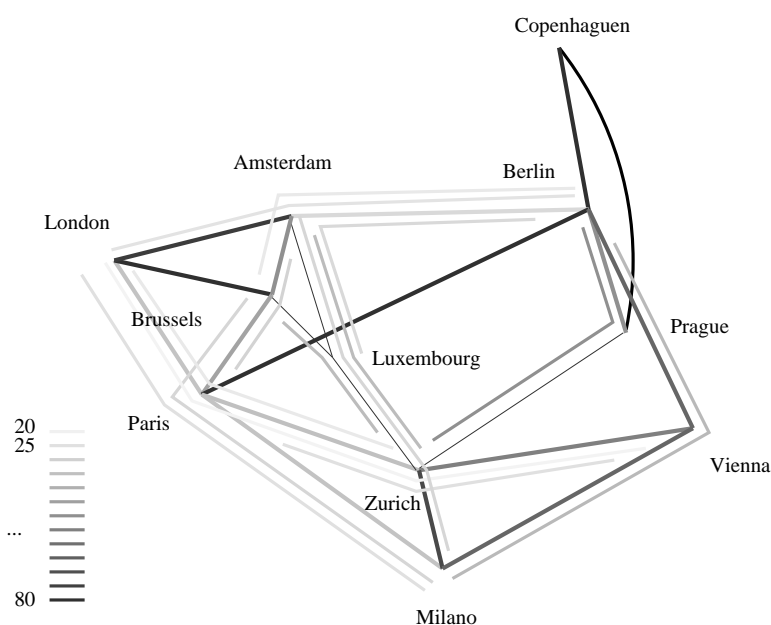


FIG. 4.6 – réseau COST : solution.

Les allocations des connexions Berlin-Prague et Berlin-Prague-Zurich sont égales et valent 50% de la capacité des liens car elles sont en compétition pour le même lien congestionné (la raison est donc la même que pour les connexions 1 et 4 du réseau précédent).

#### 4.4.2.2 Optimisation network-aware

Nous avons ensuite considéré le cas de l'optimisation network aware associée au routage fractionnaire. Dans cet exemple, nous avons ajouté trois liens supplémentaires : Londres-Copenhague, Amsterdam-Copenhague et Luxembourg-Prague afin d'augmenter la diversité des chemins possibles. Nous nous sommes de plus intéressés à différentes valeurs de  $\alpha$ . Je présente ici dans cette section les résultats obtenus pour les valeurs  $\alpha = 0.5$  et  $\alpha = 5$ .

La capacité de chaque lien est de 15 unités et nous avons considéré les 110 demandes possibles<sup>3</sup>. Les demandes ne sont plus identiques, comme illustré dans le tableau 4.1.

Ici, bien que la taille du problème soit ainsi sensiblement plus grande que dans l'exemple précédent (du fait du nombre de connexions et de l'augmentation du nombre

<sup>3</sup>Correspondant à l'ensemble de toutes les paires sources-destinations.

Paris - Milano	5	Paris - Zurich	6	Paris - Luxembourg	1	Amsterdam - Luxembourg	1
Paris - London	10	Paris - Vienna	2	Paris - Amsterdam	5	Paris - Copenhague	1
Paris - Prague	1	Paris - Berlin	11	Milano - Amsterdam	2	Milano - Luxembourg	1
Paris - Brussels	6	Milano - Berlin	9	Milano - Copenhague	1	Zurich - Luxembourg	1
Milano - Vienna	3	Milano - Brussels	2	Amsterdam - Copenhague	1	Luxembourg - Brussels	1
Milano - Zurich	6	Milano - London	3	Zurich - Amsterdam	3	Zurich - Copenhague	1
Zurich - Vienna	3	Zurich - Berlin	11	Prague - Copenhague	1	Berlin - Amsterdam	8
Zurich - London	3	Zurich - Brussels	6	Prague - Luxembourg	1	Prague - Amsterdam	1
Prague - Zurich	1	Prague - Brussels	1	Luxembourg - London	1	Luxembourg - Copenhague	1
Prague - Berlin	2	Prague - Vienna	1	Vienna - Amsterdam	1	Vienna - Luxembourg	1
Prague - Milano	1	Vienna - London	2	Vienna - Copenhague	1	London - Amsterdam	5
Vienna - Brussels	1	Vienna - Berlin	9	London - Copenhague	1	Berlin - Luxembourg	2
London - Prague	1	London - Berlin	8	Amsterdam - Brussels	4	Brussels - Copenhague	1
London - Brussels	4	Berlin - Brussels	6	Copenhague - Berlin	3		

TAB. 4.1 – réseau COST : Demandes

de liens), le temps de calcul est resté très inférieur à la minute.

Les résultats sont donnés pour  $\alpha = 0.5$  et  $\alpha = 5$  dans les tableaux 4.2 et 4.3 respectivement et représentés graphiquement dans les figures 4.7 et 4.8.

Lien	Conn	Lien	Conn	Lien	Conn	Lien	Conn
Paris - Milano	57	Zurich - Prague	87	Vienna - Berlin	96	Amsterdam - London	61
Paris - Zurich	68	Prague - Vienna	55	Berlin - Amsterdam	92	Amsterdam - Copenhague	29
Paris - Berlin	97	Prague - Berlin	93	Berlin - Copenhague	81	Amsterdam - Brussels	90
Paris - Brussels	68	Milano - Brussels	73	Prague - Luxembourg	32	Amsterdam - Luxembourg	62
Paris - London	91	Zurich - Vienna	41	Prague - Copenhague	61	London - Copenhague	87
Milano - Zurich	61	Milano - Vienna	81	Zurich - Luxembourg	73	Luxembourg - Brussels	61
Brussels - London	62						

TAB. 4.2 – Résultats pour  $\alpha = 0.5$ 

Lien	Conn	Lien	Conn	Lien	Conn	Lien	Conn
Paris - Milan	52	Milan - Zurich	73	Zurich - Luxembourg	76	Amsterdam - Londres	73
Paris - Zurich	73	Prague - Vienne	76	Berlin - Amsterdam	92	Amsterdam - Copenhague	47
Paris - Berlin	92	Prague - Berlin	92	Berlin - Copenhague	92	Luxembourg - Brussels	73
Paris - Brussels	76	Zurich - Prague	76	Prague - Luxembourg	47	Amsterdam - Luxembourg	74
Paris - Londres	78	Prague - Copenhague	74	Amsterdam - Brussels	77	Londres - Copenhague	73
Milan - Brussels	76	Vienne - Berlin	92	Brussels - Londres	74	Zurich - Vienne	47
Milan - Vienne	73						

TAB. 4.3 – Résultats pour  $\alpha = 5$ 

On observe sur les figures 4.7 et 4.8 que lorsque le paramètre  $\alpha$  croît, l'on tend vers une coloration plus homogène des liens. Ceci est en accord avec notre intuition, puisqu' $\alpha$  est le paramètre d'équité. Quand  $\alpha$  croît, les valeurs sont plus équilibrées, et donc nous avons une utilisation des liens plus égale. Ainsi, le taux d'utilisation du lien le plus chargé décroît, passant de 97.36% à 92%.

On peut noter que, quand  $\alpha$  croît de 0.5 à 5, les liens deviennent plus remplis en moyenne (le taux moyen d'utilisation des liens croît de 70.38 à 73.83), ce qui pourrait paraître surprenant, puisque la bande passante allouée reste constante ainsi que les capacités des liens. On peut en fait expliquer ce résultat de la façon suivante. Lorsque  $\alpha$  augmente, le trafic est d'avantage scindé entre différents chemins afin d'harmoniser l'utilisation des liens. Ce faisant, le nombre moyen de liens utilisés par connexion augmente également et donc l'utilisation du réseau aussi.

Il est également intéressant de noter que le taux d'utilisation de certains liens, comme Berlin-Copenhague, a augmenté presque jusqu'à leur maximum d'utilisation dans la solution pour  $\alpha = 5$ , qui est supposé être davantage équitable! Ceci montre que la recherche de l'équité à une échelle globale mène quelquefois quelques utilisateurs dans des situations critiques alors qu'ils sont à l'intérieur de leurs bornes dans des cas moins équitables.



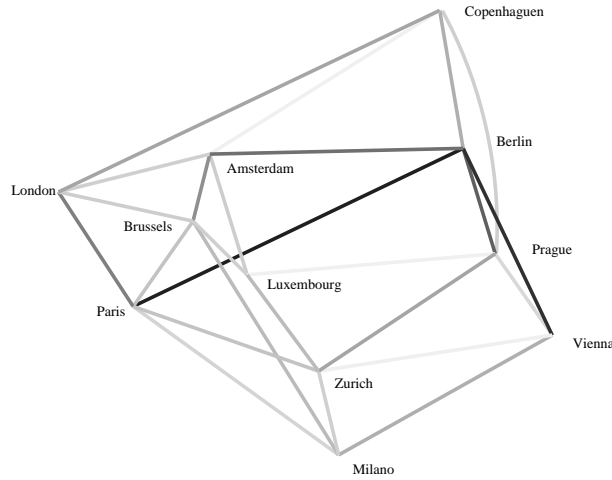


FIG. 4.7 – Allocation de bande passante pour le réseau COST network,  $\alpha = 0.5$ .

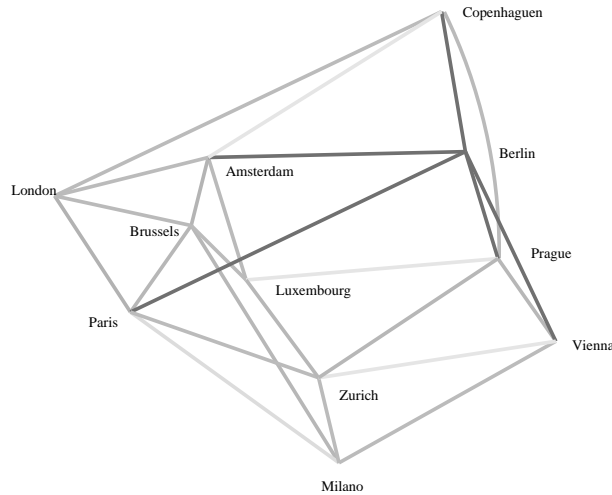


FIG. 4.8 – Allocation de bande passante pour les réseaux COST,  $\alpha = 5$ .

## 4.5 Conclusion

Nous avons présenté dans ce chapitre une méthode de résolution centralisée basée sur la programmation SDP. Une telle approche est donc adaptée aux problèmes de dimensionnement de réseaux. Nous avons présenté une méthode de construction possible de la matrice semi définie positive et avons montrés quelques résultats numériques dans des contraintes différentes (routage fractionnaire ou fixé, optimisation network-aware ou connection-aware).

Nous avons vu que la SDP à l'avantage de pouvoir résoudre les problèmes d'allocation équitables dans les réseaux appliquées à des fonctions d'utilité quadratiques pour des valeurs du paramètre d'équité  $\alpha$  et une topologie de réseau quelconques. Ceci contraste avec la méthode de relaxation Lagrangienne que nous avons présentée au chapitre précédent qui ne permettait pas l'obtention des allocations équitables autres que l'équité proportionnelle.

Enfin, nous avons vu que la taille des matrices est très creuse puisque diagonale par bloc et que les blocs sont au plus de taille 2. De plus la nombre des blocs reste raisonnable

ce qui permet en pratique d'utiliser la SDP pour résoudre des problèmes de taille réelle. A titre d'exemple, considérons l'exemple du routage fixe (dans le cas du routage fractionnaire, le nombre de blocs de taille 1 est légèrement supérieur). Soit  $N$  le nombre de connexions et  $L$  le nombre de liens. Dans le cas où  $\alpha \neq 1$ , je note  $p$  l'entier servant dans l'approximation de  $\alpha$  comme présenté dans la section 4.2.1. On a alors (dans le cas de l'optimisation avec routage fixé) :

taille du bloc	1	2
$\alpha = 1$	$4N + L - 4$	$4N - 5$
$\alpha \neq 1$	$2N + L + 1$	$p \lceil N/2 \rceil$

avec  $\lceil N/2 \rceil$  représentant la valeur entière immédiatement supérieure à  $N/2$ . On voit ainsi une propriété très intéressante de cette méthode : le nombre de blocs est dans tous les cas en  $\mathcal{O}(N + L)$ .



# Conclusion

Nous avons présenté dans ce chapitre des schémas d'allocations équitables de bande passante dans les réseaux terrestres. Notre approche ne prend pas en compte les débits associés à chaque connexion ou les quantités de bande passante résiduelle des liens, mais l'utilité que ces grandeurs représentent. Ce type de concept d'équité coïncide avec la définition de théorie des jeux de l'équité donnée par le NBS et est une généralisation intéressante des critères d'équité qui ont été proposés jusqu'alors dans le contexte des télécommunications. De plus, ceci est tout à fait justifié, puisque, comme nous avons pu le voir dans notre introduction générale, les diverses applications utilisées dans les réseaux ont des besoins en bande passante extrêmement différents. Aussi, il est intéressant de pouvoir enfin prendre cette diversité en compte dans un partage équitable.

Nous avons proposé une approche décentralisée basée sur une méthode de Lagrangien, et une approche centralisée de calcul de l'allocation équitable, basée sur la SDP.

## Comparaison des deux méthodes.

L'avantage de l'approche décentralisée est son implémentation facile dans les réseaux, chaque utilisateur n'ayant pas besoin de connaître ni l'architecture du réseau, ni le comportement ou le nombre des autres connexions mais uniquement le coût associé aux liens qu'il utilise. Dans le cas des fonctions d'utilités linéaires, le réseau, de son côté, n'a pas à connaître les fonctions d'utilité des utilisateurs présents. Dans le cas des fonctions d'utilités quadratiques cependant cette connaissance est nécessaire. De plus, notons que cette approche n'est pas généralisable à un critère quelconque d'équité.

En revanche, la méthode centralisée peut permettre, de prévoir le comportement futur des connexions sur un réseau et ainsi d'apporter des solutions efficaces et rapides aux problèmes de dimensionnements. Cette méthode a l'avantage d'être facile à implémenter en utilisant n'importe quel programme général de SDP. De plus des conditions supplémentaires (par exemple celles liées à la programmation entière ou toute autre requête de télécommunications) peuvent être ajoutées sans nécessiter d'autres études sur la stabilité ou la convergence de l'algorithme. Ceci est une nette amélioration comparée à bien d'autres méthodes très spécifiques et en particulier aux méthodes itératives qui ont précédemment été proposées pour résoudre des problèmes d'équité.

## Extensions possibles des travaux d'équité dans les réseaux terrestres.

Les travaux sur les réseaux terrestres restent un domaine de recherche toujours très ouvert. Il serait intéressant d'introduire à notre modèle d'autres notions fondamentales en télécom-

munications. Par exemple, les contraintes inhérentes à l'utilisation du *multi-cast* auraient tout à fait leur place dans cette étude.

De plus, dans notre vue du routage fractionnaire, tous les chemins possibles entre deux points sont acceptables. Or, en particulier pour les grands réseaux, on pourrait imaginer de *borner la longueur des routes*, borne dépendant, par exemple, des longueurs des routes minimales.

De plus, de nouveaux *protocoles de contrôle de congestion* pourraient être développés dont les allocations tendraient vers des équilibres équitables pour de nouvelles valeurs du paramètre  $\alpha$  (et notamment des valeurs différentes de 0 et de 1).

Enfin, des études des *équilibres* obtenus dans des réseaux de topologies diverses par les différents protocoles existants actuellement et réalisant une certaine forme d'équité pourrait être effectuée grâce à la SDP (en particulier pour TCP Vegas dont il a été montré que les allocations étaient proportionnellement équitables).

### Le protocole TCP

Une extension assez directe de nos travaux serait d'utiliser la SDP pour simuler les équilibres obtenus par des sources utilisant la variante Vegas de TCP.

Rappelons que la version de TCP actuellement déployée dans les réseaux est TCP Reno. TCP Vegas a été introduit en 1994 comme une alternative à TCP Reno. Il présente des améliorations sensibles par rapport à son aîné. Par exemple, alors que Reno utilise la probabilité de perte comme indication de congestion, TCP Vegas utilise le délai. Ainsi, il permet d'obtenir un meilleur débit et des pertes plus faibles que TCP Reno sous nombre de scénarios. Une description détaillée du protocole TCP Vegas se trouve dans [72].

Il a été montré ([72]) que l'équilibre obtenu par TCP Vegas est une certaine forme d'équité proportionnelle :

$$\max_x \prod x_n^{d_n \cdot \alpha_n}$$

Avec  $d_n$  le RTT associé à la connexion et  $\alpha_n$  un paramètre, représentant la différence maximale entre le débit attendu et le débit réel minimum auquel la source veut voir sa connexion maintenue<sup>4</sup>. (elle est typiquement de  $1/d_n$ ). Ainsi, l'équilibre obtenu par des sources TCP Vegas est typiquement la solution de :

$$\max_x \prod_{n \in \mathcal{N}} x_n$$

sous les contraintes de capacités du système. Dans le cas où  $d\alpha \neq 1$ , on peut de nouveau appliquer la SDP car on a vu que l'on pouvait intégrer certaines familles de polynômes. Cependant, l'application n'est pas immédiate et nécessite une recherche plus poussée.

Avec TCP Reno, la fonction optimisée à l'équilibre est une fonction inverse de tangente. L'équilibre ne correspond plus à l'un de nos critères d'équités. En revanche, tout laisse à croire que certaines approximations de la fonction tangente peuvent être utilisées afin

---

<sup>4</sup>En effet, contrairement à Reno, dans lequel la taille de la fenêtre est modifiée à chaque RTT, dans Vegas, la taille de la fenêtre ne change que si la valeur de cette différence est supérieure à un certain seuil.

d'obtenir une bonne approximation par la méthode SDP des équilibres obtenus par des sources TCP.

Enfin, on peut noter que TCP est un protocole qui a été très largement étudié, probablement du fait de son déploiement important. De nombreuses variantes ont été proposées. Par exemple [47] propose de modifier TCP pour se rapprocher de l'équité max-min. La encore, l'on peut croire (espérer) que la méthode de SDP pourrait permettre d'obtenir des solutions centralisée des équilibres obtenus par de tels protocoles proposés comme des développement futurs de TCP et de pouvoir étudier leur comportement.

### Conclusion générale.

L'intérêt de l'utilisation de la SDP est donc double. On voit d'une part que la SDP peut permettre d'obtenir des solutions rapides et centralisées aux problèmes d'équités rencontrés dans les réseaux filaires afin de répondre à des problèmes de dimensionnement.

Elle peut permettre également, de part son caractère général, d'étudier les allocations obtenues dans des réseaux ou se côtoient des protocoles différents, équitables où non. Elle pourrait ainsi par exemple permettre de comprendre l'impact qu'aurait l'introduction de protocoles non équitables au milieu de sources qui elles, le sont où d'étudier les équilibres obtenus dans un réseau partagé par des sources achevant divers critères d'équités.



Deuxième partie  
Les réseaux mobiles





# *Introduction*

Les réseaux de téléphonie mobile ont connus un franc succès ces dernières années. Dans les réseaux de seconde génération (comme le GSM en Europe) chaque mobile se voit affecter un débit constant prédéfini. On voit ainsi que dans un tel système, il n'existe qu'un seul type de connexion. Ceci est du en partie au fait qu'à l'époque de leur déploiement, le seul service proposé était le transport de la parole. La puissance d'émission du signal était quand à elle ajustée périodiquement en fonction de la qualité du signal.

Aujourd'hui, avec l'arrivée des téléphones de troisième génération (technologie UMTS par exemple), on considère la possibilité d'utiliser des applications diverses telles que la transmission de vidéo ou l'accès à Internet. Ceci justifie le fait que les nouveaux protocoles offrent la possibilité d'allouer dynamiquement des débits différents.

L'importance dans de tels réseaux des applications temps réel nous incite à appliquer nos concepts d'équité au partage des ressources radio, c'est-à-dire du débit.

Cependant, nous allons voir que dans les communications par voie radio le débit que l'on peut allouer à une connexion dépend d'une autre ressource du système : la puissance d'émission. En effet, le débit qu'une source peut utiliser dépend à la fois de la puissance à laquelle elle émet et de celles utilisées par les autres sources du système.

Nous nous intéressons donc dans cette partie à appliquer nos concepts d'équité introduits dans le premier chapitre aux réseaux cellulaires de troisième génération, par exemple de type UMTS. L'intérêt d'une telle étude sur le plan algorithmique est double. D'une part, les contraintes du systèmes ne se traduisent plus comme dans les réseaux terrestres par un système linéaire. De plus, nous pouvons désormais étudier un problème d'optimisation joint de deux grandeurs (le débit et la puissance).

Cette partie s'articule de la façon suivante. Dans un premier chapitre, nous présentons les travaux récents d'allocations effectués dans le domaine des réseaux mobiles. Nous présenterons dans un second notre approche et en particulier le modèle théorique que nous avons développé et qui permet de synthétiser les différents types de connexions du réseau. Enfin, dans un troisième chapitre nous appliquerons ce modèle théorique pour chacun des types. Nous concluons finalement en présentant quelques résultats numériques.



# Chapitre 5

## Précédents travaux sur le contrôle de taux dans les réseaux sans fil

Nous mentionnons brièvement dans ce chapitre quelques récents travaux sur le contrôle de débit dans les systèmes cellulaires de troisième génération. Ces réseaux ont des caractéristiques très différentes de leurs prédécesseurs.

Tout d'abord, dans ces systèmes, les taux de transmission peuvent être alloués dynamiquement aux connexions, à la fois aux applications temps-réel et non temps-réel.

De plus la méthode d'accès de tels réseaux est désormais le CDMA (Code Division Multiple Access), par opposition au TDMA (Time Division Multiple Access). En deux mots, la méthode d'accès gère la façon dont les utilisateurs se partagent le canal radio. Dans la technique TDMA, on divise le temps en petites périodes (appelées slots). On affecte alors à chaque connexion un slot ou un groupe de slot. Au contraire, dans les systèmes CDMA, les connexions ont accès en permanence au canal. Afin que chaque récepteur puisse néanmoins obtenir le signal qui lui est adressé, chaque connexion utilise un codage différent. On voit que d'une façon générale, dans les deux systèmes, plus le nombre d'utilisateur est grand, plus le débit sera réduit. En effet, dans les systèmes TDMA, le temps consécutif entre deux slots augmentera, tandis que en mode CDMA, une augmentation de la taille des codes sera nécessaire afin que le décodage puisse se faire sans erreur. Une explication plus détaillée des méthodes d'accès se trouve à la section 8.1.6.

Plus précisément dans les réseaux UMTS, un certain nombre de codes sont utilisés par les utilisateurs en mode CDMA (on dit également en mode circuit). Un code supplémentaire existe et est partagé par les utilisateurs utilisant la méthode d'accès TDMA (on dit alors qu'ils utilisent le mode paquet).

La première section a pour but de montrer quels sont les travaux qui ont déjà été effectués sur l'équité dans le cadre des réseaux sans fils. Nous présentons ensuite (Section 5.2) des références à quelques articles étudiant des allocations optimales de débits.

La seconde ressource cruciale dans les connexions radio est la puissance d'émission. En effet, nous montrerons dans le chapitre suivant qu'une source ne peut émettre à un certain débit que si sa puissance d'émission est suffisamment importante par rapport à celle des autres mobiles. En fait, la qualité d'un lien radio se mesure par son rapport signal sur bruit (SIR) qui est une fonction de la puissance. Le débit maximal qu'un mobile peut utiliser est alors proportionnel au SIR.

Je mentionnerai donc quelques travaux relatifs au contrôle de puissance (Section 5.3), puis au contrôle joint de puissance et de débit (Section 5.4).

## 5.1 L'équité dans les réseaux sans fils.

Quelques articles se sont déjà penchés sur l'allocation équitable de ressources dans les réseaux sans-fil.

L'équité a en effet été introduite pour proposer des solutions à des problèmes d'ordonnement dans des systèmes TDMA et CDMA.

Ainsi, l'article [74] s'intéresse à un ordonnancement en temps qui est une variante de l'équité max-min pondérée. Le système présenté ici est équitable à chaque instant mais si le mobile ne peut émettre lorsque vient son tour (par exemple à cause d'un problème de connexion avec la station) alors sa part de temps est allouée à un ou plusieurs autres mobiles. Plus tard, la part laissée par le mobile sera en partie compensée. Les auteurs présentent alors une implémentation possible de cet algorithme. Un article très proche est [15].

Dans la même famille d'article se trouve [37] qui considère différents algorithmes d'allocations de slots dans une trame montante TDD/TCDMA (Time Division Duplexing / Time and Code Division Multiple Access) semblable à une trame UMTS. Dans une telle trame, la multiplexion se fait en temps et en code. A titre d'exemple, dans le cas de l'UMTS, la durée de la trame est de 10ms, elle est divisée en 15 slots et dans chaque time-slot un terminal peut utiliser l'un des 16 codes possibles. L'article s'intéresse à 4 classes de services : la voix, les données rapides, les données non-rapides et les données sans contraintes de temps (c'est-à-dire un service non-temps réel). Les quatre classes ont des besoins différents (en terme de nombre de slots par trame de temps maximal d'émission et débit d'information). Le but de l'ordonneur est alors d'allouer à chaque connexion des paires (slots, code) qui respectent les besoins des connexions.

Similairement, l'article [12] s'intéresse au problème d'ordonnement dans les systèmes TCDMA. A l'intérieur de chaque slot TDMA chaque mobile peut émettre en utilisant un ou plusieurs codes. Les codes possibles sont orthogonaux<sup>1</sup> entre eux. Notons qu'un système TDMA peut être vu comme un cas particulier du TCDMA dans lequel il n'existe qu'un code disponible. L'algorithme considère pour chaque mobile deux paramètres : le rapport signal sur bruit et la file d'attente des paquets (ceux en attente d'émission ou de réémission). Les auteurs proposent alors un ordonneur, appelé CHAOS qui donne priorité aux mobiles ayant de grandes files d'attente, sous réserve que l'affaiblissement de leur signal dans le canal est acceptable.

Ainsi, dans cette famille de travaux, l'équité apparaît sous la forme d'informations d'agencement des paquets dans les buffers [15, 82, 37]. On voit ainsi que dans ces travaux les taux de transmission moyens des différentes sources sont donnés et fixés alors que notre étude veut, au contraire, les allouer équitablement.

Nous citons enfin [32] qui considère le lien descendant d'un système avec trois types canaux : des canaux de circuit dédiés (pour les applications temps réel), un canal d'overhead

---

<sup>1</sup>Pour une définition précise de ce terme, on se référera à la section 8.1.6. Grosso modo, l'idée est que des codes orthogonaux peuvent être séparés, c'est-à-dire que chaque récepteur peut correctement décoder le message qui lui est adressé des autres signaux.

et un canal de paquet (pour les données). Le contrôle de débit est uniquement utilisé sur les canaux de paquets. L'idée est que chaque canal de paquet se voit affecté la même puissance. Si celle-ci ne permet pas à l'utilisateur d'obtenir un SIR supérieur à une valeur cible, alors la puissance qui lui était allouée est transférée à un autre utilisateur. Le facteur d'équité introduit est le rapport entre le plus fort et le plus faible débit alloué. Les mesures de performances considérées sont la probabilité d'avoir épuisé les codes disponibles, la probabilité de blocage d'appel, le débit asymptotique et le facteur d'équité.

On voit donc que certaines intégration des concepts d'équité dans les réseaux mobiles ont déjà été entreprises dans un cadre différent du notre. Les articles cités dans les sections suivantes considèrent des problèmes d'allocation de ressources dans les réseaux sans fil mais ne sont pas directement liés aux concepts d'équité.

## 5.2 Choix des débits

Je présente dans cette section quelques travaux effectués sur le contrôle des débits dans les réseaux sans fil.

### Trafic non temps réel.

Plusieurs articles étudient les allocation d'optimisation de débit pour le trafic non temps réel. Ainsi, on cherche dans [107] à maximiser le débit effectif, qui est fonction du taux de transmission et du SIR. On remarque que le débit effectif est le même si on alloue à une source un taux de transmission  $R$  avec 0% de perte, ou si l'on transmet avec  $2R$  et perdons la moitié de l'information. La maximisation du débit effectifs n'est intéressante que pour les *applications non temps réels* (NRT). En effet, pour les applications temps réel (RT) les utilisateurs des deux scénarios mentionnés auraient des perceptions de qualité différentes. Les auteurs utilisent une approche lagrangienne et ne considèrent par de valeur SIR cible a priori. Une section considère à la fois des applications temps réel et non temps-réel mais seul le débit des applications non temps-réel est optimisé. Enfin, le système étudié se limite au lien montant et à une seule cellule.

Dans [98], les auteurs considèrent le lien montant CDMA avec deux classes d'utilisateurs : de temps réel et non temps réel. Les mobiles de la première classe (RT) transmettent continûment alors que la transmission des autres est partagée en temps. Les bénéfices du partage en temps sont étudiés ainsi que les conditions sous lesquelles on peut imposer aux applications non-temps réel l'arrêt de transmission. L'objectif étudié est la maximisation des débits tout en maintenant le SIR à des taux acceptables. L'article prend en compte le fait que lorsqu'un mobile est silencieux (c'est-à-dire ne transmet pas d'information) il a tout de même besoin d'un certain débit pour transmettre des messages de synchronisation. Nous remarquons aussi que le montant de l'information transmise n'est pas affecté par l'ordonnement.

La constatation simple que le débit alloué n'a pas la même conséquence sur le volume de donné échangé est faite dans [6]. En effet, dans les applications non temps-réel le débit de la connexion n'influe pas sur la quantité d'information transférée. En revanche les connexions temps-réel ont des contraintes sur la durée de transmission. Une diminution du débit alloué

s'accompagne donc d'une diminution du volume de données transféré. Ainsi, un problème d'optimisation statique tel que formulé dans [62] n'est pas adéquat pour décrire l'effet de l'allocation de débit pour les applications non temps réel. Altman décrit alors l'utilisation du réseau par le calcul de la capacité d'Erlang<sup>2</sup> en fonction du débit alloué aux applications.

### Allocation de taux instantannés.

Une autre direction de recherche liée a été l'allocation de taux de transmission instantannés des paquets au niveau des buffers dans les systèmes CDMA. Dans ces articles, le véritable taux de transmission de la source n'est pas contrôlé. Par exemple trois profils de trafics sont considérés dans [121]. Les deux premiers, le CBR (Constant Bit Rate) et le VBR (Variable Bit Rate) représentent des applications temps réel. En revanche, le ABR (Available Bit Rate) n'impose pas de contraintes sur le délai. Les taux de transmission instantannés des connexions des différents profils sont calculés en fonction de bornes sur les pertes de paquets dans les buffers des mobiles. D'autres articles très proches sont [30, 76, 115].

## 5.3 Contrôle de puissance

Nous nous penchons ici sur les travaux faits dans le domaine du contrôle de puissance dans les réseaux sans fil.

Certains articles se concentrent sur les caractéristiques des liens radio. Par exemple Berry montre qu'un tel lien peut se modéliser comme un filtre passe bande dont les propriétés varient dans le temps de façon aléatoire à la sortie duquel s'ajoute un bruit ([13]). Le but de l'auteur est de minimiser la puissance d'émission de façon à satisfaire une certaine qualité de service (qui s'exprime en termes de probabilité d'erreur et de délais des paquets). Il s'intéresse principalement au cas où il y n'y a qu'un émetteur et un récepteur. L'auteur montre le compromis qui existe, pour un taux d'erreur donné, entre la puissance moyenne utilisée et le délai. Notons que le codage de canal est un paramètre important dans le calcul du délai. En effet, il est important de minimiser la taille des codes afin de réduire le temps passé par les paquets dans les buffers avant d'être émis sur le canal. L'auteur considère alors le contrôle combiné de la puissance d'émission et du codage de canal de sorte à minimiser à la fois la puissance et le délai. Aucune retransmission et ni ordonnancement n'est considéré.

On peut citer également l'article [90] qui présente une étude des différentes modélisations existantes du canal radio. Il introduit en particulier les modèles de propagations classiques de prédiction de signal en intérieur et extérieur.

### Algorithmes actuels pour déterminer la puissance d'émission des mobiles.

La façon dont le contrôle de puissance est effectué dans les systèmes CDMA actuels est

---

<sup>2</sup>Erlang : Unité de mesure de l'intensité du trafic sur une liaison, qui sert à évaluer l'occupation optimale d'une ligne. Elle a été introduite dans les réseaux de téléphonie fixe. C'est le produit du nombre moyen de sessions (ou de communications) établies pendant une heure et de la durée moyenne des sessions. Compris entre 0 et 1 (à 1, la ligne est constamment occupée). A partir de cette valeur, il est ensuite possible de déterminer la valeur optimal du nombre de ligne nécessaires pour que la qualité de service (taux de refus) soit acceptable.

décrite dans [45]. Une catégorie d'algorithmes est basée sur un contrôle de puissance appelé *inner loop*. Le principe est qu'à la fois le mobile et la station de base mesurent la puissance du signal qu'ils reçoivent. Ainsi, le mobile choisit sa propre puissance d'émission en fonction de celle du signal qu'il reçoit de la station car les pertes de puissance dues à la distance et au masquage sont indépendantes de la fréquence utilisée. En revanche, comme les mobiles et les stations émettent à des fréquences différentes, les évanouissements de leurs signaux sont différents. Ainsi, la station de base mesure également la puissance du signal reçu, le compare à un seuil et envoie périodiquement au mobile un message lui demandant d'augmenter ou de diminuer sa puissance.

Un autre algorithme est le contrôle *outer loop*. Dans ce cas, on ne mesure plus la puissance des signaux, mais leur qualité par l'intermédiaire de leur *SIR*. Ainsi la station regarde si chaque trame reçue contient des erreurs. La valeur du *SIR* cible est augmentée si tel est le cas et diminuée sinon. Notons enfin que les auteurs considèrent le cas de la macrodiversité dans le lien montant, c'est-à-dire le cas où le signal du mobile est reçu par plusieurs stations de base et recomposé. Un article proche est [23].

### **Etude du lien montant pour une cellule isolée et à débit fixé.**

Plusieurs articles se sont intéressés au contrôle de puissance dans le lien montant (c'est-à-dire quand le mobile envoie des informations à la station de base) dans le cas d'une cellule isolée. Dans ces articles, le débit des connexions est fixé.

Par exemple, dans [76] chaque mobile peut utiliser un certain nombre de canaux ayant des débits et des seuils minimaux de *SIR* différents. Les auteurs présentent une formulation sous forme matricielle du problème.

Les auteurs de [69] s'intéressent à l'ensemble formé des puissances acceptables pour des débits de transmission donnés dans une cellule isolée et pour le lien montant. Ils montrent que ce n'est pas un ensemble convexe. Il le devient pourtant lorsque l'on utilise un ordonnancement approprié des paquets. Les auteurs s'intéressent alors à des ordonnancements possibles de façon à respecter des bornes prédéfinies sur les délais ou des contraintes sur les différents trafics de profils.

### **Contrôle d'admission.**

[11] s'intéresse au problème de contrôle d'admission dans un réseau DS-CDMA. L'article se concentre sur le contrôle de puissance uniquement. L'admission d'un nouvel appel se fait en 2 étapes. Dans un premier temps, les puissances allouées aux mobiles déjà en communication sont augmentées de façon proportionnelle afin que l'arrivée du nouvel appel ne modifie pas la qualité de leur lien. Dans un second temps, le nouvel appel est accepté et on lui affecte un niveau de puissance choisi de sorte à optimiser une certaine mesure du réseau comme le débit total ou l'énergie par bit. Les auteurs montrent également que dans le lien montant, pour un état du système donné (c'est-à-dire un *SIR* cible et des gains de liens donnés) il existe un vecteur de puissance minimal (composante par composante).

### **Autres articles.**



Nous citons deux autres articles intéressants en matière de contrôle de puissance.

Yates a présenté une étude statique du lien montant. Il propose un algorithme itératif pour trouver le vecteur de puissance minimal ([118]). Le débit des connexions étant fixé, le système est de type :  $p \geq f(p)$  pour chaque mobile avec  $p$  la puissance d'émission et  $f$  une fonction du SIR. A chaque itération de l'algorithme,  $p_{k+1}$  est choisi comme étant  $p_{k+1} = f(p_k)$ . L'auteur propose alors des conditions assurant la convergence de l'algorithme.

Enfin [5] considère le lien montant d'un système CDMA. Le système est présenté comme un jeu non coopératif dans lequel chaque joueur cherche à minimiser sa fonction de coût, différence entre le prix payé (proportionnel à la puissance utilisée) et son utilité (fonction de son rapport signal sur bruit). Les auteurs se limitent au cas d'une seule cellule et montrent l'existence d'un unique équilibre de Nash dont ils étudient les propriétés (de convergence et de stabilité notamment).

## 5.4 Contrôle joint des débits et des puissances

Nous achevons notre tour d'horizon des articles traitants des problèmes d'allocation de ressources dans les réseaux sans fils par des travaux qui ont considéré le contrôle joint des débits et des puissances.

Ainsi, [62] considère le problème de l'optimisation des taux de transmission et des puissances dans le cas de débits disponibles discrets. Le problème est formulé mathématiquement comme une programmation linéaire-entière mixte, pour laquelle des algorithmes d'optimisation polynômiaux ne sont pas disponibles. Deux méthodes d'approximations itératives basées l'une sur une relaxation Lagrangienne et l'autre sur l'idée simple d'allouer le débit maximal possible à chaque itération sont proposées et testées. Leur but est de maximiser le débit total du système.

Un autre article associé est [92], dans lequel les auteurs étudient le contrôle optimal d'à la fois la puissance et le *spreading gain* (augmenter ce dernier est équivalent à diminuer le débit). Les auteurs restreignent leur étude au cas d'une cellule isolée et au cas montant. Leur but est de maximiser le débit total des sources NRT. Le modèle inclut le codage de canal (FEC) et une fonction générale du BER (Bit Error Rate) (c'est une fonction du SIR). Des modèles similaires sont disponibles dans la littérature [93, 91]. On peut remarquer cependant que comme leur but est d'augmenter le débit global, leur algorithme tend à affecter un débit maximal aux sources bénéficiant des canaux de meilleure qualité tandis que les terminaux se trouvant dans des plus mauvaises conditions de réception se voient refuser l'accès au réseau. On observe ainsi une conséquence classique des algorithmes d'optimisations globales (cf Section 1.3.1).

## 5.5 Conclusion

Nous pouvons finalement citer l'article [30] qui propose une méthode d'allocation de ressources général en adoptant un modèle statistique. Chaque mobile génère un certain trafic (de voix) qu'il envoie à la station. Le but est alors pour elle multiplexer au mieux ce trafic (qui est un processus on-off), de façon à maximiser le nombre d'utilisateurs simultanés tout en garantissant une faible probabilité de blocage des appels, de terminaison d'appels et de

perte des paquets. Le modèle est valable pour des réseaux sans-fils comme pour des réseaux filaires où les utilisateurs sont mobiles. Les spécificités du canal radio ne sont pas pris en compte ici.

Ainsi parmi toutes les directions de recherche que nous avons mentionné dans ce chapitre, notre axe est plus proche de [62] et [92] car il s'intéresse à l'allocation véritable des débits de transmission aux sources (plutôt qu'à l'intérieur du réseau) des applications temps-réel. Nous considérons cependant ici un environnement multi-cellules général avec une attention particulière des contrôle de taux des liens montant et descendant et nous incluons une étude de la macro-diversité. (ceci contraste avec [92] qui ne considère que le cas du lien montant sur un seul lien, ou de [32] dont le modèle paraît plus adapté au lien montant du cas multi cellulaire). De plus, un aspect majeur de nos travaux est l'introduction des considérations d'équité dans le problème d'allocation des débits.



# Chapitre 6

## Allocation jointe de débit et de puissance : cas général

### 6.1 Introduction

Les opérateurs de réseaux de télécommunications cherchent à maximiser le nombre d'utilisateurs pouvant être servis sur un réseau et pour une qualité donnée, cette qualité pouvant être variable selon les utilisateurs. Dans les réseaux de téléphonie de type CDMA, chaque utilisateur est une source de bruit pour les autres. Si le nombre d'utilisateurs devient trop important, le bruit augmente et entraîne une chute de la qualité de la communication. Cependant, un contrôle de débit de transmission des mobiles peut permettre de contrôler et limiter ce bruit et d'améliorer les performances du réseau. Pour ces raisons, il est fondamental pour un opérateur d'optimiser la gestion des débits des mobiles ainsi que d'optimiser l'admission des utilisateurs dans un réseau CDMA de manière à maximiser leur nombre tout en assurant une qualité de service (QoS) donnée.

#### 6.1.1 Pourquoi une allocation jointe entre débit et puissance ?

Les systèmes peuvent en effet désormais utiliser des taux différents pour les transferts de données, les applications temps réel de voix et de vidéo en choisissant un codec approprié. A titre d'exemple, l'UMTS utilisera le codec AMR (Adaptive Multi-rate) qui offre huit taux de transmission différents entre 4.75 kbps et 12.2 kbps qui peuvent être dynamiquement changés toutes les 20 msec.

Le débit que l'on peut utiliser est limité par la qualité du signal radio comme nous le verrons dans la sous-section 6.3.3. Or cette qualité est une fonction des puissances d'émission des différentes sources. Pour en comprendre le principe, je propose de considérer une analogie avec les signaux sonores. Considérons une salle fermée dans laquelle plusieurs personnes parlent simultanément. On entend d'autant mieux son interlocuteur que celui parle fort par rapport aux autres. De plus, dans une telle situation, moins on pourra entendre distinctement notre interlocuteur (à cause des autres personnes situées dans la salle), plus on devra lui demander de se répéter et donc moins le débit d'information (utile) sera grand.

On peut remarquer que, pour ce faire mieux entendre, chacun a une tendance naturelle à hausser la voix. Seulement, lorsque chacun adopte ce comportement, le volume total de

bruit augmente et le signal n'est pas plus clair pour autant.

Il en est de même pour les communications par voix radio : la puissance d'émission a deux effets opposés sur la qualité du lien. En effet, en augmentant la puissance d'émission d'un mobile, on renforce la qualité de sa transmission, et on dégrade en même temps celle de tous les autres.

Cette simple observation nous pousse à considérer une allocation jointe des débits et des puissances. A l'heure actuelle, on utilise des mécanismes d'allocation de débits et l'on adapte ensuite les puissances des différents émetteurs afin de pouvoir utiliser ces débits. Ces allocations sont très largement sous-optimales au sens où l'allocation de débit ainsi obtenue n'est pas Pareto-optimale. Par un contrôle joint des ressources judicieux, on peut donc espérer pouvoir augmenter globalement les performances du système.

Un problème déjà étudié est celui du choix des taux de transmission de chaque individu afin de maximiser le débit total du système dans des configurations assez simples, cf [62, 92]. Il présente cependant un grave inconvénient, au niveau de l'équité. En effet, on sait que les mobiles dont les signaux subissent de fortes atténuations doivent augmenter leur puissance d'émission afin de pouvoir communiquer avec la station de base. En faisant, ils créent davantage d'interférences avec les autres terminaux. Ainsi, lorsque l'on cherche à maximiser le débit total du système, pénalise fortement les mobiles dont les signaux sont fortement dégradés.

Ce raisonnement justifie ainsi tout à fait l'utilisation de critères d'équité dans les réseaux cellulaires. De plus, du fait de la diversité des applications utilisées nous proposons de les utiliser non aux débits directement, mais aux utilités qui leur sont associées.

### 6.1.2 Un réseau, différents liens.

Je donne ici une brève introduction au fonctionnement des réseaux cellulaires. La zone géographique couverte par un tel système est divisée en *cellules*. Elles sont modélisées en général par des disques ou des hexagones. A l'intérieur de chacune d'elles se trouve une *station de base*. Les stations de base sont reliées entre elles dans un réseau (filaire). Lorsqu'un mobile veut se joindre au réseau, il communique par voie radio avec la station de base de la cellule dont il dépend. On considère quelquefois qu'une station de base est en fait un regroupement de trois antennes, chacune couvrant un angle de  $120^\circ$ , comme illustré à la figure 6.1. On parle alors de trois *secteurs*.

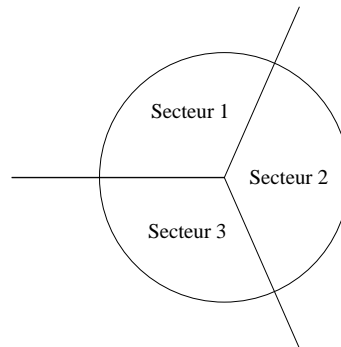


FIG. 6.1 – Un exemple de cellule avec 3 secteurs

On considère alors deux types de liens entre le mobile et la station de base. Lorsque le mobile envoie des signaux à la station, on parle de *lien montant* (uplink). Dans le sens contraire, on parle de *lien descendant* (downlink).

Considérons pour finir le petit exemple de la figure 6.2. Ce réseau se compose de deux cellules, numérotées 1 et 2 dans lequel évoluent 3 terminaux, représentés par des formes géométriques différentes. On voit que le premier mobile est dans la cellule 1, il sera donc

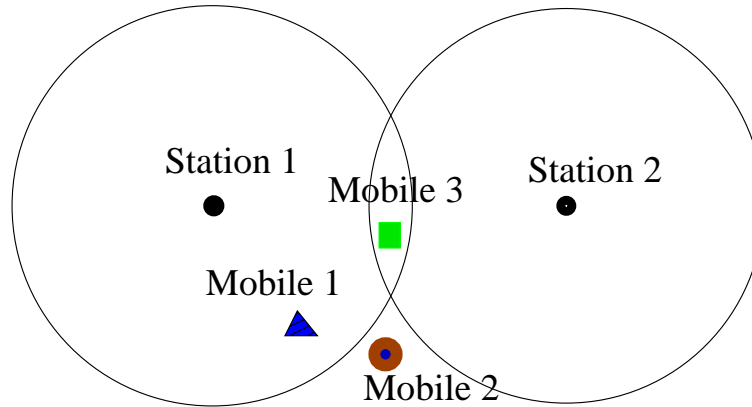


FIG. 6.2 – Un exemple de petit réseau sans fil avec 2 cellules

couvert par la station de base correspondante (représentée par un disque noir au centre de la cellule). Le deuxième mobile se situe en dehors de la zone de couverture, le réseau cellulaire ne pourra donc pas accepter de communications avec lui. Reste alors le troisième terminal. On voit qu'en fait, les zones de couvertures des zones se recouvrent légèrement, afin d'éviter des "trous" dans la couverture. De plus, on comprend aisément qu'en limite de couverture la réception soit mauvaise, car les mobiles sont alors dans les positions les plus éloignées des stations. Alors, quand un terminal se trouve comme le terminal 3, c'est-à-dire grossièrement "entre deux cellules", il peut communiquer avec deux stations, afin d'augmenter les chances de succès de la communication. On parle de "handover", ou *macrodiversité*. Nous y reviendrons dans la section 7.3 du prochain chapitre.

Nous allons montrer un modèle mathématique qui peut représenter les contraintes d'un système cellulaire de troisième génération. Ce modèle est général dans la mesure où il est adapté à la fois aux liens montants et descendants, et pour des mobiles situés au sein d'une cellule, comme le terminal 1 ou en limite de couverture, comme le mobile 3.

### 6.1.3 La suite...

La structure du chapitre est la suivante. Dans la prochaine section, nous présentons globalement le but de notre étude et montrons comment elle se situe par rapport aux travaux déjà réalisés dans le domaine. Nous introduisons alors dans la Section 6.3 notre modèle et les notations que nous utiliserons dans la suite. Nous analysons alors le système dans la Section 6.4 et exhibons quelques propriétés intéressantes. Notamment, nous prouvons que pour chaque allocation de débit correspond un vecteur de puissance minimal, que nous explicitons. De plus, nous donnons une condition nécessaire et suffisante pour qu'un vecteur

de taux soit acceptable, c'est-à-dire qu'il puisse être réalisé dans le système moyennant le choix judicieux du vecteur de puissance.

## 6.2 Discussion générale sur l'intérêt de notre proposition

Nous nous intéressons ici à la question de comment allouer des taux de transmission de façon à la fois optimale et équitable. Comme nous l'avons vu dans notre introduction sur les concepts d'équité :

- Pour les faibles valeurs du paramètre d'équité  $\alpha$  on tend vers des équilibres qui maximisent la somme des allocations. On tend dans ce cas vers la maximisation du débit total, qui sacrifie éventuellement l'accès au réseau à certains utilisateurs,
- Au contraire, pour les fortes valeurs de  $\alpha$ , on tend vers une allocation max-min équitable dont l'allocation est la plus égalitaire possible. En fait, on peut remarquer en étudiant la formule d'équité (Equation 1.4 ou 1.8) que dès que  $\alpha \geq 1$ , les allocations obtenues minimisent le nombre de connexions qui se voient refuser l'accès au réseau.

Ainsi, selon le choix du paramètre d'équité, notre stratégie d'allocation favorisera le débit total ou les taux d'admissions ou encore n'importe quel compromis possible entre les deux.

Notre est ainsi de gérer en temps réel les ressources radio ainsi que l'admission des mobiles, tout en leur offrant une garantie de débit. L'intérêt a tout particulièrement été mis en évidence pour les applications temps-réel.

### 6.2.1 Inconvénients des techniques actuelles

Les constructeurs fournissent des équipements qui mettent en œuvre des algorithmes de contrôle du débit et le contrôle de l'admission. Les solutions actuellement adoptées pour accroître le taux de d'admission des mobiles dans un réseau CDMA ou W-CDMA sont basées sur l'adaptation du débit du mobile entrant, pour les mobiles utilisant des services qui peuvent supporter des débits variables. Les politiques d'admission peuvent être différentes. Par exemple on peut adapter le débit du mobile entrant en lui garantissant un débit donné (Alcatel). Cette politique induit des taux de rejet importants, et limite la capacité du réseau. On peut aussi adapter le débit du mobile entrant de manière à toujours l'admettre, même avec un débit nul (Nokia). Cette politique ne garantit pas de Qualité de Service (QoS). Enfin, ces stratégies sont sous-optimales au sens de Pareto car l'on ne touche pas au débits des terminaux déjà dans le système.

Ainsi, aucune technique existante ne permet d'explorer toute la gamme de stratégies de gestion dynamique de débits des mobiles dans le système, ni d'admission dans le réseau située entre un taux de rejet important (pour une QoS garantie), et un taux de rejet faible (avec une QoS non garantie).

Les stratégies optimales-équitables introduites dans le contexte des réseaux fixes [7] ne peuvent pas être appliquées directement aux réseaux CDMA-WCDMA puisqu'elles ne

permettent pas la prise en compte des spécificités de ces réseaux, en particulier de l'interface radio.

### 6.2.2 But des travaux

Par opposition aux méthodes existantes développées par les constructeurs (par exemple Alcatel et Nokia), nous désirons avoir une gestion plus souple du réseau en fonction de critères de QoS qui réponde en temps réel aux choix de l'opérateur. Cela se traduit par la re-négociation, à l'admission d'un mobile, des débits de certains mobiles judicieusement choisis parmi les mobiles présents dans le réseau. Cette re-négociation permettra alors au réseau d'allouer les débits de manière Pareto-optimale pour la définition d'équité choisie. Elle rend possible une gestion globale de l'allocation des débits et non mobile par mobile.

Ce procédé permet donc un accroissement de la capacité du réseau (en terme de nombre de mobiles) par rapport aux stratégies basées sur la garantie d'un débit minimal et accroissement des débits pour les stratégies basées sur une QoS garantie.

Enfin, une conséquence de cette méthode est la suivante : un opérateur peut valoriser les débits associés à des services, et proposer une équité basée sur la valeur des services proposés.

Nous avons montré que notre méthode, dans le cas de l'UMTS, peut en pratique être implantée aux RNC (Radio Network Controller), sur les Node B, voire les OMC.

Si  $D_t$  représente le débit courant du mobile, et  $D_{tmin}$  la valeur minimale autorisée pour ce débit, on a l'algorithme de la figure 6.3.

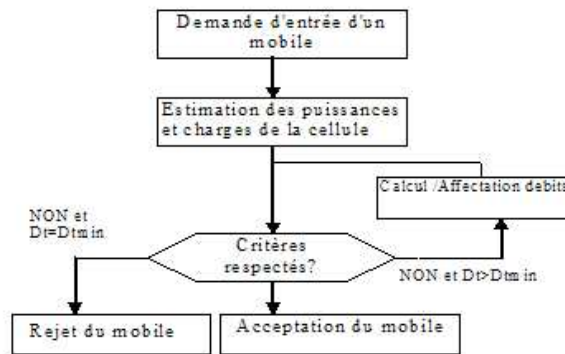


FIG. 6.3 – Algorithme de contrôle d'admission.

## 6.3 Le modèle

On s'intéresse à développer un cadre général qui peut s'appliquer à la fois au lien montant et descendant. Notre point de départ fut le modèle de [62] que nous avons grandement élargi.

Le réseau se compose d'un ensemble  $\mathcal{N}$  de connections. Soit  $\overrightarrow{\mathcal{E}}$  l'ensemble des émetteurs et  $\overleftarrow{\mathcal{E}}$  celui des récepteurs. Une connection  $n \in \mathcal{N}$  est une paire  $\{s, t\}$  avec  $s \in \overrightarrow{\mathcal{E}}$  et  $v \in \overleftarrow{\mathcal{E}}$ .



Si l'on étudie le lien montant, alors  $\mathbf{s}$  désigne un mobile et  $\mathbf{t}$  une station de base. Les rôles sont inversés dans le lien descendant.

Je vais montrer dans la suite que l'optimisation équitable et optimale est un problème joint entre deux ressources : le débit ( $\vec{r}$ ) et la puissance ( $\vec{p}$ ) allouée.  $\vec{r}$  et  $\vec{p}$  sont des vecteurs de taille  $|\mathcal{N}|$  (le nombre de liens  $\{\mathbf{s}, \mathbf{t}\}$ ) tel tout élément  $r_n$  (respectivement  $p_n$ ) représente le débit instantané (respectivement la puissance) associé à la connection  $n, n \in \mathcal{N}$ .

Je présente dans la suite de cette section la fonction d'optimisation correspondant à l'allocation équitable des débits dans le contexte des réseaux mobiles (sous-section 6.3.1) puis les contraintes existantes sur les vecteurs de puissances (sous-section 6.3.2). Nous verrons dans la suite (sous-sections 6.3.3 et 6.3.4) comment les débits et les puissances allouées sont liées. Enfin, je résumerai le problème d'optimisation d'allocation équitable et optimale utilisé.

### 6.3.1 Allocation des débits

La qualité de la liaison perçue par l'utilisateur est fonction de son débit alloué. C'est donc lui que nous allons chercher à allouer équitablement. L'application du critère d'équité pondéré nous dit alors que l'allocation optimale-équitable est la solution de :

$$\max_{\vec{r}} \sum_{n \in \mathcal{N}} \frac{r(n)}{1 - \alpha}.$$

On suppose dans la suite que la connection  $n = \{\mathbf{s}, \mathbf{t}\}$  peut utiliser n'importe quelle valeur de débit entre une valeur minimale garantie  $MR_n$  et une valeur maximale  $PR_n$ . Ceci peut être obtenu si un mode paquet est utilisé avec l'ordonnancement adéquat (voir par exemple [12, 37] et les références incluses).

### 6.3.2 Modèle de puissance

Chaque source  $\mathbf{s}$  peut transmettre avec une puissance maximale totale  $\overline{p_{tot,s}}$ . Il n'existe pas de minimum de puissance. Ainsi, la puissance  $p_{tot,s}$  qui lui sera effectivement allouée, appartient à l'intervalle  $[0, \overline{p_{tot,s}}]$ .

Il se peut que la source émette à destination de plusieurs récepteurs. Par exemple, dans le cas descendant chaque station de base envoie des signaux à l'ensemble des terminaux de la cellule. On appelle alors *canal* chaque signal de la source. Nous noterons dans la suite  $T_s$  l'ensemble des destinations d'une source et  $p_{\{\mathbf{s}, \mathbf{t}\}}$  la puissance du signal émis en vu de  $\mathbf{t}$ . Enfin,  $p_{NC,s}$  représente les canaux non contrôlés en puissance. Ainsi :

$$p_{tot,s} = \sum_{t' \in T_s} p_{s,t'} + p_{NC,s}. \quad (6.1)$$

Le vecteur de puissance maximale s'écrit alors  $\overrightarrow{P_{tot}}$ .

### 6.3.3 Lien entre puissance et débit

Considérons désormais un récepteur  $t$ . Le signal qu'il reçoit de la source  $s$  a une certaine qualité, que l'on mesure en quantité de rapport **SIR** (Signal over Interference Ratio). C'est le rapport entre la puissance utile de signal qu'il reçoit (c'est-à-dire le signal qui lui est envoyé par la source) et la somme des puissances des signaux interférents. Ces signaux interférents représentent les signaux considérés comme parasites par le récepteur. Je note dans la suite  $SIR(\{s, t\})$  le **SIR** exprimé au niveau du récepteur  $t$  du signal émis par  $s$  (exprimé en terme de puissance).

On peut alors montrer qu'un récepteur  $t$  peut correctement décoder le signal issu de l'émetteur  $s$  si la relation suivante est vérifiée (je note  $n$  la connection  $\{s, t\}$ ) :

$$\delta_n \leq \frac{SIR(n)}{r(n)}, \quad (6.2)$$

impliquant que  $\delta_n r(n) \leq SIR(n)$ .<sup>1</sup>  $\delta_{\{s,t\}}$  représente la valeur minimale de  $SIR_{\{s,t\}}$  par bit transmis et par seconde acceptable. C'est quelquefois une caractéristique du récepteur, donnée par le constructeur.

La relation 6.2 lie donc les vecteurs de débit et de puissance.

### 6.3.4 Expression analytique du SIR

On explicite ici la relation entre  $\vec{r}$  et  $\vec{p}$ . En fait, on voit que pour chaque vecteur  $\vec{r}$  correspond un ensemble de vecteurs  $\vec{p}$  satisfaisant la contrainte du SIR.

Les liens et donc les problèmes d'allocations dans les liens montant et descendants sont indépendants<sup>2</sup>, nous pouvons donc les résoudre de façon indépendante. Soit  $m$  un mobile et  $b$  une station de base, alors  $\exists s \in \vec{\mathcal{E}}, t \in \overleftarrow{\mathcal{E}}$  tel que :

$$(m, b) = \begin{cases} (s, t) & \text{Si l'on s'intéresse au problème du lien montant.} \\ (t, s) & \text{Si l'on s'intéresse au problème du lien descendant.} \end{cases}$$

Comme l'on considère les problèmes liés aux liens montants et descendant de façon indépendante, on ne s'inquiétera pas de l'ordre des variables dans l'expression des connections. Ainsi, on notera indifféremment une connection sous la forme  $\{m, b\}$  ou  $\{b, m\}$ . L'indice ( $m$  ou  $b$ ) correspondant au nœud source ou destination sera rendu évident par le contexte.

Nous montrerons dans le chapitre suivant que dans le cas du lien montant comme du lien descendant et de la macrodiversité, on peut toujours écrire le rapport signal sur bruit

<sup>1</sup>Pour obtenir cette équation, nous avons implicitement supposé que le rapport entre énergie par bit et densité spectrale de puissance d'interférence, notée  $(E_b/I_o)_m$  ne dépend pas du taux de transmission  $r(m)$ . C'est une hypothèse courante dans la littérature de la modélisation, voir par exemple [62]. En pratique, cependant il peut dépendre de  $r(m)$ , voir par exemple [50, p. 151, 222, 239]. Mais, comme nous voyons depuis [50, Fig. 10.4, p. 222], il est proche d'une constante sur un grand ensemble de valeur de débit. Par exemple, entre 16Kbps et 256Kbps, la variation maximale autour de la valeur médiane est inférieure à 20%. Nous proposons donc de prendre comme valeur  $(E_b/I_o)_m$  sa moyenne sur l'intervalle  $[MR_m, PR_m]$ . Cependant, si une dépendance exacte est disponible analytiquement, elle peut être incluse dans notre modèle.

<sup>2</sup>Car les fréquences utilisées dans les deux sens sont différentes.

sous la forme :

$$SIR(\{m, b\}) = \frac{p'_{\{m,b\}}}{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m, m' \neq m} p'_{\{m',b\}}} \quad (6.3)$$

où :

- $N(\{m, b\}) > \mathbf{0}$  est la puissance du bruit ambiant au niveau du récepteur. Il représente non seulement le bruit thermique mais également les interférences radio dues aux canaux non contrôlés en puissance.
- $C$  est une constante multiplicative. Elle peut être par exemple le facteur d'orthogonalité.
- $p'_{\{s,t\}}$  est une fonction de la puissance  $p_{\{s,t\}}$  de la source de la transmission  $s$  pour la destination  $t$ .
- Enfin, soit  $m$  le mobile considéré (qui peut être  $s$  ou  $t$  selon que l'on se trouve dans le cas du lien montant ou descendant), alors  $\mathcal{M}_m$  représente l'ensemble des terminaux se situant dans la même cellule que  $m$ .

Les valeurs de  $C$ ,  $N$  et  $p'$  seront explicitées dans le chapitre suivant pour chaque cas (lien montant, descendant et macrodiversité).

### 6.3.5 Résumé : formulation du problème d'équité

Nous résumons ici l'expression du problème d'allocation équitable et optimale des ressources dans les réseaux mobiles.

Notre problème est donc de

$$\text{Trouver } (\vec{p}, \vec{r}) \in \Pi^c \text{ qui maximise } \sum_{n \in \mathcal{N}} \frac{r(n)^{1-\alpha}}{1-\alpha}.$$

L'ensemble de contraintes  $\Pi^c = (P, R)$  est défini par :

$$\forall n \in \mathcal{N}, n = \{m, b\} \left\{ \begin{array}{l} \delta_{\{m,b\}} r(\{m, b\}) \leq \frac{p'_{\{m,b\}}}{N(\{m, b\}) + C \sum_{\substack{m' \in \mathcal{M}_m, \\ m' \neq m}} p'_{\{m',b\}}}, \\ MR_n \leq r(n) \leq PR_n, \\ 0 \leq p_{tot,s} \leq \overline{p_{tot,s}} \\ \text{avec } s = \begin{cases} m & \text{dans le lien montant,} \\ b & \text{dans le lien descendant.} \end{cases} \end{array} \right. \quad (6.4)$$

On peut ainsi voir la difficulté de ce problème d'optimisation par rapport à celui associé aux réseaux terrestres. La différence est due à deux facteurs.

- L'allocation est un problème joint entre deux variables  $\vec{p}$  et  $\vec{r}$ ,
- Le système de contraintes est plus complexe (rappelons-nous en effet que dans le cadre des réseaux terrestres, les contraintes pouvaient s'écrire sous la forme d'un système de contraintes linéaires).

Nous avons donc dans cette section proposé un modèle représentant les contraintes d'allocation dans les réseaux mobiles de type CDMA. Nous justifierons l'adéquation de ce modèle à chacun des liens dans le chapitre suivant.

Dans la section suivante, nous étudions le système général que nous venons d'explicitier et mettons en exergue quelques propriétés fondamentales.

## 6.4 Propriétés du système

Nous étudions dans cette section les propriétés du système de contraintes 6.4. Nous commençons dans la sous-section suivante par donner une formulation équivalente du système de contraintes. L'intérêt est de pouvoir ensuite énoncer une condition simple pour assurer concavité de la fonction objectif tout en ayant un système aux contraintes convexes.

Dans la seconde sous-section nous montrons tout d'abord l'existence, pour tout vecteur d'allocation de débit acceptable d'un vecteur de puissance minimal dans l'ensemble des vecteurs de puissance satisfaisant les contraintes du système. Dans un second temps, nous explicitons la forme de ce vecteur de puissance et nous déduisons trois conditions nécessaires et suffisantes pour qu'un vecteur de débit soit acceptable. Ainsi, notre ensemble de contraintes  $\Pi^c$  se ramène à un ensemble de dimension 1.

### 6.4.1 Problème équivalent et concavité de la fonction objectif

Nous proposons ici (Sous-section 6.4.1.1) une formulation équivalente du problème basée sur un simple changement de variables. L'intérêt est de rendre alors le système de contraintes convexe. En revanche, la fonction objectif obtenue n'est plus alors concave. Nous verrons cependant dans la Sous-section 6.4.1.2 que l'on peut émettre une hypothèse simple pour s'en assurer.

#### 6.4.1.1 Problème équivalent.

**Lemme 6.1** *Le problème est équivalent à trouver  $(\vec{p}, \vec{\rho}) \in \hat{\Pi}^c = (P, \rho)$  qui maximise*

$$Z(\vec{\rho}) := \sum_{n \in \mathcal{N}} \frac{1}{1 - \alpha} \left( \frac{\rho(n)}{1 - \delta_n \rho(n)} \right)^{1 - \alpha} \quad \text{où } \hat{\Pi}^c = (P, \rho) \text{ est donné par :}$$

$$\forall n \in \mathcal{N}, n = \{m, b\}, \left\{ \begin{array}{l} \delta_{\{m, b\}} \rho(\{m, b\}) \leq \frac{p'_{\{m, b\}}}{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}, \\ \frac{MR_n}{1 + \delta_n MR_n} \leq \rho(n) \leq \frac{PR_n}{1 + \delta_n PR_n}, \\ 0 \leq p_{tot, s} \leq \overline{p}_{tot, s} \\ s = \begin{cases} m & \text{dans le lien montant,} \\ b & \text{dans le lien descendant.} \end{cases} \end{array} \right. \quad (6.5)$$

**Démonstration :** Soit  $\vec{\rho}$  le vecteur de taille  $|\mathcal{N}|$  tel que  $\forall n \in \mathcal{N}, \rho(n) = \frac{r(n)}{1 + C\delta_n r(n)}$ .

Nous notons que  $\delta_n \rho(n) = 1 - \frac{1}{1 + \delta_n r(n)}$  et faisons simplement le changement de variable de  $\vec{r}$  à  $\vec{\rho}$  dans le système (6.4). ■

#### 6.4.1.2 Concavité de la fonction objectif

Une fois le changement de variable effectué, nous pouvons énoncer une condition simple sur  $\vec{\rho}$  assurant la concavité de la fonction objectif  $Z$ .

**Lemme 6.2** *La fonction objectif est concave si pour tout  $\vec{\rho}$  dans l'ensemble des solutions faisables nous avons :  $\forall n \in \mathcal{N}, 2\delta_n \rho(n) \leq \alpha$ .*

**Démonstration :** Notons que le dénominateur  $1 - \delta_n \rho(n)$  est positif ou nul sur l'ensemble des solutions faisables (d'après la deuxième inégalité du système (6.5)). Pour déterminer si la fonction objectif est concave, nous dérivons deux fois  $Z$  par rapport à  $\rho(\mathbf{n})$ , pour toute connection  $\mathbf{n}$  et obtenons

$$\frac{\partial^2 Z(\vec{\rho})}{\partial \rho(n)^2} = \frac{2\delta_n \rho(n) - \alpha}{(1 - \delta_n \rho(n))^3 \rho(n)} \left( \frac{\rho(n)}{1 - \delta_n \rho(n)} \right)^{-\alpha}.$$

Si  $2\delta_n \rho(n) \leq \alpha$ , alors cette dérivée est négative ou nulle pour tout  $\rho(n)$  faisable. La fonction objectif est alors concave. ■

**Remarque 6.1**  $\alpha \geq 2$  est une condition suffisante pour que la fonction objectif soit concave. Cette condition peut ensuite être relaxée. Soit

$$w = \max_{n \in \mathcal{N}} \frac{\delta_n PR_n}{1 + \delta_n PR_n}.$$

Alors une condition suffisante pour que la **fonction objectif soit concave est  $\alpha \geq 2w$** . Assez souvent  $w$  est proche de zéro (voir à ce sujet la discussion avant le lemme 1 dans [11]).

Dans la suite, nous appellerons *vecteur de taux de transmission acceptable* tout vecteur  $\vec{r}$  (respectivement  $\vec{\rho}$ ) qui accepte au moins une allocation de puissance  $\vec{p}$  satisfaisant les contraintes (6.4) (respectivement (6.5)). L'ensemble des vecteurs  $\vec{p}$  acceptables correspondant est alors noté  $P(\vec{r})$ .

### 6.4.2 Propriétés des vecteurs de taux acceptables

Le but de cette sous-section est d'exhiber des propriétés des vecteurs de taux acceptables. Le but étant évidemment de découpler les équations en  $\vec{p}$  et en  $\vec{\rho}$ .

Dans la première sous-section nous montrons que pour un vecteur de transmission acceptable  $\vec{r}$  (ou de façon équivalente  $\vec{\rho}$ ) correspond un vecteur de puissance qui soit minimal composante par composante dans  $P(\vec{r})$ .

Nous montrons dans une seconde sous-section une condition nécessaire et suffisante pour qu'un vecteur de débit soit acceptable et explicitons alors la forme du vecteur minimal de puissance correspondant.

#### 6.4.2.1 Vecteur de puissance acceptable minimal

Commençons par remarquer que :

**Lemme 6.3** *Soit  $\mathbf{n}$  une connection,  $\mathbf{n} = \{m, b\}$ . Alors, pour tout  $\vec{\rho}$  acceptable*

$$C \sum_{m' \in \mathcal{M}} \delta_{\{m', b\}} \rho(\{m', b\}) < 1.$$

**Démonstration :** La preuve est évidente. En effet, considérons la première inégalité du système (6.5) :

$$\forall n \in \mathcal{N}, n = (\{m, b\}), \delta_{\{m, b\}} \rho(\{m, b\}) \leq \frac{p'_{\{m, b\}}}{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}.$$

Comme  $N(\{m, b\}) > 0$ , alors  $\delta_{\{m, b\}} \rho(\{m, b\}) < \frac{p'_{\{m, b\}}}{C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}$ . Finalement :

$$C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\}) < \frac{C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}{C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}} = 1.$$

■

Nous pouvons en déduire le lemme suivant :

**Lemme 6.4** *Soit  $\vec{\rho}$  un vecteur acceptable. Considérons le système linéaire suivant :*

$$\delta_{\{m, b\}} \rho(\{m, b\}) = \frac{p'_{\{m, b\}}}{N(\{m, b\}) + C \sum p'_{\{m', b\}}}. \quad (6.6)$$

Alors, ce système linéaire de  $|\mathcal{N}|$  équations et  $|\mathcal{N}|$  variables admet une et une seule solution. Nous l'appelons  $\vec{\mathbb{P}}$ . On a évidemment  $(\vec{\mathbb{P}}, \vec{\rho}) \in \hat{\Pi}^c$ .

**Démonstration :** Il suffit de montrer que les  $|\mathcal{N}|$  équations sont linéairement indépendantes. Elles peuvent être réécrites ainsi :

$$\forall n \in \mathcal{N}, n = \{m, b\} \delta_n \rho_n N(n) = p_n - C \delta_n \rho_n \sum_{m' \in \mathcal{M}_m} p_{\{m', b\}}. \quad (6.7)$$

Supposons que l'on numérote les connections  $\mathbf{n}$  de 1 à  $|\mathcal{N}|$ . Alors on peut écrire les équations sous forme matricielle  $\mathbf{A}\mathbf{X} = \mathbf{Y}$  avec

- $\mathbf{Y} = (\delta_i \rho_i N(i))_i, i \in [1, |\mathcal{N}|]$ ,
- $\mathbf{X} = (p'_i)_i, i \in [1, |\mathcal{N}|]$
- $\mathbf{A} = Id_M - \mathbf{B}$
- $Id_M$  la matrice d'identité de taille  $M$

–  $B$  la matrice :

$$\begin{pmatrix} b_1 & \dots & b_1 \\ b_2 & \dots & b_2 \\ & \dots & \\ b_M & \dots & b_M \end{pmatrix}. \quad (6.8)$$

– et  $\forall i, b_i = C\delta_i\rho_i$ .

Alors, si  $\mathbf{U}$  est un vecteur propre de  $\mathbf{A}$  associé à la valeur propre  $\lambda$ ,  $AU = \lambda U = U - BU$ . Ainsi  $\mathbf{U}$  est un vecteur propre de  $\mathbf{B}$  avec la valeur propre  $\mathbf{1} - \lambda$ . Mais  $\text{rank}(\mathbf{B}) = \mathbf{1}$  et  $\text{trace}(\mathbf{B}) = \sum_i b_i$ . Alors  $\mathbf{A}$  a uniquement deux valeurs propres qui sont :  $\mathbf{1}$  et  $\mathbf{1} - C \sum_i \delta_i \rho_i$ .

Comme  $1 - C \sum_i \delta_i \rho_i \neq 0$  par le Lemme (6.3), alors  $A$  est inversible et donc  $\mathbb{P}$  existe et est unique. ■

**Proposition 6.1** *A chaque vecteur de taux de transmission fixé  $\vec{\mathbf{r}}$  (respectivement  $\vec{\boldsymbol{\rho}}$ ) correspond un unique vecteur de puissance minimale  $\vec{\mathbf{P}}^{\text{min}}$  composante par composante dans l'ensemble  $\mathbf{P}(\vec{\mathbf{r}})$ . De plus  $\vec{\mathbf{P}}^{\text{min}} = \vec{\mathbb{P}}$ , avec  $\vec{\mathbb{P}}$  tel que défini dans le lemme 6.4.*

**Démonstration :** Comme les problèmes (6.4) et (6.5) sont équivalents, il nous suffit de prouver la proposition dans le premier cas.

Nous étendons la preuve du Lemme 1 de [11] qui ne considère qu'une seule cellule. Supposons qu'il existe un vecteur d'allocation de puissance faisable  $(P')^0$  satisfaisant les contraintes (6.4). Nous construisons une suite de vecteurs de puissance  $(P')^i$  défini par  $\forall i \in \mathbb{N}, \forall n \in \mathcal{N}, n = \{m, b\}, (p'_{\{m,b\}})^{i+1} = \delta_{\{m,b\}} r(\{m, b\}) \left( N(\{m, b\}) + C \sum_{\substack{m' \in \mathcal{M}_m \\ m' \neq m}} (p'_{\{m',b\}})^i \right)$ . Nous avons  $0 \leq (p'_{\{m,b\}})^{i+1} \leq (p'_{\{m,b\}})^i$ . Donc la suite est décroissante (composante par composante) et minorée (chaque vecteur est en effet borné - composante par composante - par le vecteur nul). Elle converge donc vers une allocation  $\vec{\mathbf{P}}^{\text{min}}$  satisfaisant  $\forall i \in \mathbb{N}, \vec{\mathbf{P}}^{\text{min}} \leq (P')^i$  composante par composante et

$$\forall n \in \mathcal{N}, n = \{m, b\}, P'_{\{m,b\}}^{\text{min}} = \delta_{\{m,b\}} r(\{m, b\}) \left( N(\{m, b\}) + C \sum_{\substack{m' \in \mathcal{M}_m \\ m' \neq m}} P'_{\{m',b\}}^{\text{min}} \right).$$

Nous avons donc montré dans cette sous-section que si  $\vec{\mathbf{r}}$  est un vecteur acceptable alors il existe un vecteur  $\mathbb{P}$  minimal dans l'ensemble des vecteurs de puissance acceptables. De plus, ■

– Le vecteur est minimal composante par composante.

– Il vérifie

$$\forall n \in \mathcal{N}, n = \{m, b\}, P'_{\{m,b\}}{}^{min} = \delta_{\{m,b\}} r(\{m, b\}) \left( N(\{m, b\}) + C \sum_{\substack{m' \in \mathcal{M}_m \\ m' \neq m}} P'_{\{m',b\}}{}^{min} \right).$$

Dans la dernière sous-section, nous allons donner une expression analytique du vecteur  $\overrightarrow{P'{}^{min}}$  (Lemme 6.6) et une nouvelle expression de l'ensemble de contraintes  $\mathbf{\Pi}^c$  ne dépendant que de  $\overrightarrow{\rho}$ . Nous pouvons donc finalement écrire le problème d'allocation jointe des débits et des puissances dans les réseaux mobiles de type CDMA comme un problème d'optimisation d'une seule variable,  $\overrightarrow{\rho}$ .

#### 6.4.2.2 Conditions nécessaires et suffisantes sur les vecteurs de débits et puissances minimales associées

**Lemme 6.5** Soit  $\overrightarrow{\rho}$  un vecteur de taux acceptable fixé. L'ensemble des allocations de puissances faisables  $\mathbf{P}'$  satisfait  $\forall \mathbf{p}' \in \mathbf{P}', \forall n \in \mathcal{N}, n = \{m, b\}$  :

$$\begin{aligned} \delta_n \rho(n) \left[ C \sum_{m' \in \mathcal{M}_m} \delta_{\{m',b\}} \rho(\{m', b\}) [N(\{m', b\}) - N(\{m, b\})] + N(n) \right] \\ \leq \left( 1 - C \sum_{m' \in \mathcal{M}_m} \delta_{\{m',b\}} \rho(\{m', b\}) \right) p'_{\{m,b\}}. \end{aligned} \quad (6.9)$$

**Démonstration :** Je considère une connexion  $\mathbf{n} = \{\mathbf{m}, \mathbf{b}\}$  et réécrit la première inégalité du système (6.4) :

$$p'_{\{m,b\}} \geq \delta_{\{m,b\}} \rho(\{m, b\}) \left( N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} p'_{\{m',b\}} \right). \quad (6.10)$$

Il vient, en sommant sur tous les mobiles de l'ensemble  $\mathcal{M}_m$  :

$$\left( \sum_{m'' \in \mathcal{M}_m} p'_{\{m'',b\}} \right) \cdot \left( 1 - C \sum_{m'} \delta_{\{m',b\}} p'_{\{m',b\}} \right) \geq \sum_{m' \in \mathcal{M}_m} \delta_{\{m',b\}} \rho(\{m', b\}) N(\{m', b\}).$$

Nous combinons finalement ceci avec (6.10) afin d'obtenir l'inégalité (6.9). ■

De la proposition 6.1 et du lemme 6.5, nous pouvons finalement obtenir l'expression du vecteur minimal de puissance acceptable :

**Lemme 6.6** Pour un vecteur  $\overrightarrow{\mathbf{r}}$  acceptable (respectivement  $\overrightarrow{\rho}$ ),  $\overrightarrow{P'{}^{min}}$  est donné par :

$$P'_{\{m,b\}}{}^{min} = \delta_{\{m,b\}} \rho(\{m, b\}) \frac{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} \delta_{\{m,b\}} \rho(\{m, b\}) [N(\{m, b\}) - N(\{m, b\})]}{1 - C \sum_{m' \in \mathcal{M}_m} \delta_{\{m,b\}} \rho(\{m, b\})}. \quad (6.11)$$



Nous concluons finalement que :

**Théorème 6.1** *Un vecteur de taux  $\vec{\rho}$  est acceptable si et seulement si il satisfait les trois conditions suivantes :*

$$(C1) \quad \forall n \in \mathcal{N}, n = \{m, b\}, \frac{MR_{\{m,b\}}}{1 + \delta_{\{m,b\}}MR_{\{m,b\}}} \leq \rho(\{m, b\}) \leq \frac{PR_{\{m,b\}}}{1 + \delta_{\{m,b\}}PR_{\{m,b\}}},$$

$$(C2) \quad C \sum_{n \in \mathcal{N}} \delta_n \rho(n) < 1,$$

$$(C3) \quad 0 \leq p_{tot,s}^{min} \leq \overline{p_{tot,s}} \text{ avec } s = \begin{cases} m & \text{dans le lien montant,} \\ s & \text{dans le lien descendant} \end{cases} \text{ et } \overrightarrow{p_{tot}^{min}} \text{ associé à } \overrightarrow{P'^{min}}.$$

**Démonstration :** Si  $\vec{\rho}$  est acceptable, alors (C1) est vérifié. La proposition 6.1 affirme que si  $\overrightarrow{P'^{min}}$  est une solution du système alors (C2) et (C3) sont également satisfaites.

Réciproquement, si les conditions (C2) et (C3) sont satisfaites, alors  $\overrightarrow{P'^{min}}$  est un vecteur solution (il satisfait la deuxième et la troisième inégalité de (6.5)). Finalement, (C1) est la seconde inégalité de (6.5). ■

## Conclusion

Dans cette section, nous avons fourni un système d'équations équivalent au problème initial basé sur un changement de variables (Lemme 6.1). Nous avons exprimé une condition suffisante pour que la fonction objectif soit concave (Lemme 6.2 et Remarque 6.1).

Nous avons alors montré trois propriétés de ce système (Proposition 6.1, Lemme 6.6 et Théorème 6.1). En particulier nous avons montré que si un vecteur de taux est acceptable (c'est-à-dire si au moins une allocation faisable  $\vec{P}$  lui correspond), alors tous les vecteurs de puissance correspondants sont plus grands (composante par composante) que le vecteur de puissance faisable  $\overrightarrow{P'^{min}}$  donné par  $\forall n \in \mathcal{N}, n = \{m, b\}$ ,

$$P'_{\{m,b\}}^{min} = \delta_{\{m,b\}} \rho(\{m, b\}) \frac{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} \delta_{\{m,b\}} \rho(\{m, b\}) [N(\{m, b\}) - N(\{m, b\})]}{1 - C \sum_{m' \in \mathcal{M}_m} \delta_{\{m,b\}} \rho(\{m, b\})}.$$

Il est intéressant de noter que pour chaque vecteur  $\overrightarrow{P'^{min}}$  correspond un vecteur  $\overrightarrow{P_{tot}}$  unique. Nous avons finalement conclu par une condition nécessaire et suffisante pour qu'un vecteur de taux soit acceptable.

## Conclusion générale

Nous avons présenté dans ce chapitre le contexte du problème que nous nous proposons d'étudier, c'est-à-dire le problème d'allocation jointe des débits et des puissances dans un

réseau cellulaire de troisième génération de type UMTS. Nous désirons faire cette allocation en temps réel.

Dans de tels systèmes, les utilisateurs pourront avoir accès à une grande variété de services. Il est donc intéressant de pouvoir leur allouer des débits différents en fonction de leurs besoins. Cependant, dans les communications radio, le débit que l'on peut utiliser dépend de la qualité du lien radio, lui-même fonction de la puissance d'émission utilisée ainsi que de celle des autres émetteurs présents. Cela est très comparable à ce qui peut se passer dans une salle fermée avec un certain nombre de personnes communiquant simultanément. Lorsque que l'une d'elle s'exprime, son message est d'autant plus audible que la personne parle fort. Cependant, plus un (ou plusieurs) individu(s) élève(nt) la voix, plus il est difficile de suivre la conversation de quelqu'un d'autre.

De même dans les communications radio, la qualité d'un canal est d'autant meilleure que la puissance d'émission est grande. Cependant lorsque l'on augmente la puissance d'un signal, alors on augmente les interférences perçues par toutes les autres connexions et donc leur qualité diminue.

C'est pourquoi nous nous intéressons ici à une allocations jointe de puissance et de débit. Nous avons vu qu'actuellement on ne gère pas de façon globale les débits et les puissances des terminaux dans un réseau mobile. Ainsi les systèmes actuels ne sont pas optimaux.

Nous avons introduit dans ce chapitre un modèle mathématique représentant les contraintes présentes dans ces réseaux sans chercher à le justifier. Nous avons vu, par son étude, que pour chaque vecteur de débit que le système peut accepter correspond un vecteur de puissance minimale. Nous avons enfin présenté un système de conditions nécessaires et suffisantes pour qu'un vecteur de débit soit valide, c'est-à-dire qu'il puisse être utilisé dans le système conjointement à un vecteur de puissance.

Le but du chapitre suivant est de justifier l'adéquation de ce modèle mathématique aux contraintes réelles des systèmes cellulaires tant dans l'étude du lien montant que descendant.



# Chapitre 7

## Application du modèle aux différents liens et approximations possibles

Nous montrons dans ce chapitre comment le système que nous avons présenté peut être effectivement utilisé pour modéliser les contraintes de débit et de puissance dans les liens (montant et descendant) des réseaux CDMA tels que l'UMTS.

Nous étudierons dans les trois premières sections les cas des liens montant, descendant et le cas de la macro-diversité.

Nous avons vu dans le chapitre précédent que la fonction d'utilité  $Z$  n'était concave que pour des valeurs de  $\alpha$  supérieures à une certaine borne, que nous avons appelée  $w$ . Il serait donc intéressant de proposer quelques schémas d'approximation pour les faibles valeurs de  $\alpha$  et en particulier pour  $\alpha = 0$ . C'est l'objet de notre quatrième section. Nous présenterons, en toute fin de chapitre, quelques résultats numériques.

### 7.1 Cas du lien montant

Nous nous intéressons dans cette section à l'étude du lien montant qui concerne, rappelons-le aux signaux envoyés par les mobiles à destination des stations de base dont ils dépendent.

Nous faisons l'hypothèse ici que chaque mobile  $m$  envoie des messages (par voie radio) à une seule station de base,  $b(m)$ .

Dans la sous-section suivante, nous présentons le système correspondant au lien montant et les hypothèses effectuées et montrons que le modèle du chapitre précédent est adéquat. Dans la seconde, nous appliquons les résultats obtenus.

#### 7.1.1 Notations et hypothèses.

Considérons un mobile  $m$  dans le système. Il n'utilise qu'un seul canal de transmission, utilisé pour communiquer avec la station de base  $b(m)$ . Ainsi, sa puissance totale d'émission (définie dans l'équation 6.1) peut s'écrire :

$$p_{tot,m} = P_{\{m,b(m)\}}$$

Dans le canal radio, comme dans tout canal de transmission, le signal subit un affaiblissement. Dans les réseaux mobiles, ces pertes de puissance sont principalement dues aux obstacles se situant entre le terminal mobile et la station de base<sup>1</sup>. Dans les réseaux satellitaires nous verrons qu'au contraire le relief joue un rôle peu important et que les pertes sont principalement dues aux fortes distances séparant le terminal du satellite avec lequel il communique.

Notons  $g_{\{m,b(m)\}}$  le gain du lien entre la source  $m$  (le mobile) et la station de base destination  $b(m)$ . Cela peut s'interpréter de la façon suivante : si le mobile envoie un signal avec la puissance  $\alpha$  alors le signal reçu au niveau de la station de base sera  $\alpha g_{\{m,b(m)\}}$ . On a évidemment  $0 \leq g_{\{m,b(m)\}} \leq 1$ . Comme les principales causes de perte de puissance sont les obstacles physiques situés autour du terminal, on conçoit aisément que le gain du signal peut varier très vite avec le mouvement du mobile, c'est pourquoi nous devons supposer que les intervalles de temps dans lesquels nous faisons les mesures et réactualisons le partage équitable sont suffisamment faibles pour que  $g_{\{m,b(m)\}}$  soit constant.

Ainsi,  $g_{\{m,b(m)\}}P_{\{m,b(m)\}}$  est la puissance du signal reçu au niveau de la station de base en provenance du mobile  $m$ .

Le bruit perçu au niveau de la station de base est égal à la somme des puissances des signaux interférents et du bruit thermique, noté ici  $\nu_b$ . Les signaux interférents sont ceux reçus au niveau de la station de base considérée issus de tous les terminaux du système différents de  $m$ . Alors le  $SIR_{\{m,b\}}$  de la connexion  $\{m, b\}$  s'écrit :

$$SIR_{m,b} = \frac{g_{\{m,b\}}P_{\{m,b\}}}{\nu_b + \sum_{\substack{m' \text{ in any cell} \\ m' \neq m}} g_{\{m',b\}}P_{\{m',b\}}}. \quad (7.1)$$

Nous faisons alors l'approximation suivante, qui est très souvent utilisée pour le cas du lien montant (voir par exemple [67]) :

**Hypothèse 7.1** *Pour toute station de base  $b$ , l'interférence causée par les mobiles des autres cellules est proportionnelle à l'interférence due aux mobiles de la cellule. De façon plus formelle,  $\exists \lambda \in \mathbb{R}^+, \forall b, I_{m',b(m') \neq b} = \lambda I_{m',b(m')=b}$ , c'est-à-dire que :*

$$\sum_{m',b(m') \neq b} g_{\{m',b\}}P_{\{m',b\}} = \lambda \cdot \sum_{m',b(m')=b} g_{\{m',b\}}P_{\{m',b\}}. \quad (7.2)$$

Sous l'hypothèse 7.1, le lien montant peut donc être modélisé par le système (6.4) (ou de façon équivalente par (6.5)) avec :

$$C = \lambda + 1, \quad N(\{m, b\}) = \nu_b, \quad p'_{\{m,b\}} = g_{\{m,b\}}P_m.$$

Nous avons présenté dans cette section notre modélisation du lien montant. Elle a en fait déjà été utilisée dans la littérature (voir par exemple [67]). Nous avons utilisé une hypothèse classique sur les signaux interférents nous permettant de ramener le système à l'étude d'une cellule isolée.

<sup>1</sup>Une introduction aux problèmes de bilan de liaison et atténuation de signal se trouve dans le chapitre d'introduction aux réseaux satellitaires, 8.1.5.

Le système de contraintes peut alors s'écrire conformément au modèle du chapitre précédent (système 6.4). On peut remarquer que c'est en fait un cas relativement simple du fait que chaque émetteur (donc chaque mobile) n'utilise qu'un canal pour émettre. De plus, dans ce cas, l'expression du bruit est réduite au simple bruit thermique.

### 7.1.2 Utilisation des résultats précédents

Nous appliquons désormais les résultats obtenus dans l'analyse théorique du chapitre précédent. En particulier, nous obtenons (voir également [67]) :

$$\forall m, n = \{m, b(m)\}, P_n^{\prime min}(\rho) = \frac{1}{g_n} \left( \frac{\nu_b \delta_n \rho(n)}{1 - (1 + \lambda) \sum_{\substack{m' \\ b(m')=b(m)}} \delta_{\{m', b(m)\}} \rho(\{m', b(m)\})} \right).$$

Soit  $c$  une cellule,  $b$  la station de base associée et  $N_c$  le nombre de mobiles qui s'y trouvent, sumérotés de 1 à  $N_c$ . Le problème peut alors s'écrire :

Trouver  $\vec{\rho}$  sur chaque cellule qui maximise

$$Z(\vec{\rho}) := \sum_{m'=1}^{N_c} \frac{1}{1 - \alpha} \left( \frac{\rho(\{m', b\})}{1 - \delta_{\{m', b\}} \rho(\{m', b\})} \right)^{1-\alpha} \quad \text{tel que}$$

$$\left\{ \begin{array}{l} \forall m, 1 \leq m \leq N_c, \frac{MR_{\{m,b\}}}{1 + \delta_{\{m,b\}} MR_{\{m,b\}}} \leq \rho(\{m, b\}) \leq \frac{PR_{\{m,b\}}}{1 + \delta_{\{m,b\}} PR_{\{m,b\}}}, \\ (1 + \lambda) \sum_{m'=1}^{N_c} \delta_{\{m', b\}} \rho(\{m', b\}) < 1, \\ \forall m, 1 \leq m \leq N_c, 0 \leq \nu_b \delta_{\{m,b\}} \rho(\{m, b\}) \leq g_{\{b,m\}} \overline{p_{\{m,b\}}} \left( 1 - (1 + \lambda) \sum_{m'=1}^{N_c} \delta_{\{m', b\}} \rho(\{m', b\}) \right). \end{array} \right.$$

On remarque que dans ce cas :

**Remarque 7.1** *L'ensemble des contraintes est un polytope (convexe).*

De plus, si les quantités  $\mathbf{PR}$  et  $\delta$  vérifient une condition simple, alors la première condition n'est plus une contrainte du système. En effet :

**Remarque 7.2** *Une condition suffisante pour tous les vecteurs de taux pour satisfaire (C1) est :*

$$(1 + \lambda) \sum_{m'=1}^{N_c} \frac{\delta_{m',b} PR_{m'}}{1 + \delta_{m',b} PR_{m'}} < 1.$$

Nous concluons que pour  $\alpha \geq 2w$  (et en particulier pour  $\alpha \geq 2$ , voir le lemme 6.2 et la remarque 6.1), le problème multi-cellulaire du contrôle joint de puissance et de taux de transmission dans le lien montant peut se réduire à un problème standard de *minimisation avec des contraintes linéaires et une fonction objectif concave*. Il peut donc être résolu soit par des algorithmes décentralisés tels que les Lagrangiens, soit par des méthodes centralisées

basées par exemples sur la SDP (Semi-Definite Programming). Nous avons implémenté une solution par Programmation Semi-Définie positive, que je ne détaille pas ici.

On peut finalement noter que pour le cas d'une seule cellule, la solution précédente est une solution exacte (car nous n'utilisons pas l'hypothèse (7.1)).

## 7.2 Solution du lien descendant

Nous nous intéressons désormais au lien descendant, c'est-à-dire les signaux issus de la station de base à destination des différents mobiles. Notre expression est plus précise que dans [49] car notre modèle d'interférence est plus complet. De plus, les auteurs utilisent plusieurs hypothèses que nous ne considérons pas ici.

Comme dans la section précédente, je suppose que chaque mobile reçoit des signaux d'une seule station de base. On note comme précédemment  $\mathbf{b}(\mathbf{m})$  la station de base associée au terminal  $\mathbf{m}$ .

La première sous-section est consacrée à l'étude du système et aux hypothèses faites et montre que la encore le système du chapitre précédent peut être utilisé pour modéliser les contraintes associées aux débits et puissances des terminaux dans les réseaux de type CDMA. La seconde sous-section applique les résultats de l'étude théorique au système.

On peut noter que l'étude du lien descendant est moins aisée que celle du lien montant car chaque station utilise différents canaux.

### 7.2.1 Notations et hypothèses

Nous présentons dans cette sous-section notre modélisation des contraintes de débits et de puissance dans le lien descendant. Cette étude est découpée en plusieurs paragraphes. Nous verrons tout d'abord l'expression de la puissance totale émise par chaque station de base, la valeur des gains de transmission et une première expression du rapport signal sur bruit. Puis, nous expliciterons la valeur des puissance des différents signaux interférents. Comme déjà mentionnée, ce calcul est plus complexe que dans le cas du lien montant à cause de la multiplicité des canaux de chaque station de base et de leurs différents statuts. Enfin, nous émettrons une hypothèse sur la puissance des interférences inter et intracellulaire, comparable à celle faite dans le cas du lien montant.

#### 7.2.1.1 Puissance d'émission

Chaque station de base utilise un certain nombre de canaux. On note alors :

- $p_{\{\mathbf{b},\mathbf{m}\}}$  la puissance de transmission à la station de base  $\mathbf{b}$  pour le mobile  $\mathbf{m}$  (en général appelé Dedicated Physical Channel (DPCH)),
- $P_{SCH,\mathbf{b}}$  la puissance du canal de synchronisation de la station  $\mathbf{b}$ ,
- $P_{CCH,\mathbf{b}}$  la puissance du canal commun de la station  $\mathbf{b}$

La puissance totale émise par une station de base  $\mathbf{b}$  s'écrit alors tout naturellement :

$$p_{tot,\mathbf{b}} = \sum_{m',b(m')=\mathbf{b}} p_{\{\mathbf{b},m'\}} + P_{CCH,\mathbf{b}} + P_{SCH,\mathbf{b}}$$

### 7.2.1.2 Gain de transmission

On note  $h_{b,m}$  le gain du signal descendant. Bien que cela puisse paraître à première vue surprenant, ce gain est en général différent du gain du lien montant ( $g_{m,b}$ ). Cela peut s'expliquer de la façon suivante. Comme précédemment mentionné, les atténuation du signal sont dues principalement aux obstacles se situant entre la station de base et le terminal mobile. Or les fréquences utilisées dans les deux liens (montants et différents) sont différentes et l'on sait que les atténuations face aux obstacles varient avec la fréquence utilisée.

### 7.2.1.3 Première expression du rapport signal sur bruit

Je note comme précédemment  $\nu_m$  le bruit thermique au niveau du mobile receveur  $m$  et  $I_{\{m,b\}}$  l'interférence à laquelle est soumise la connexion  $\{m, b\}$  (cette interférence est mesurée au niveau du récepteur, c'est-à-dire du mobile  $m$ ).

La formule générale du rapport signal sur interférence entre le mobile  $m$  et la station  $b$  s'écrit :

$$SIR_{\{m,b\}} = \frac{p_{\{m,b\}} h_{\{m,b\}}}{\nu_m + I_{\{m,b\}}}.$$

Il nous reste alors à trouver une expression de l'interférence  $I_{\{m,b\}}$ .

### 7.2.1.4 Valeurs des interférences

#### Deux types d'interférences.

Les interférences perçues au niveau du mobile  $m$  sont de 2 types :

- Les interférences *intra-cellulaires* dues aux signaux envoyés par la station de base  $b(m)$  aux autres mobiles de la station ainsi sur les canaux communs et
- Les interférences *inter-cellulaires* dues aux réceptions des émissions des autres stations de base que  $b(m)$ .

Nous notons  $I_{inter}(\{m, b\})$  et  $I_{intra}(\{m, b\})$  les valeurs des puissances des interférences inter-cellulaire et intra-cellulaire respectivement reçues au niveau du mobile  $m$ . Le rapport signal sur bruit s'écrit alors :

$$SIR_{\{m,b\}} = \frac{p_{\{m,b\}} h_{\{m,b\}}}{\nu_m + I_{inter}(\{m, b\}) + I_{intra}(\{m, b\})}.$$

#### L'interférence inter-cellulaire.

Elle s'écrit simplement :

$$I_{inter} = \sum_{b' \neq b} p_{tot,b'} h_{\{b',m\}}.$$



### L'interférence intra-cellulaire.

A priori, elle s'écrirait sous la forme :

$$h_{\{b',m\}} \left( \sum_{m' \neq m, b(m')=m} p_{\{b,m'\}} + P_{CCH,b} + P_{SCH,b} \right).$$

En fait, dans les réseaux CDMA, afin de réduire la quantité d'interférence entre les signaux, on utilise des canaux orthogonaux (voir l'explication de ce terme dans la section 8.1.6). Dans la pratique cependant, les signaux ne sont pas parfaitement orthogonaux et l'on introduit alors  $\beta$  le facteur de synchronisation. On a bien évidemment  $0 \leq \beta \leq 1$ . Plus  $\beta$  est petit et plus les signaux sont orthogonaux, c'est-à-dire bien séparés et donc moins l'interférence est grande. En contrepartie, cela nécessite une augmentation de la taille des codes utilisés et donc un plus faible débit utile d'information.

L'interférence dues aux signaux envoyé par  $\mathbf{b}(m)$  aux autres terminaux de la cellule s'écrit :

$$\beta h_{\{b,m\}} \sum_{m', b(m')=b} p_{\{b,m'\}}.$$

De même, les interférences dues au canal commun s'écrivent :

$$\beta h_{\{b,m\}} P_{CCH,b}.$$

Enfin, l'autre canal commun (le canal de synchronisation) n'est pas orthogonal aux autres signaux. L'interférence qu'il génère au niveau du mobile  $m$  s'écrit alors :

$$h_{\{b,m\}} P_{SCH,b}.$$

Finalement, l'interférence intracellulaire totale s'écrit :

$$I_{intra} = \beta(p_{tot,b} - p_{b,m})h_{\{b,m\}} + (1 - \beta)p_{SCH,b}h_{b,m},$$

Et donc l'expression du  $SIR$  devient :

$$SIR_{\{m,b\}} = \frac{p_{\{b,m\}}}{\left( P_{SCH,b} + \beta P_{CCH,b} + \beta \sum_{m' \neq m} p_{\{b,m'\}} + \frac{1}{h_{\{b,m\}}} \left[ \nu_m + \sum_{b'=1, b' \neq b} p_{tot,b'} h_{\{b',m\}} \right] \right)}.$$

#### 7.2.1.5 Hypothèses

Comme pour le lien montant, posons  $F_{b,m}$  le rapport entre les puissance des signaux correspondant aux interférences inter-cellulaire et intra-cellulaire :

$$F_{b,m} = \frac{I_{intra}}{I_{inter}}$$

alors le SIR s'écrit :

$$SIR_{\{m,b\}} = \frac{P_{\{b,m\}}}{(1 + F_{\{b,m\}}) \left( \beta \sum_{m' \neq m} P_{\{b,m'\}} + P_{SCH,b} + \beta P_{CCH,b} \right) + \frac{\nu_m}{h_{\{b,m\}}}}. \quad (7.3)$$

Alors, comme dans [49], nous faisons les hypothèses suivantes :

**Hypothèse 7.2** *On suppose que :*

- $F_{\{b,m\}}$  est approché par sa valeur moyenne  $F$  sur l'ensemble des cellules du système
- $P_{CCH,b}$  et  $P_{SCH,b}$  sont des paramètres connus et ne sont pas soumis au contrôle de puissance.

On peut remarquer cependant que dans [49] d'autres hypothèses sont effectuées que nous ne considérons pas dans notre modèle.

### 7.2.2 Lien avec la formulation générale

Grâce aux hypothèses faites dans la sous-section précédente, on voit donc que l'on peut, comme dans le cas du lien montant, représenter le problème d'allocation jointe des débits et des puissances par le modèle présenté dans le chapitre précédent.

Ainsi, dans le lien descendant, les taux de transmissions dans le lien descendant par la résolution du problème (6.4) (ou de façon équivalente le problème (6.5)) avec :

$$N_D(\{b, m\}) = (1 + F)(P_{SCH,b} + \beta P_{CCH,b}) + \frac{\nu_m}{h_{\{b,m\}}}$$

$$C_D = (1 + F)\beta \quad \text{et} \quad p'_{\{m,b\}} = p_{\{b,m\}}.$$

Comme dans la section suivante, on peut résoudre le problème d'optimisation au niveau de chaque cellule. Soit  $\mathbf{b}$  une station de base. En appliquant les résultats de la section 6.4, nous obtenons le problème d'optimisation suivant : trouver  $\vec{\rho}$  maximisant  $\mathbf{Z}(\vec{\rho})$  tel que :

$$\left\{ \begin{array}{l} \forall m, b(m) = b, \frac{MR_{\{m,b\}}}{1 + \delta_{\{m,b\}} MR_{\{m,b\}}} \leq \rho(\{m, b\}) \leq \frac{PR_{\{m,b\}}}{1 + \delta_{\{m,b\}} PR_{\{m,b\}}}, \\ C_D \sum_{m, b(m)=b} \delta_{\{m,b\}} \rho_{\{m,b\}} < 1, \\ 0 \leq p_{SCH,b} + p_{CCH,b} + \sum_{m, b(m)=b} \delta_{\{m,b\}} \rho(\{m, b\}) \times \\ \frac{N_D(\{m, b\}) + C_D \sum_{m', b(m')=b} \delta_{\{m',b\}} \rho(\{m', b\}) [N_D(\{m', b\}) - N_D(\{m, b\})]}{1 - C_D \sum_{m', b(m')=b} \delta_{\{m',b\}} \rho(\{m', b\})} \leq \overline{p_{tot,b}}. \end{array} \right. \quad (7.4)$$

On voit que le système de contraintes ne paraît pas linéaire. Cependant, en y regardant de plus près, on s'aperçoit que pour n'importe quelle valeur de  $\delta$ ,  $\rho$ ,  $\mathbf{N}$  et  $\mathbf{C}_D$  nous avons de façon évidente :

$$\sum_{m,b(m)=b} \delta_{\{m,b\}} \rho(\{m,b\}) \left( N_D(\{m,b\}) + C_D \sum_{m',b(m')=b} \delta_{\{m',b\}} \rho(\{m',b\}) [N_D(\{m',b\}) - N_D(\{m,b\})] \right) = \sum_{m,b(m)=b} \delta_{\{m,b\}} \rho(\{m,b\}) N_D(\{m,b\}).$$

De plus, comme  $1 - C_D \sum_{m'} \delta_{m',b} \rho(m',b) > 0$ , alors pour toute cellule  $\mathbf{c}$  couverte par une station de base  $\mathbf{b}$  et dans laquelle évoluent les mobiles  $m, 1 \leq m \leq N_b$ , le problème d'allocation équitable et optimale des débits peut s'écrire :

Trouver  $\vec{\rho}$  qui maximise  $Z(\vec{\rho}) := \sum_{m=1}^{N_b} \frac{1}{1-\alpha} \left( \frac{\rho(\{m,b\})}{1-\delta_{\{m,b\}} \rho(\{m,b\})} \right)^{1-\alpha}$  tel que

$$\left\{ \begin{array}{l} \frac{MR_{\{m,b\}}}{1+\delta_{\{m,b\}} MR_{\{m,b\}}} \leq \rho(\{m,b\}) \leq \frac{PR_{\{m,b\}}}{1+\delta_{\{m,b\}} PR_{\{m,b\}}}, \\ (1+F)\beta \sum_{m=1}^{N_b} \delta_{\{m,b\}} \rho(\{m,b\}) < 1, \\ \sum_{m=1}^{N_b} \delta_{\{m,b\}} \rho(\{m,b\}) N_D(\{m,b\}) \leq (\overline{p_{tot,b}} - p_{SCH,b} - p_{CCH,b}) \left( 1 - C_D \sum_{m'=1}^{N_b} \delta_{\{m',b\}} \rho(\{m',b\}) \right). \end{array} \right.$$

Nous avons alors de nouveau un problème de maximisation avec des contraintes linéaires. Et pour  $\alpha \geq 2\mathbf{w}$  la fonction objective est concave. Ce système peut donc être aisément résolu. Dans ce cas encore, nous avons implémenté une méthode de résolution basée sur la SDP.

Nous avons donc vu dans cette section que le modèle du chapitre précédent permettait de représenter les contraintes de débits et de puissance associé au cas du lien descendant d'un système de type CDMA. On peut remarquer que notre modèle est plus complet que ceux que nous avons pu trouver dans la littérature, notamment grâce à la prise en compte des différents canaux existants et du faible nombre des hypothèses considérées. On peut remarquer également que le système de contrainte reste malgré tout linéaire.

## 7.3 Macro-diversité dans le lien descendant

Nous considérons dans cette section une extension du système de la section précédente permettant de prendre en compte la possibilité de hand-over dans les systèmes cellulaires de troisième génération.

Les systèmes de seconde génération utilisaient le *hard-handover*. Cela signifie que lorsque la puissance de signal entre un mobile et la station de base devenait inférieure à

un certain seuil, la communication basculait alors vers une seconde station de base, dont le signal était plus clair (c'est-à-dire de plus grand  $SIR$ ) si une telle station était disponible.

Ce genre de technique permettait alors à un mobile de se déplacer entre plusieurs cellules. Elle n'est pourtant pas très efficace lorsqu'un mobile réside en limite de zone de couverture. Il n'existe alors aucune station de base dont le signal reçu aura un  $SIR$  satisfaisant. Le risque est alors que la qualité de la communication soit dégradée.

Dans les systèmes tels que l'UMTS, on utilise désormais le *soft-handover*. Dans ce cas, un terminal peut avoir un lien actif avec deux stations de bases simultanément, ou deux secteurs de stations. Chaque mobile en limite de couverture peut alors écouter les signaux des deux stations de base et les reconstituer ensuite.

On considère ici l'utilisation du soft-handover uniquement dans le cas descendant. C'est-à-dire que l'on donne la possibilité au terminal de superposer les signaux qu'il reçoit de deux stations afin d'obtenir un signal de meilleure qualité. On ne considère pas en revanche de macrodiversité dans le lien montant car cela requièrerait une collaboration entre les stations ce qui est en pratique difficile.

### 7.3.1 SIR d'un terminal en situation de soft-handover

Considérons un mobile  $\mathbf{k}$  ayant un lien actif avec deux stations de base ou secteurs  $\mathbf{l}$  et  $\mathbf{s}$ . Les rapports signal sur bruit s'ajoutent ([49]) et l'on note  $SIR_{\{k\}} \geq \delta_{\{k\}}r(\{k\})$  leur somme :

$$\delta_{\{k\}}r(\{k\}) \leq SIR_{\{k\}} = SIR_{\{k,l\}} + SIR_{\{k,s\}}.$$

Supposons que le lien avec la station  $\mathbf{s}$  soit de meilleur qualité, c'est-à-dire que  $SIR_{\{k,l\}} \leq SIR_{\{k,s\}}$ . Je pose alors :

$$\Delta_k := \frac{SIR_{worst\ link}}{SIR_{best\ link}} = \frac{SIR_{\{k,l\}}}{SIR_{\{k,s\}}} \leq 1.$$

Afin de résoudre le problème joint de contrôle de puissance et de taux de transmission, nous faisons l'hypothèse simplificatrice que  $\Delta_{\mathbf{k}}$  ne dépend pas de  $\mathbf{k}$  (par exemple, nous proposons de prendre comme valeur de  $\Delta$  la valeur moyenne des  $\Delta_{\mathbf{k}}$  sur l'ensemble des mobiles qui sont en soft handover dans le système).

### 7.3.2 SIR des mobiles dépendants d'une station de base donnée

Nous séparons l'ensemble des mobiles qui sont en communication avec la station de base  $\mathbf{b}$  en trois groupes. Soient :

- $\mathbf{I}$  l'ensemble des mobiles de la cellule qui ne font pas de handover. Pour de tels mobiles nous avons :

$$SIR_{\{i\}} = SIR_{\{i,b\}} \text{ avec } \mathbf{SIR}_{\{i,b\}} \text{ donné par la formule 7.3.}$$

- $\mathbf{J}$  les mobiles en situation de soft handover dont le meilleur lien est celui avec la station de base  $\mathbf{b}$ . Alors,

$$SIR_{\{j\}} = (1 + \Delta_j)SIR_{best\ link} = (1 + \Delta_j)SIR_{\{j,b\}}. \text{ De nouveau, la valeur de } \mathbf{SIR}_{\{j,b\}} \text{ est donné par la formule (7.3).}$$

- $\mathbf{K}$  l'ensemble des mobiles connaissant également le soft handover, mais dont le plus mauvais lien est celui avec la station  $\mathbf{b}$ . Pour ceux la :

$$SIR_{\{k\}} = \frac{\Delta_k}{1+\Delta_k} SIR_{\text{worst link}} = \frac{\Delta_k}{1+\Delta_k} SIR_{\{k,b\}}.$$

### 7.3.3 Application des résultats du chapitre précédent

L'équation (6.1) s'écrit dans ce cas :

$$p_{tot,b} = \sum_{i \in I} p_{\{i,b\}} + \sum_{j \in J} p_{\{j,b\}} + \sum_{k \in K} p_{\{k,b\}} + P_{CCH,b} + P_{SCH,b}.$$

Posons :

$$\begin{cases} N_I(b, m) = N_D(b, m), & C_I = C_D \\ N_J(b, m) = \frac{1}{\Delta+1} N_D(b, m), & C_J = \frac{1}{\Delta+1} C_D \\ N_K(b, m) = \frac{\Delta}{\Delta+1} N_D(b, m), & C_K = \frac{\Delta}{\Delta+1} C_D \end{cases}$$

avec  $\mathbf{N}_D$  tel que défini dans la Section 7.2.

Soit enfin  $\mathbf{i}(\mathbf{m})$  l'indice de l'ensemble auquel appartient le mobile  $\mathbf{m}$  ( $i(\mathbf{m}) \in \{I, J, K\}$ ). Nous obtenons, pour chaque station de base  $\mathbf{b}$ , le problème d'optimisation suivant (je note encore  $\mathbf{N}_b$  le nombre de mobiles présents dans la cellule couverte par  $\mathbf{b}$ ) :

Trouver  $\vec{\rho}$  qui maximise  $Z(\vec{\rho}) := \sum_{m=1}^{N_b} \frac{1}{1-\alpha} \left( \frac{\rho(\{m\})}{1-\delta_{\{m\}}\rho(\{m\})} \right)^{1-\alpha}$  tel que

$$\begin{cases} \frac{MR_{\{m,b\}}}{1+\delta_{\{m\}}MR_{\{m,b\}}} \leq \rho(\{m\}) \leq \frac{PR_{\{m,b\}}}{1+\delta_{\{m\}}PR_{\{m,b\}}}, \\ \sum_{m=1}^{N_b} C_{i(\mathbf{m})} \delta_{\{m\}} \rho(\{m\}) < 1, \\ \sum_{m=1}^{N_b} \delta_{\{m\}} \rho(\{m\}) N_{i(\mathbf{m})}(\{m, b\}) \leq (\bar{p}_{tot,b} - P_{SCH,b} - P_{CCH,b}) \left( 1 - C_{i(\mathbf{m})} \sum_{m'=1}^{N_c} \delta_{\{m'\}} \rho(\{m'\}) \right). \end{cases} \quad (7.5)$$

### 7.3.4 Conclusion

Nous remarquons que nous obtenons donc encore un problème d'optimisation avec une fonction d'objectif concave et des contraintes linéaires, ce qui est un problème standard à résoudre et pour lequel il existe des algorithmes de résolutions efficaces. Nous l'avons également implémenté en programmation SDP.

Cette expression, quoique d'apparence simpliste est en fait plus générale que ce que nous avons pu trouver dans la littérature. Par exemple, si nous regardons la méthode utilisée dans [49], on remarque que les auteurs ne distinguent pas les ensembles  $J$  et  $K$ . C'est pourquoi leurs équations diffèrent. Aussi, ils supposent qu'il y a le même nombre de mobiles dans chaque cellule et que

$\forall i, j, (\{i, b\}, \{i, s\}, \{j, b\}, \{j, s\}) \in \mathcal{N}^4, \delta_{\{i,b\}} r(\{i, b\}) = \delta_{\{j,s\}} r(\{j, s\})$ . Ce qui est une grossière approximation.

Enfin, on peut noter que dans cette partie, l'hypothèse a été faite que  $\Delta_{\mathbf{k}}$  était indépendant de  $\mathbf{k}$ . En fait, on peut, si l'on désire avoir une image plus fine de la situation du réseau, discrétiser les valeurs possibles de  $\Delta_{\mathbf{k}}$ . Cela peut se faire simplement en sous-divisant les ensembles  $\mathbf{J}$  et  $\mathbf{K}$  en  $\mathbf{J}_p$  et  $\mathbf{K}_p$ ,  $1 \leq p \leq n$  avec  $\forall p$ ,

$$\begin{cases} N_{J_p}(b, m) = \frac{1}{\Delta_{p+1}} N_D(b, m), & C_{J_p} = \frac{1}{\Delta_{p+1}} C_D \\ N_{K_p}(b, m) = \frac{\Delta_p}{\Delta_{p+1}} N_D(b, m), & C_{K_p} = \frac{\Delta_p}{\Delta_{p+1}} C_D \end{cases}$$

On peut par exemple prendre en compte 3 ou 4 valeurs types de  $\Delta_k$  pour représenter les divers scénarios. On se rend compte que augmenter le nombre possibles de valeurs de  $\Delta$  n'augmente pas la complexité du problème d'optimisation mais celui de l'implémentation pratique. En effet, cela nécessiterait une mesure précise de tous les rapports signal sur bruit des terminaux en situation de soft-handover, ce qui est peut réalisable en pratique.

## 7.4 Approximations possibles

Comme nous l'avons vu, la condition  $\alpha \geq 2w$  ne couvre pas le cas intéressant de  $\alpha = 0$  qui correspond au problème de la maximisation du débit global. Nous proposons alors dans cette section deux approximations.

### 7.4.1 Premier schéma : approximation de la fonction objectif

Nous pouvons approcher la fonction  $Z(\vec{\rho})$  par  $Z^{app}(\rho) = \sum_m \frac{(\rho(\{m, b\}))^{1-\alpha}}{1-\alpha}$ , c'est-à-dire négliger le terme  $\delta_{\{m, b\}} \rho(\{m, b\})$  du dénominateur, vu qu'il est assez souvent beaucoup plus petit que 1 (comme mentionné plus tôt). Avec cette nouvelle fonction objectif remplaçant la précédente, nous obtenons un problème d'optimisation convexe pour tout  $\alpha > 0$  (excepté  $\alpha = 1$ ). Nous pouvons donc résoudre le problème pour n'importe quel critère d'équité avec un paramètre  $\alpha$  arbitrairement proche de 0.

L'ensemble des contraintes (et donc l'ensemble des solutions faisables) pour cette méthode d'approximation est le même que dans le problème initial. La solution  $(\vec{\rho}^{app}, \vec{\rho}^{app})$  obtenue par cette méthode d'approximation est donc une paire valide pour le problème initial.

De plus, on peut remarquer que la valeur obtenue par cette approximation  $Z(\vec{\rho}^{app})$  est sous optimale, c'est-à-dire que  $\exists(\vec{p}, \vec{\rho}) \in \Pi^c, Z(\vec{\rho}) \geq Z(\vec{\rho}^{app})$ . La solution trouvée par ce schéma d'approximation est donc une borne inférieure de la solution du problème d'optimisation original. Cette borne est d'autant meilleure que  $\delta_{\{m, b\}} \rho(\{m, b\})$  est négligeable devant 1<sup>2</sup>.

<sup>2</sup>Ceci est en pratique vérifié lorsque le système contient un nombre raisonnablement grand de mobiles. En effet, rappelons que  $\sum_n \delta_n \rho_n \leq 1$ .

### 7.4.2 Second schéma d'approximation : approximation de l'ensemble des contraintes

Une autre approximation possible est obtenue en considérant la formulation originelle (6.4) en termes de vecteur de taux  $\vec{r}$  et non  $\vec{\rho}$ , dans laquelle la fonction objectif est concave mais l'ensemble des contraintes n'est pas convexe (pour plus de détails sur cet ensemble, se référer à [69] qui considère le case d'une cellule unique). Notre approximation consiste alors à remplacer la dernière contrainte dans (6.4) par :

$$\delta_{\{m,b\}r}(\{m,b\}) \leq \frac{g_{\{b,m\}}P_{\{m,b\}}}{N_{\{m,b\}} + \sum_{m' \in \mathcal{M}_m} g_{\{m',b\}}P_{\{m',b\}}}. \quad (7.6)$$

Nous avons étudié ce schéma d'approximation dans [113] dans le cas du lien montant. Dans le cas général, on s'aperçoit que l'approximation revient à remplacer  $\sum_{m' \in \mathcal{M}_m, m' \neq m} g_{\{m',b\}}P_{\{m',b\}}$  par  $\sum_{m' \in \mathcal{M}_m} g_{\{m',b\}}P_{\{m',b\}}$ . Ainsi l'on remplace un ensemble de contraintes par un strict sous ensemble de celui-ci. En d'autres termes, si  $(\vec{p}^{app2}, \vec{r}^{app2})$  satisfait les contraintes de ce schéma d'approximation, alors il satisfait également les contraintes du système initial. De plus on peut également dire que l'approximation est raisonnable quand le système contient un nombre raisonnable de mobiles. En effet, l'approximation consiste à négliger la puissance reçue d'un signal devant celle de tous les autres. Dans un système où les puissances reçues sont relativement homogènes<sup>3</sup> et de nombre de terminaux par cellule raisonnable, cette approximation paraît tout à fait justifiée, comme le confirme nos expériences numériques.

Ainsi, comme pour le schéma d'approximation précédent, le résultat trouvé par cette approximation est une solution acceptable mais sous optimale pour le problème initial d'allocation équitable et optimale des débits et des puissances dans les systèmes de type CDMA.

### 7.4.3 Conclusion

Nous avons proposé deux méthodes d'approximations du problème d'allocation optimal et équitable des débits et des puissances dans les réseaux CDMA. L'intérêt de ces deux méthodes est de produire des solutions qui sont acceptables, c'est-à-dire des couples  $(\vec{p}, \vec{r})$  que l'on peut en pratique affecter aux mobiles du système car ils respectent les contraintes de celui-ci.

De plus, les problèmes associés à ces approximations sont implémentables en programmation SDP et valides pour des systèmes relativement larges.

---

<sup>3</sup>C'est-à-dire qu'il n'existe pas un mobile dont la puissance  $p'$  et de gain  $g'$  est telle que pour toute autre mobile de puissance  $p''$  et de gain  $g''$  alors  $\frac{g'p'}{g''p''} \ll 1$ .

## 7.5 Quelques résultats numériques et conclusion

### Premiers résultats numériques.

Nous avons implémenté les solutions correspondants au lien montant pour le problème exact et approché. Je montre dans les figures 7.1 et 7.2 les premiers résultats obtenus dans le cas d'une cellule (pour  $\alpha = 0$  et 1 respectivement). Le disque noir du centre de la figure représente la station de base concernée et les mobiles sont représentés par des cercles centrés à leur réelle position géographique et de rayon et proportionnel à la puissance (ou la bande passante) allouée.

Pour les  $N$  mobiles, nous avons fixons  $\delta = 1$ ,  $\nu = 1$ ,  $\bar{p} = 1/N$ ,  $MR_i = 1/4N$ , et  $PR_i = 1$ . La position des mobiles est prise au hasard dans le carré  $[-1; +1] \times [-1; +1]$  et le gain est égal à  $1/d^2$ , où  $d$  est la distance au centre du carré. Nous prenons  $N = 50$ .

Nous voyons clairement sur les figures que les mobiles situés près de la station de base tendent à recevoir plus de bande passante. Cependant de subtiles différences sur les valeurs des rayons apparaissent entre les deux critères d'équité. En particulier, dans l'équité proportionnelle aucun mobile ne se voit affecter sa puissance maximale et quelque uns ne reçoivent quasiment que leur demande minimale. Dans le cas  $\alpha = 0$  en revanche, on voit clairement que les utilisateurs situés près de la station de base peuvent émettre avec la puissance maximale et donc avec un débit maximal. Parallèlement, les utilisateurs situés loin du centre du carré se voient systématiquement allouer leur demande minimale car ce sont eux qui souffrent des plus grandes atténuations de signal (et ont donc les plus faibles valeurs de  $g$ ).

### Conclusion

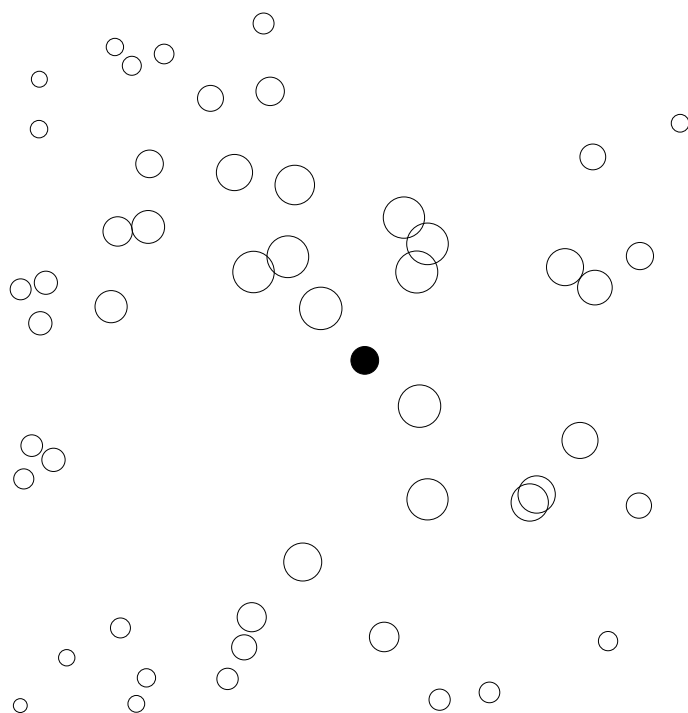
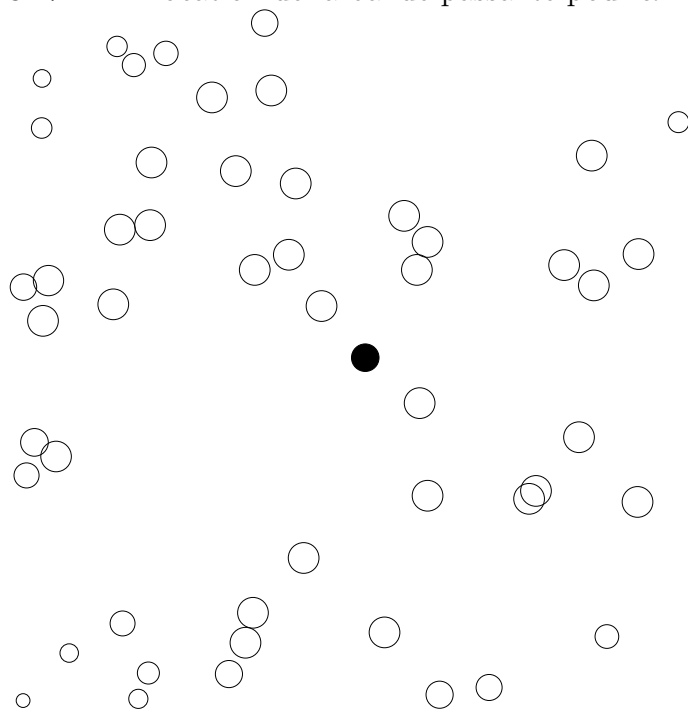
Nous avons montré dans ce chapitre que les contraintes portant sur les débits et les puissances des signaux utilisés dans les liens montants et descendants des réseaux de type CDMA pouvaient effectivement être modélisés par le système de contraintes du chapitre précédent. Nous avons également proposé une mise en équation rigoureuse du système de contraintes imposées par l'utilisation du soft-handover dans de tels réseaux.

Nous avons de plus montré que les contraintes se résument à un système d'équation linéaire, et qu'il est ainsi possible de les résoudre, par exemple par l'utilisation de la SDP.

Enfin, nous avons vu précédemment que pour les faibles valeurs de  $\alpha$ , la fonction d'optimisation n'était pas convexe. Nous avons donc proposé deux schémas d'approximations possibles pour traiter les valeurs de  $\alpha$  arbitrairement proches de 0.

Enfin, nous avons montré quelques résultats numériques. Ils tendent à montrer que pour un nombre raisonnablement grand de mobiles par cellule les allocations obtenues avec les approximations faites pour les faibles valeurs de  $\alpha$  sont proches des allocations optimales.



FIG. 7.1 – Allocation de la bande passante pour  $\alpha = 0$ .FIG. 7.2 – Allocation de la bande passante pour  $\alpha = 1$ .

# Conclusion

Nous nous sommes intéressés ici à l'allocation jointe des débits et des puissances d'émissions dans les réseaux sans fil afin d'être à la fois équitable et optimal. Nous avons présenté une modélisation qui est valide à la fois pour le lien montant et descendant. Nous avons mis en valeur les similitudes entre les deux cas en prouvant des théorèmes communs. Nous avons ensuite étudié la structure particulière de chaque lien afin d'obtenir des solutions spécifiques. Nous avons finalement présentée une solution simple du problème de macrodiversité dans le lien descendant.

Rappelons que l'idée de la macrodiversité dans le lien descendant est de permettre à un mobile en limite de couverture de superposer les signaux qu'il reçoit de deux stations afin d'améliorer la qualité du signal qu'il reçoit. Dans ce cas, la valeur du *SIR* de son signal est simplement égal à la somme des *SIR* des signaux des deux stations.

Notre système peut être résolu de façon efficace pour les valeurs de  $\alpha$  supérieures à un seuil, que nous avons explicité. Nous avons fourni, pour les autres valeurs du paramètre d'équité deux méthodes d'approximations que nous avons validées par des expériences numériques.

Le modèle que nous avons proposé est à notre connaissance le plus précis et aux contraintes les plus minimalistes que l'on peut raisonnablement considérer. De plus, de part le contrôle des puissances d'émissions des mobiles notre méthode augmente considérablement l'efficacité totale des systèmes par rapport à ceux déjà implémentés dans le cadre des réseaux UMTS. Le contrôle des puissances de façon jointe aux débits est nécessaire, mais la relation entre les deux grandeurs n'est pas triviale car l'influence de la puissance sur le débit est doublée :

- D'une part le débit que peut utiliser chaque émetteur est proportionnel à la puissance qu'il utilise
- Une augmentation de chaque puissance d'émission accroît la quantité d'interférence des autres connexions et donc limite leur débit de chacun.

Nous avons alors montré que pour chaque vecteur de débits que le système peut supporter, il existe un vecteur de puissance minimal correspondant. Nous avons alors donné une expression analytique de ce vecteur de puissance minimal et proposé un système d'optimisation permettant d'obtenir une allocation optimale et équitable des débits.

Le programme SDP utilisé ne présente pas en lui-même de réelles innovations, la méthode de construction des matrices étant sensiblement identiques à celle utilisée dans les réseaux terrestres. L'intérêt majeure de cette étude était donc d'introduire les problèmes d'équité dans un système où deux ressources sont couplées.

Nos travaux font suite à beaucoup de références ([61, 77, 85, 117]) qui considèrent l'allocation équitable du débit lui-même dans d'autres contextes de télécommunications. Une

question à laquelle nous ne répondons pas est celle de savoir comme atteindre ces débits en pratique si un mode paquet est utilisé ou, en d'autres termes comme ordonnancer les paquets de façon à obtenir ces débits qui étaient alloués équitablement. Cette seconde question à été très bien étudiée, voir par exemple [15, 82, 74].

On peut noter que nous avons dans cette partie appliqué les concepts d'équité aux débits des connexions. En fait, nous voyons que l'on peut utiliser notre concept d'équité généralisé et appliquer les concepts d'équité aux utilités des connexions comme nous l'avons fait précédemment. En effet, cela ne pose pas de problème d'implémentation.

Il serait intéressant toutefois d'étudier plus avant les performances de nos méthodes d'approximation pour les faibles valeurs du paramètre d'équité.

Enfin, des travaux d'implémentations sont en cours, dans le cadre de travaux menés en collaboration avec France Telecom R&D. L'idée est effectivement d'utiliser notre critère pour la gestion en temps réel le débits des connexions ainsi que les décisions d'admissions d'appels. Le principe est, comme nous l'avons mentionné plus haut, d'utiliser l'algorithme à chaque demande d'admission d'un nouveau mobile afin de déterminer un nouvel état optimal et équitable du réseau en modifiant si nécessaire, les débits des mobiles présents dans le réseau. Dans ces travaux, nous considérons un ensemble de débits discrétisés, ce qui soulève de nouvelles questions algorithmiques. Dans la version actuelle, nous calculons l'allocation optimale et équitable comme présentée dans cette thèse puis allouons le débit discrétisé dont la valeur est la plus proche immédiatement inférieure au débit. Cette méthode est donc une sous-optimale. On pourrait peut être l'améliorer en proposant par exemple d'utiliser cette technique d'allocation puis de chercher s'il est possible d'augmenter le débit de l'une des connexions sans sortir des contraintes du système. On voit alors que le choix de la connexion en question est crucial.

**Troisième partie**  
**Les réseaux satellites**



# Introduction

Cette dernière partie s'intéresse aux communications par satellite. Dans de tels réseaux les transmissions se font par canal radio, tout comme dans les réseaux mobiles. Cependant les systèmes satellitaires ont des caractéristiques physiques bien différentes, notamment à cause de l'éloignement géographique séparant les terminaux des stations émettrices. Les technologies utilisées dans les terminaux, ainsi que dans les satellites sont donc de nature très différentes, ce qui justifie une étude spécifique.

Le système étudié dans cette partie est celui qui nous a été proposé par Alcatel Space Industries. Il s'agit d'un système satellite géostationnaire multispot MF/TDMA dans lequel un manager veut allouer de la bande passante entre différents opérateurs.

Dans un système MF/TDMA, la ressource radio est découpée en intervalle de temps et de fréquence. Ces surfaces élémentaires sont appelées *slots*. Rappelons qu'au contraire, dans un système CDMA tel que l'UMTS, tous les terminaux utilisent la même fréquence simultanément, mais que les codes qu'ils utilisent diffèrent.

Les satellites géostationnaires évoluent à grande distance de la terre (approximativement 36000 km). Ils couvrent donc un grand nombre de stations terrestres et une grande superficie (typiquement, 3 satellites suffisent à couvrir toute la surface du globe). On ne considère pas alors dans de tels systèmes les problèmes de hand-over ni de macro-diversité. On ne s'intéresse plus non plus aux interférences entre les terminaux mais entre les différentes zones géographiques dans lesquelles ils évoluent.

De plus, de nouvelles contraintes apparaissent, en particulier des notions d'intégrité : les terminaux utilisent des porteuses de largeur de bande différentes mais de taille prédéfinies.

Cette partie se compose de trois chapitres. Le premier constitue une introduction générale sur les systèmes satellitaires, mettant en exergue leurs spécificités et les travaux déjà effectués dans le cadre de l'allocation de ressources. Notre approche, quant à elle se décompose en deux phases. Dans la première nous déterminons le nombre et le type de porteuses à affecter à chacun des opérateurs du système en prenant en compte la bande passante disponible totale et introduisons des notions d'équité. C'est la phase de *planification radio*. Elle fait l'objet du chapitre 7. Puis, dans une seconde phase, appelée *allocation des slots* nous convertissons cette quantité de ressource en nombre de time-slots que nous positionnons dans l'espace constitué bidimensionnel temps/fréquence en prenant en compte les conditions de propagations radio et d'interférences.



# Chapitre 8

## Introduction aux réseaux satellitaires.

Cette partie introductive se compose de trois sections. La première est une introduction générale sur les réseaux satellitaires, fortement inspirée du livre [7]. Le but est de comprendre quels sont les spécificités de tels systèmes, par rapport à ceux déjà présentés dans cette thèse.

Le second chapitre présente une vue d'ensemble des travaux précédents effectués dans le cadre de l'allocation de ressources dans les systèmes TDMA et MF/TDMA (Multiple-Frequency TDMA, cf sous-section 8.1.6).

Enfin, nous présentons dans un troisième chapitre le système que nous nous proposons d'étudier.

### 8.1 Qu'est-ce qu'un réseau satellitaire ?

Les satellites ont vu le jour dans les années 60. L'intérêt était d'assurer une fonction de relais de télécommunications. En effet, ils permettent d'établir des communications dans des zones inaccessibles pour les réseaux, notamment dans les zones désertiques et les océans.

Les autres avantages des satellites sont principalement de pouvoir couvrir de grandes zones géographiques, d'offrir des accès et des destinations multiples pour une même communication et de permettre un déploiement rapide de services.

Leurs concurrents terrestres conservent de nombreux atouts, notamment grâce à la qualité de leur média de transmission. Ils offrent (en particulier depuis l'introduction de la fibre optique) un débit important avec peu d'erreurs.

Les services offerts par les satellites sont nombreux. L'importance du débit qu'ils offrent peut permettre d'interconnecter des réseaux locaux. La télévision par satellite est aussi un service répandu, rendu possible notamment par la taille des couvertures géographiques. En outre, les progrès technologiques permettent aujourd'hui d'utiliser des terminaux de taille réduite et donc l'accès à des services mobiles, de téléphonie notamment. En effet, les téléphones portables satellitaires sont aujourd'hui de tailles comparables à leurs homologues GSM. De plus, des terminaux de la taille d'une valise permettent d'accéder des applications multimedia haut débit.

Nous présentons dans un premier temps les orbites possibles des satellites. Nous détaillons ensuite les bandes de fréquences utilisées, puis les spécificités des systèmes géostationnaires. Enfin, nous terminons cette présentation par quelques éléments de compréhension du bilan de liaison et enfin les différentes méthodes d'accès.



### 8.1.1 Orbites et empruntes.

Un satellite est un équipement placé en orbite autour de la terre qui dispose de peu de moyens de propulsion propres. Ceux dont il dispose servent uniquement à corriger les perturbations éventuelles de l'orbite.

Il est important de remarquer qu'il n'est en général pas concevable d'aller modifier le matériel embarqué dans un satellite, comme on pourrait le faire avec des équipements terrestres. C'est pourquoi ces matériels sont hautement reconfigurables.

#### Orbites idéales des satellites.

Une orbite est la trajectoire idéale que suit un satellite en l'absence de perturbations, c'est-à-dire quand il n'est soumis qu'à l'attraction de la terre. En fait, la trajectoire sans propulsion d'un appareil autour de la terre ne dépend pas de son poids.

On peut en fait montrer que les orbites des satellites sont des ellipses, dont la terre est l'un des foyers. On appelle inclinaison  $i$  est l'angle entre le plan de l'orbite du satellite et le plan équatorial. L'apogée est le point où le satellite est le plus éloigné de la terre, le périégée est celui où il en est le plus proche.

**Remarque 8.1** (*Petit rappel sur les ellipses*) On rappelle qu'une ellipse est l'ensemble des points  $(x, y)$  qui vérifient :

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

où  $a$  est appelé demi-axe principal et  $b$  le demi-petit axe. L'excentricité d'une ellipse est le rapport  $e = c/a = \sqrt{a^2 - b^2}/a$ . Lorsque  $e = 0$ , l'orbite est circulaire.

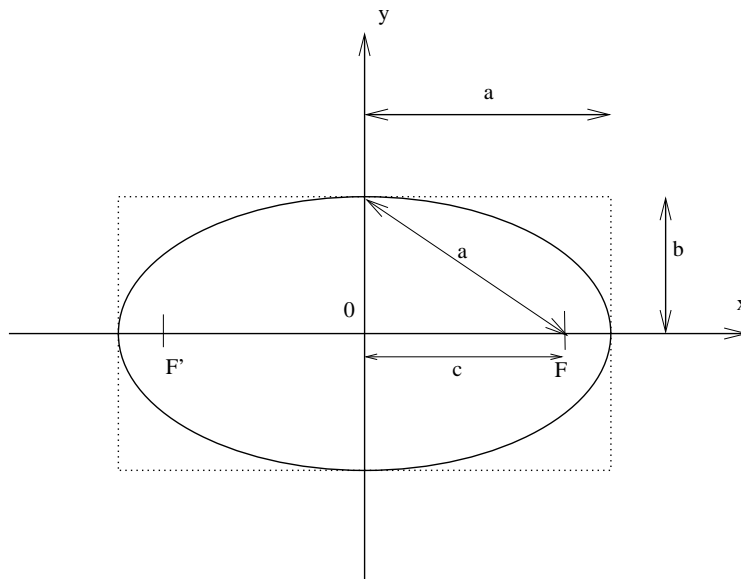


FIG. 8.1 – Allure générale d'une ellipse.

Ceci décrit des orbites parfaites. En fait, deux types de perturbations viennent modifier ces trajectoires : les *perturbations gravitationnelles* (dus au volume de la terre, à la présence d'autres astres à proximité du satellite, aux marées, etc) et les *perturbations non gravitationnelles* (dus par exemple au frottement dans l'atmosphère). Pour cette raison la position du satellite est régulièrement évaluée par des centres de contrôle et réajustée suivant les souhaits de l'exploitant. A l'heure actuelle le carburant le plus couramment utilisé par les satellites est l'ergol. Malheureusement le combustible est une ressource limitée du satellite : une fois épuisé, l'engin est considéré comme inutilisable parce que l'on ne peut plus contrôler sa position, et cela même si les équipements à bord fonctionnent correctement.

### 8.1.2 Altitudes possibles

La terre contient des champs magnétiques importants : certaines zones présentent des radiations suffisamment importantes pour endommager les composants électroniques qui les franchissent.

Il existe ainsi deux zones considérées comme dangereuses pour les équipements électroniques des satellites. Elles sont appelées "ceintures de Van Allen". En outre, un satellite situé à une altitude inférieure à 1000 km subit les effets du frottement par l'air. C'est pourquoi aucune orbite de satellite ne se situe à moins de 200 km d'altitude. On distingue quatre régions principales de mise en orbite des satellites :

- la zone *LEO* (Low Earth Orbit) entre la fin de l'atmosphère et la première ceinture de Van Allen, de 400 km à 1500 km d'altitude,
- la zone *MEO* (Medium Earth Orbit) entre les deux ceintures de Van Allen, de 5000 à 13000 km d'altitude,
- la zone *HEO* (High Earth Orbit) dont l'apogée est au-delà des ceintures de Van Allen, mais qui traverse les zones précédentes,
- la zone *GEO* (Geostationary Earth Orbit), qui peut se voir comme un cas particulier de HEO, pour les satellites à orbite circulaire à une altitude de 35786 km. Les satellites ils évoluent à la même vitesse que la terre, de sorte que le point de la terre au-dessus duquel il se trouve est fixe.

Les satellites LEO et GEO ont des orbites circulaires. Les satellites HEO ont des orbites elliptiques. Les satellites MEO peuvent avoir des orbites circulaires ou elliptiques.

Un inconvénient des satellites géo-stationnaires est leur délai de transmission. Ce délai correspond au temps de propagation des ondes électromagnétiques dans le vide, qui est, avec une très bonne approximation égal à la vitesse de la lumière dans le vide. Ainsi un aller-retour vers un satellite géostationnaire prend environ 240 msec. C'est en particulier un frein majeur pour les communications téléphoniques. Une des motivations principales des satellites à orbite basse est de réduire les délais de communications.

### Emprunte

Le point de la surface terrestre le plus proche d'un satellite est celui qui se trouve exactement à sa verticale. Plus on s'éloigne de ce point, plus la réception se dégrade et ceci indépendamment du relief. En effet, étant donné l'altitude minimale à laquelle les satellites évoluent, le relief de notre planète peut être considéré comme inexistant. La zone de cou-

verture d'un satellite est donc un disque sur la surface terrestre. Ce disque porte le nom d'empreinte (*footprint*).

Réciproquement, l'élévation est l'angle avec lequel un utilisateur à la surface de la terre "voit" le satellite. Plus cet angle est petit, c'est-à-dire plus la satellite parait près de l'horizon, et plus la reception est mauvaise.

### 8.1.3 Bandes de fréquence utilisées.

Un des points stratégiques dans les télécommunications par satellites est l'allocation des ressources du spectre radio. Le spectre est découpé en plusieurs zones appelées "bandes". Ce découpage est dû aux propriétés physiques d'absorption de l'atmosphère, et à l'histoire du développement des communications par voie radio. Les bandes fréquemment utilisées sont :

- AM autour de 1 mHz,
- HF autour de 10 mHz,
- VHF autour de 100 mHz,
- UHF de 300 mHz à 900 mHz.

Les bandes de fréquences pour les communications par satellite se répartissent selon le schéma suivant (en GHz).

1	2	4	8	12.25	18	26.5	46	56
L	S	C	X	Ku	K	Ka	V	

Le choix de la bande utilisée à un impact sur l'utilisateur. En effet :

- la taille de l'antenne diminue quand la fréquence augmente. Ainsi par exemple, les antennes utilisant la bande Ka sont deux fois plus petites que celle de la bande Ku.
- la disponibilité des bandes passantes croît avec la fréquence,
- la qualité de la propagation se dégrade avec les fortes fréquences. On remarque ainsi globalement que lorsque l'on parcourt l'ensemble des bandes de fréquence par ordre croissant (c'est-à-dire de la bande L à la bande Ka), les ondes sont de plus en plus sensibles aux conditions atmosphériques (en particulier à la pluie) mais connaissent moins d'atténuation lors de leur traversée dans la ionosphère.
- la mobilité prend place dans les faibles fréquences,
- le haut-débit se positionne dans des bandes plus élevées. Ainsi la bande Ku dispose de trois fois plus de capacité que la bande C.

Certaines utilisations sont caractéristiques des bandes. Ainsi, la bande L sert de support aux services mobiles, de télévision UHF et de téléphonie cellulaire. La bande C aux applications de télévision par satellite et la la bande X était traditionnellement utilisée par les applications militaires.

On voit ainsi que l'utilisation que l'on souhaite avoir du système influe non seulement sur le choix de l'orbite mais également sur la bande de fréquence utilisée.

### 8.1.4 Spécificité des systèmes GEO

Ils sont caractérisés par un grand délai de transmission. De plus, de tels gardent une position fixe au dessus de l'équateur. Leur altitude importante leur permet de couvrir une large partie du globe. Typiquement, trois satellites GEO peuvent assurer une couverture maximale.

Cependant ces systèmes posent d'importants problèmes de couverture. En effet, il leur est impossible, pour des raisons de simple visibilité, de couvrir les pôles, ainsi que toute position qui a une latitude de plus de 75 degrés en pratique. Cependant, même à des latitudes comprises entre 45 et 75 degrés, l'élévation du satellite est petit et pose des difficultés dès que le relief n'est pas plat. Ainsi, de graves problèmes de blocage peuvent apparaître à cause des grands immeubles dans les zones urbaines. Or, par exemple, les villes de Paris, Londres, Berlin et Moscou sont comprises dans cette zone.

Il serait possible de remédier à ce défaut en utilisant des orbites légèrement inclinées les unes aux autres. Les trajectoires des satellites se présentent alors sous la forme d'un "huit" décrit quotidiennement. Il s'agit de trajectoires géosynchrones non géostationnaires. Ce système a cependant l'inconvénient majeur d'utiliser trois fois plus de satellites.

Un autre inconvénient de tels systèmes est l'orbite géostationnaire a déjà été amplement utilisée pour de nombreux systèmes, et les autorisations de mise en orbite sont de plus en plus difficiles à obtenir. La puissance radio nécessaire pour franchir les quelques 35800 km qui les séparent de la terre est considérable, voire dissuasive pour les systèmes qui réclament du haut débit ou de la mobilité. Enfin, les technologies actuelles ne permettent pas une grande réutilisation des fréquences radio depuis une orbite aussi éloignée, les cellules de communication étant très larges.

### 8.1.5 Bilan de liaison

Chaque satellite contient un certain nombre d'*antennes* (typiquement 2 comme chez Globalstar ou 3 chez Iridium), chacune pouvant transmettre un certain nombre de *faisceaux* (16 chez Iridium). La couverture du satellite est divisée en *spots*, chacun étant couvert par un faisceau. Le nombre de faisceaux radio disponibles limite le nombre de communications admise de façon concurrente dans le réseau.

Avant d'être émis, un signal doit subir un traitement de modulation et d'amplification. Au niveau du récepteur, ce signal est réamplifié et démodulé. Le rapport  $P_r/P_e$  entre la puissance reçue et émise vaut

$$\frac{P_r}{P_e} = G_r \cdot G_e \left( \frac{c}{4\pi r F} \right)^2$$

avec  $G_e$  et  $G_r$  les gains des antennes d'émission et de réception,  $c$  la vitesse de lumière,  $F$  la fréquence en Hz,  $r$  la distance en mètres.

Dans le domaine des communications satellitaires, l'affaiblissement dû à la distance peut être très important. Il vaut typiquement 205 dB pour une transmission entre la terre et un satellite géostationnaire. Or, à titre d'exemple, une bonne antenne parabolique (3.20 m de diamètre) apporte un gain réel de 40.2 dB seulement.

Afin de garder un rapport signal/bruit acceptable, l'émetteur (respectivement le récepteur) doit donc fournir une puissance (respectivement avoir une sensibilité) suffisante. La sen-

sibilité représente la plus petite valeur qu'il faut donner au signal d'entrée du démodulateur (au niveau du récepteur) pour obtenir un rapport signal/bruit déterminé.

Le *bilan de liaison* permet d'établir le rapport signal/bruit d'un système donné. Il se calcule en plusieurs étapes.

La *PIRE* (Puissance Isotropique Rayonnée par l'Émetteur) est la puissance de sortie de l'émetteur dans une direction donnée. Elle est donnée par :

$$PIRE(dBW) = \text{Puissance de Transmission (dBW)} \\ - \text{Pertes du guide d'onde d'émission (dB)} + \text{Gain de l'antenne d'émission (dBi)}.$$

Les pertes du guide d'ondes concernent l'acheminement du signal vers l'antenne. Suivant la qualité du support (paire de cuivre, câble coaxial, fils en or) et son isolement et sa longueur, une perte plus ou moins grande peut intervenir. Il est toujours préférable de rapprocher l'émetteur de l'antenne pour limiter les dissipations.

On calcule ensuite la puissance reçue, dépendant de l'absorption atmosphérique, de la PIRE et du gain de l'antenne en réception.

Enfin, la puissance de bruit du récepteur dépend de la largeur de bande du signal et des conditions de réception. La contribution principale en est la composante thermique. D'ordinaire les satellites comprennent un système de régulation thermique qui vise à répartir la chaleur dissipée par les équipements sur l'ensemble de l'appareil. En pratique, la température de bruit d'un équipement est donnée et mesurée par le constructeur.

Finalement, le rapport signal sur bruit est alors donné par :

$$\text{Rapport Signal/Bruit (dB)} = \text{Puissance du signal reçu (dBW)} - \text{Puissance de bruit (dBW)}.$$

Cette valeur met en valeur la puissance minimale que l'on doit fournir pour tirer partie de la bande passante. D'autres notions de rapport signal/bruit sont parfois utilisées. En particulier le rapport énergie de bit sur densité spectrale de bruit donne la marge nécessaire pour transmettre un seul bit d'information.

En fait, comme on l'a vu dans le cadre des réseaux mobiles, la valeur du rapport signal sur bruit détermine le débit que l'on pourra utiliser sur le réseau. On a en effet la capacité de Shannon :

$$C \leq \frac{W}{2} \log_2 \left( 1 + \frac{S}{B} \right)$$

avec  $C$  le débit,  $W$  la largeur de bande,  $S/B$  le rapport signal sur bruit et  $K$  la probabilité d'erreur tolérée. On remarque alors que le rapport signal/bruit doit augmenter exponentiellement pour donner une augmentation linéaire du débit. En fait il y a des améliorations possibles de la capacité de Shannon. La polarisation permet d'augmenter la capacité d'un facteur 2. D'autres approches sont possibles pour démultiplier la capacité.

### 8.1.6 Méthode d'accès

Dans les réseaux satellitaires, de même que pour les réseaux mobiles, le lien est un canal radio partagé par toutes les sources et toutes les destinations. Le problème du partage de cette ressource est plus délicat que dans les réseaux terrestres.

On distingue plusieurs approches pour l'accès au lien : le partage statique, le partage par demande et l'accès aléatoire. Dans le partage statique, une quantité fixe des ressources est pré-allouée à une connexion. Les méthodes traditionnelles de partage statique sont :

- l'accès multiple à répartition de fréquence (FDMA - Frequency Division Multiple Access)
- l'accès multiple à répartition de temps (TDMA - Time Division Multiple Access)
- l'accès multiple à répartition par codes (CDMA - Code Division Multiple Access)

Les méthodes de partage par demande sont au contraire dynamiques. Elles allouent des ressources (temps de transmission, fréquence, etc) selon les besoins ponctuels de chaque source et la disponibilité des ressources.

Enfin, dans la méthode par accès aléatoire, les sources peuvent avoir des accès aux liens indépendants. Par conséquent, il peut en résulter des collisions de paquets et donc des pertes.

En pratique, on trouve souvent des méthodes d'accès qui combinent plusieurs approches.

### 8.1.6.1 Systèmes FDMA

On découpe la bande de fréquence en sous bandes. Chaque source a sa propre bande de fréquence où elle peut émettre indépendamment des autres. Cette méthode souffre de problèmes d'intermodulation entre les fréquences, qui croissent rapidement avec la puissance utile. On peut alors perdre jusqu'à la moitié de la capacité de transmission par rapport à un accès unique. Pour éviter ce problème on laisse des "trous" entre les bandes pour améliorer l'atténuation d'autres fréquences. L'avantage est qu'elle n'a pas besoin de synchronisation temporelle entre les sources.

### 8.1.6.2 Systèmes TDMA

On définit ici un cycle, aussi appelé trame, découpé en tranches de temps. Dans chaque cycle, chaque source transmet pendant une tranche de temps prédéterminée. Pour pouvoir allouer une bande passante différente à des sources distinctes, on définit aussi quelquefois une version généralisée : une source peut transmettre pendant plus d'une tranche de temps. Les inconvénients de cette méthode sont la nécessité d'une synchronisation temporelle entre les sources et le fait que cette méthode demande des puissances instantanées de transmission plus élevées. En effet, si on veut transmettre à une puissance moyenne donnée et que la transmission n'est possible que pendant une fraction  $a$  de temps, il faut transmettre pendant cette période une puissance supérieure d'un facteur  $1/a$ .

### 8.1.6.3 Systèmes CDMA

Imaginons que l'on veuille transmettre une séquence de bits, où les 0 et les 1 sont représentés par des valeurs de  $-1$  et  $1$  respectivement. La durée de transmission d'un bit d'information (appelée temps bit) est divisé en intervalles de temps appelés chips. Le nombre de chips par temps bit est typiquement  $N=64$  ou  $128$ .

Chaque source  $a$  utilise une séquence  $a^i$  de  $N$  chips. Pour transmettre un 0, la source émet la séquence  $a^i$ , et pour un 1, elle transmet la séquence  $-a^i$ . Les séquences sont orthogonales, c'est-à-dire qu'elles vérifient  $\forall s, t, s \neq t, a^s \cdot a^t = \sum_{i=1}^N a_i^s \cdot a_i^t = 0$ . Elles sont également orthonormées :  $\forall s, (a^s)^2 = \sum_{i=1}^N (a_i^s)^2 = 1$ . Supposons que deux sources  $s$  et  $t$  transmettent en même temps, les bits  $b^s$  et  $b^t$ . A la fin des  $N$  chips, la destination a reçu  $b^s \cdot a^s + b^t \cdot a^t = b^s$ .

C'est ainsi que, alors que les sources émettent à la même fréquence et pendant le même intervalle de temps, la destination peut séparer les informations par simple calcul de produit scalaire. Cependant, cette technique implique une synchronisation des sources.

En pratique, les séquences de chips multipliées par les bits d'informations sont souvent utilisées pour moduler un signal sinusoidal de fréquence élevée  $f$  avant la transmission du signal.

Le CDMA a des avantages par rapport au TDMA et FDMA. Il notamment plus robuste aux phénomènes d'atténuation par des multi-chemins et aux interférences.

## 8.2 Précédents travaux sur l'allocation de slot dans les systèmes TDMA et MF-TDMA

De nombreux articles traitant des problèmes d'allocation de slots dans les systèmes TDMA et MF-TDMA se trouvent dans la littérature. Nous en donnons ici un bref aperçu.

Nous présentons dans la sous-section suivante le problème d'allocation de slots (TSA). La sous-section 8.2.2 présente un deuxième problème souvent étudié dans la littérature : l'affectation des burst. Nous mentionnons ensuite quelques problèmes annexes, et citons dans la dernière sous-section des articles présentant des systèmes satellitaires généraux.

### 8.2.1 Le problème TSA

Dans un système SS/TDMA (Satellite Switch Time Division Multiple Access), les stations terrestres communiquent les unes aux autres par l'intermédiaire du satellite. On distingue donc 2 types de faisceaux, correspondants aux liens montants et descendants. Les spots correspondants aux deux types de liens peuvent être différents. Ainsi un switch embarqué dans le satellite spécifie alors les interconnexions entre les faisceaux montants et les faisceaux descendants sur une base de un pour un.

Dans la méthode de multiplexage en temps, chaque station demande une quantité différente de temps pour transmettre de l'information à une autre station terrestre. Les stations d'un même faisceau communiquent les unes après les autres. La façon dont cet agencement est fait constitue ce que l'on appelle l'allocation des burst (cf. sous-section 8.2.2). En fait, chaque trame TDMA est divisée en un certain nombre de *time-slots*. La demande de chaque station s'exprime alors en un nombre entier de time-slots. Autrement dit, un time slot représente l'unité de trafic écoulé dans chaque canal pendant une unité de temps.

$d_{i,j}$  est la quantité de temps requise pour un flot de données du faisceau montant  $i$  pour être routé vers le faisceau descendant  $j$  dans la trame TDMA. La matrice  $D = (d_{i,j})$  est appelée *matrice de trafic*. On note que  $D$  est de taille  $m \times n$  où  $m$  est le nombre de canaux montants et  $n$  le nombre de canaux descendants.

Le problème d'allocation de time-slot (TSA) consiste à trouver une allocation des time-slots sans conflit dans chaque canal (ou faisceau) descendant de façon à minimiser le temps nécessaire au routage. Les entrées du problème sont le nombre d'unités de trafic à transmettre de chaque canal montant vers chaque canal descendant. Un algorithme TSA

donne alors la longueur de la trame nécessaire ainsi que la configuration du switch pendant chaque time-slot.

### 8.2.1.1 Expression matricielle

Un switch TDM est représenté à la figure 8.2. Sur l'exemple proposé, il y a 4 faisceaux dans chaque sens. On voit que dans le premier time-slot, les données issues des liens montants 1 et 4 doivent être routées vers le lien descendant 2. En conséquence, ils ne peuvent être transmis simultanément. La longueur de la trame solution sera de taille strictement supérieure à 1.

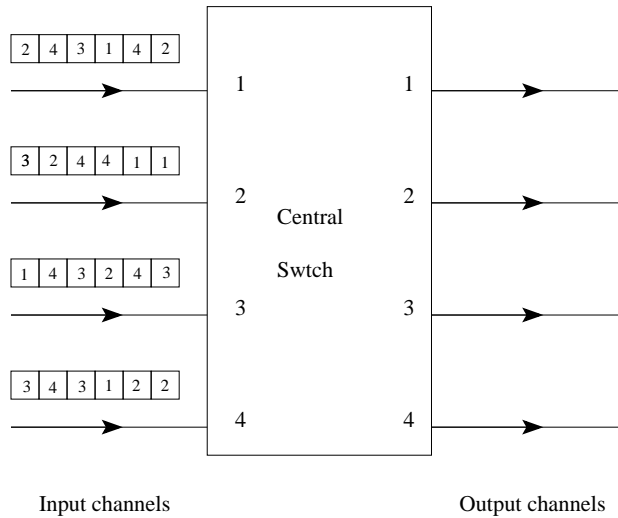


FIG. 8.2 – Un switch TDM classique

**Définition 8.1 (Matrices de trafic et de mode)** Une matrice de trafic  $D$  est une matrice de taille  $m \times n$  dans laquelle chaque élément est un entier. L'élément  $d_{i,j}$  représente le trafic, exprimé en time-slots, de la zone  $i$  à la zone  $j$ .

Une telle matrice est appelée **matrice de mode** ou **matrice de switch** si chaque ligne et chaque colonne comprend au plus un élément non nul. Elle correspond alors à une configuration possible du switch. Son élément maximal est la durée du mode.

Dans notre exemple, la matrice de trafic est :

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 1 \end{pmatrix}.$$

Le fait que dans une matrice de mode au plus un élément par colonne est non nul traduit le fait que l'on ne peut router simultanément le trafic de plusieurs lien entrant sur un lien sortant. On ne peut avoir plus d'un élément non nul par ligne également car on ne peut router simultanément deux time-slot d'un même canal montant.

Le problème TSA est alors d'exprimer une matrice de trafic donnée comme somme de matrices de mode de façon à minimiser la durée du mode, i.e. la somme des modes



des matrices de switch de la décomposition. De plus, on essaye quelquefois de minimiser le nombre de matrices de switch contenus dans la décomposition car le passage d'un mode à l'autre introduit des délais.

Plus formellement, si  $D$  est la matrice de trafic, on cherche une séquence de matrices de modes  $D_i$ , telle que :  $D = \sum_i D_i$  de façon à minimiser  $\sum_i |D_i|$ .

Dans notre exemple, une décomposition possible est :

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Dans cette décomposition, le mode est alors de 7. On voit que cette décomposition est optimale car la somme des éléments de la quatrième colonne, c'est-à-dire le nombre de time-slots de trafic destiné au 4<sup>ème</sup> faisceau descendant vaut également 7. Il est donc impossible de trouver une décomposition qui utilise une trame plus courte.

On peut résoudre le problème de la décomposition en temps polynomial. En revanche, la recherche d'une décomposition de mode minimal est un problème NP-complet. Les solutions proposées dans la littérature sont en général des méthodes itératives : à chaque itération, on regarde le trafic résiduel (c'est-à-dire non encore alloué) et l'on sélectionne une nouvelle matrice de mode que l'on rajoute à la liste déjà obtenue. On soustrait alors cette matrice de mode à la matrice de trafic résiduel et on réitère le processus avec la nouvelle matrice de trafic.

### 8.2.1.2 Premières méthodes de résolutions

L'un des plus anciens articles traitant du problème TSA est [51]. Il utilise un algorithme glouton et donne une solution (itérative) du problème dans le cas d'un système symétrique. Le nombre de modes de switch est alors borné par  $n^2 - 2n + 2$ . Une amélioration de cet algorithme est parue quelques années plus tard dans [101].

Dans [18] on introduit les problèmes dus à la limitation du nombre de transpondeurs. Un transpondeur est l'équipement associé à chaque antenne. Il a pour but de produire certaines fonctions (amplification, translation de fréquence, etc). Il se compose de 2 parties : une unité de réception et une unité de transmission. Ainsi, lorsque le nombre de transpondeurs est égal à  $K$ , on peut router au maximum  $K$  time-slots simultanément. Cela revient à dire que dans une matrice de switch le nombre d'éléments non nul est limité à  $K$ . On remarque que dans le cas de  $r$  transpondeurs il y a exactement  $r!$  matrices de modes de taille 1 possibles, correspondant au nombre possibles de permutations de  $r$  éléments. [18] présente alors un algorithme résolvant le problème TSA avec  $K$  transpondeurs. Le nombre de matrices de mode est borné par  $N^2 - N + 1$  dans le cas symétrique (c'est-à-dire de  $M = N = K$ ) et  $MN + K + 1$  dans le cas général. De nombreux résultats numériques générés à partir de matrices de trafic aléatoires sont présentés. Ils montrent qu'en moyenne, le nombre de matrices de mode est sensiblement inférieur à la valeur des bornes.

### 8.2.1.3 Variantes du problème

#### Minimisation du délai moyen

Gopal, Coppersmith et Wong remarquent dans [41] que ce genre de résolution n'optimise les performances que du point de vue du système (puisque cela revient en fait à maximiser l'utilisation des transpondeurs). Cependant, du point de vue de l'utilisateur il serait intéressant de minimiser le délai moyen entre l'émission et la transmission d'un paquet. Ce délai est entre autre dû au temps séparant le temps de début de trame et l'émission du paquet (dans le lien descendant). Les auteurs proposent alors de regrouper au mieux les paquets en début de la trame descendante pour minimiser ce temps en moyenne. Malheureusement, la solution optimale découle d'une méthode de branch-and-bound, trop coûteuse pour pouvoir être implémenté. Ils proposent alors deux heuristiques.

Pour comprendre le problème dont traité par l'article, revenons à notre petit exemple. La matrice de mode :

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

Est de durée de mode 2, en fait, elle signifie que les configurations du switch sont :

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ pendant un mode et } \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ pendant l'autre.}$$

Ainsi si l'on décompose les matrices de mode utilisées dans la décomposition proposée, on remarque que les matrices obtenues ne sont pas toutes des matrices de permutations. L'article propose donc d'utiliser des matrices 0 – 1 les plus "pleines" possibles en début de décompositions, c'est-à-dire au nombre de 1 maximal.

#### Limitation du nombre de modes de switch

Dans [70], les auteurs proposent de ne considérer un faible nombre de configuration de switch (en  $\mathcal{O}(n)$  au lieu de  $\mathcal{O}(n!)$  comme précédemment considéré). Ainsi, le nombre de configuration possibles du switch embarqué est considérablement réduit et donc les délais de changement de configuration également. Le nombre exact de matrices de switch qu'ils considèrent est de  $2n$ . Évidemment, en ne considérant qu'un nombre réduit de matrices de switch, le temps minimal de routage est augmenté par rapport aux algorithmes précédents. L'article présente une borne du temps ainsi perdu.

#### Minimisation des reconfigurations

Dans [42] les auteurs font l'hypothèse que le temps nécessaire au switch pour se reconfigurer

est grand par rapport au temps de mode. En effet, la durée de routage d'une trame peut s'écrire comme étant  $L + st$  avec  $L$  le temps de transmission,  $s$  le nombre de reconfiguration du switch et  $t$  le temps de reconfiguration. Alors que les articles précédents faisaient l'hypothèse que  $t$  était faible par rapport à  $L$ , cet article fait l'hypothèse contraire. Ils montrent d'abord que ce problème est NP-complet et proposent une heuristique basée sur la théorie des graphes minimisant le nombre de reconfiguration du switch. Leur algorithme trouve une trame avec un minimum de reconfigurations. Parmi l'ensemble des solutions trouvées, ils considèrent ensuite celle qui minimise le temps de transmission. Ils valident leur algorithme par des expérimentations numériques.

### Trafic hétérogène

Bonuccelli, Gopal et Wong considèrent dans [20] un système à deux types de trafic. L'un est du trafic de voix, supposé relativement statique. Le second est un trafic de données, constitué par nature de rafales. L'article suppose qu'une première trame TDMA est donnée, permettant de router le trafic de voix et dans laquelle, trame par trame, on essaye d'incorporer le trafic de données. Ainsi, une trame TDMA partiellement remplie et un trafic de paquet en attente de transmission étant donnés, le problème est de construire une nouvelle trame TDMA qui ne change l'allocation d'aucun des times slots déjà alloué mais alloue autant de paquets que possible pour remplir les slots vides. Ils montrent que ce problème est NP-complet et proposent 3 solutions heuristiques. Les méthodes sont itératives, c'est-à-dire qu'à chaque étape, une nouvelle matrice de switch est proposée. Seule la façon dont elle est choisie diffère d'une heuristique à l'autre. Elles sont toutes basées sur une relaxation Lagrangienne.

### Extension au multicast

Dans [29], les auteurs étendent le problème du TSA au système multicast. Le problème multicast peut être résolu en utilisant les techniques vues dans le cas unicast si les packets multicasts sont autorisés à être scindés en packets unicasts. Dans le cas contraire, on ne peut plus utiliser une représentation matricielle du trafic. A la place, chaque connexion se représente par un vecteur. Deux heuristiques sont proposées, l'une basée sur une idée de "divide and conquer", la seconde sur un algorithme glouton. Le cas particulier où l'ensemble des destinataires de deux sessions multicast est identique ou disjoints fait l'objet d'une étude spécifique. On peut noter que même ce cas particulier est NP-complet.

#### 8.2.1.4 TDM hiérarchiques

Les systèmes hiérarchiques sont présentés dans [36]. Les utilisateurs finaux utilisant une connexion à faible bande passante sont regroupés entre eux par des multiplexeurs. L'avantage des systèmes hiérarchiques est qu'ils permettent de réduire le coût des systèmes par la maîtrise de la taille du switch. En outre, la croissance de tels système est modulaire. Un tel système est représenté à la figure 8.2.1.4.

L'article [36] présente une série de conditions nécessaires et suffisantes sur le trafic des utilisateurs pour que l'on puisse affirmer qu'il existe un TSA valide pour acheminer le trafic

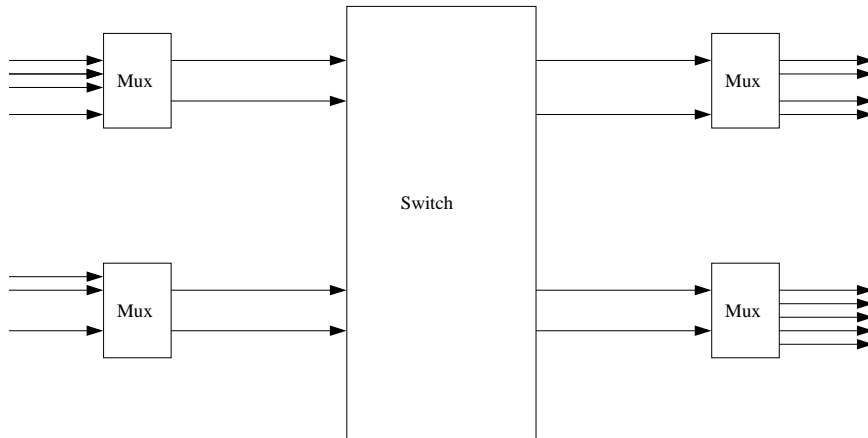


FIG. 8.3 – Un exemple de système TDMA hiérarchique

entre tous les utilisateurs finaux.

Dans [19], Bonuccelli s'intéresse au cas symétrique ( $M$  utilisateurs finaux aussi bien à l'entrée qu'à la sortie et un switch central de taille  $N \times N$ ). Il propose dans son article un algorithme itératif de complexité  $\mathcal{O}(M^5)$ . En considérant à chaque étape le réseau associé à la matrice de trafic, il se ramène à un problème circulation<sup>1</sup> sur ce réseau. Une version améliorée a été proposé par Tham dans [110] qui utilise les propriétés des graphes bipartites<sup>2</sup>. La complexité est alors de  $\mathcal{O}(M^4)$ . Un article très proche est [109].

Les auteurs de [39] proposent une heuristique basée sur l'utilisation de réseaux de neurones. Leur algorithme tourne en parallèle sur  $n^2 \times m$  processors, où  $m$  est le nombre de times-slots du problème, et  $n$  la taille du système ( $n \times n$  TDM système).

Dans [119], Yeung propose un algorithme itératif permettant de résoudre le problème en  $\mathcal{O}(N^2)$  pour le problème non-hiérarchique et en  $\mathcal{O}(M^2)$  pour le problème hiérarchique. Ici  $N$  représente la taille du switch (symétrique) et  $M$  le nombre d'utilisateurs finaux (symétriques également). L'idée est proche de celle de Inukai dans [51] (dans le cadre des systèmes non hiérarchiques). A chaque itération, la matrice de mode est construite de sorte à prendre les éléments des entrées critiques, c'est-à-dire des lignes et colonnes dont la somme est maximale. L'intérêt de cette heuristique est sa simplicité d'implémentation. Leurs simulations tendent à montrer que les résultats obtenus sont très proches de la solution optimale.

**Exemple 8.1** (*méthode des éléments critiques*) On considère la matrice suivante :

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

<sup>1</sup>La circulation dans un réseau est l'allocation de nombres à des arcs (les liens) de sorte que :

- le nombre alloué à un arc (le flot) est compris entre une borne minimale et la capacité,
- à chaque nœud, la somme des flots des arcs entrants est égale à celle des arcs sortants (loi de conservation).

<sup>2</sup>Un graphe bipartite est un graphe dont les nœuds peuvent être divisés en deux groupes  $A$  et  $B$ . Il n'existe alors que des arcs joignant un nœud du groupe  $A$  au groupe  $B$ .

Une solution optimale se trouve en 6 unités de temps. L'idée est de considérer à chaque fois la matrice de mode unitaire contenant les éléments critiques de la matrice trafic, c'est-à-dire situés sur les lignes et les colonnes dont la somme des éléments est la plus maximale.

Je représente en italique les éléments correspondants. La décomposition consiste alors à prendre à chaque étape la matrice de mode contenant des 1 aux coordonnées des éléments grisés et des 0 partout ailleurs.

$$\begin{pmatrix} 1 & \mathbf{2} & 1 & 0 \\ 2 & 0 & \mathbf{2} & 1 \\ 2 & 1 & 1 & \mathbf{2} \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{1} & 1 & 0 \\ 2 & 0 & 1 & \mathbf{1} \\ \mathbf{2} & 1 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & \mathbf{1} & 0 \\ \mathbf{1} & 1 & 1 & 1 \\ 0 & 0 & 0 & \mathbf{3} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & \mathbf{1} & 0 \\ \mathbf{2} & 0 & 0 & 0 \\ 0 & 1 & 1 & \mathbf{1} \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

### 8.2.1.5 Systèmes à faisceaux de bande passante différentes

Le système étudié dans [27] est appelé système SS/TDMA avec faisceaux de taille variable (voir Fig. 8.4). L'idée est d'utiliser des faisceaux de bande passante  $\alpha_i$  différentes. On ne peut malheureusement pas appliquer les méthodes utilisées dans les systèmes TDMA hiérarchiques pour résoudre de tels problèmes. Chalasani et Varma proposent alors deux algorithmes. Pour cela ils modélisent le problème de TSA comme un problème de flot de réseau, tout comme Bonuccelli dans [19]. L'un de leur algorithme est séquentiel, de complexité  $\mathcal{O}(M + N)^2 \cdot \min(MN\alpha, L)$ , où  $M$  est le nombre de liens montants,  $N$  celui de liens descendants,  $\alpha$  la bande passante maximale d'un lien montant ou descendant,  $L$  la longueur de la trame. Leur second algorithme tournant en parallèle sur  $L/2$  processeurs et une complexité de  $\mathcal{O}((M + N)^3 \log(L))$ .

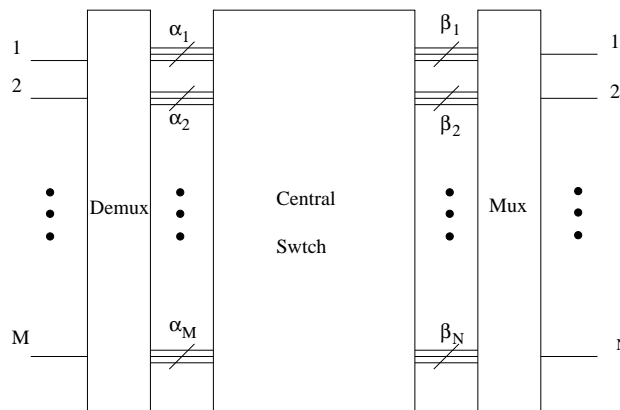


FIG. 8.4 – SS/TDMA avec faisceau de taille variable

## 8.2.2 Allocation des burst

On considère ici le problème consistant à allouer aux utilisateurs un ou des time-slots dans le lien montant.

### Approche classique

Une solution est proposée dans [96] qui prend en compte la qualité de service (QoS). Les auteurs mesurent la qualité d'un lien par la mesure du débit et du BER (Bit Error Rate). Chaque terminal calcule alors le nombre de time-slot dont il a besoin pour satisfaire sa QoS. Dans le système MF-TDMA qu'ils considèrent, la bande passante est divisée en 32 porteuses (canaux) et les trames se composent de 70 slots. Les slots sont alloués par groupes, appelés "burst". Chaque terminal a besoin d'un certain nombre de time-slots. Il faut satisfaire ses demandes et qu'en plus les times-slots d'une même source soient placés consécutivement sur une même porteuse. Les auteurs adaptent alors une heuristique classique de résolution de problème de "bin-packing" à ce type de problème.

Dans [52], les stations terrestres ont un transmetteur/émetteur qui peut passer d'une porteuse à l'autre (cette technique est appelée sauts de porteuse) en émission et en réception. La contrainte est qu'une station ne peut pas recevoir plus qu'une porteuse à la fois. En pratique, le temps que la station nécessite pour changer de fréquence est négligeable. On peut alors considérer deux types de systèmes : SCPB (Single Channel Per Burst) et MCPB (Multiple Channels per Burst). Dans le deuxième cas, on autorise une station à changer de canal au sein même d'un burst.

### Résolution jointe du problème d'allocation de burst et de TSA.

Dans [109], l'auteur considère le système suivant. La zone couverte par un satellite est divisée en "spots". Il y a en fait 2 partitionnements de l'espace, l'un pour le lien montant, l'autre pour le lien descendant. Au sein de chaque spot les stations peuvent utiliser un certain nombre de porteuses. Les porteuses utilisées pour les liens montants et descendants sont distinctes. La question est comment allouer les porteuses aux stations et pendant quel intervalle de temps de façon à minimiser la taille du mode de la matrice de trafic. Dans cette approche, les deux problèmes (de burst assignment et de TSA) sont donc considérés de façon jointe.

L'idée est de regarder la matrice de trafic liant chaque source à chaque destination. Sur chaque spot, le maximum des sommes des valeurs de chaque ligne et de chaque colonne de la matrice correspondant à chaque spot concerné divisé par le nombre de porteuses disponibles représente une borne supérieure du temps de routage nécessaire. Sur chaque spot (d'émission et de réception) l'algorithme découpe alors les stations en groupes dont le trafic sera inférieur à la borne obtenue. Le mécanisme est itératif. A chaque étape on détermine le temps de transmission maximal associé à un groupe de façon à minimiser dans le TSA le nombre de reconfigurations du switch.

### Variable Channel Per Burst - VCPB

[116] étudie le cas d'un système SS-TDMA variable-channel-per-burst (VCPB). Un certain

nombre d'appels sont en cours sur le réseau. Des "rafales" de trafic sont alloués aux stations terrestres une à une. Au début de chaque rafale se trouve une entête (overhead) et deux rafales sont séparées par un intervalle de sécurité. On augmente ainsi l'efficacité du lien en minimisant le nombre de rafales, ou, de façon équivalente en utilisant pour chaque station des rafales aussi grandes que possible, c'est-à-dire en regroupant le plus possible les émissions de chaque station.

L'idée de [116] est de transférer les slots inutilisés d'une communication à une autre n'en n'ayant pas assez. Ceci permet une meilleure utilisation du canal, car plus de time-slots sont alors actifs et que les rafales sont en moyenne plus longues. Il faut cependant minimiser le nombre de reconfigurations (car elles sont coûteuses en temps) et effectuer les reconfigurations sans interrompre l'appel en cours. L'article présente différentes stratégies de reconfigurations dont ils évaluent la fréquence ainsi que l'efficacité d'utilisation du canal.

### 8.2.3 Problèmes annexes

#### Spots interférents.

[40] considère un problème SS/TDMA avec spots interférents. Les auteurs proposent une approche en deux étapes. Dans la première, ils allouent des polarisations orthogonales aux stations pour réduire les interférences. L'allocation proprement dite des time-slots aux stations constitue la seconde phase (on considère ici un système SS/TDMA, il y a donc une porteuse par spot). La première étape est solvable en temps polynomial dans la plupart des cas. On suppose que les spots interfèrent 2 à 2. Ainsi, le but de l'allocation est de faire que 2 zones interférentes n'émettent pas en même temps. Le problème d'interférence est modélisé par un graphe dont les nœuds sont les zones. Il existe une arête entre deux nœuds si et seulement si ils interfèrent entre eux. L'objectif est toujours de minimiser la durée de l'allocation sur la trame. L'article propose des heuristiques.

#### Synchronisation des sources

Dans [24], les auteurs considèrent le problème pour les différentes sources de se synchroniser entre elles afin qu'elles transmettent leur rafale de sorte que le signal atteigne le satellite à l'instant désiré dans la trame TDMA. La difficulté du problème réside dans le fait que le temps de transmission entre les sources et le satellite dépend de leur position géographique.

### 8.2.4 Surveys et description de systèmes complets

#### 8.2.4.1 Surveys sur les systèmes satellitaires.

[120] présente une description sommaire des satellites utilisant la bande Ka classés par pays (cette étude date de 1999). [16] décrit en particulier les atténuations dues aux conditions atmosphériques dans de tels systèmes.

[58] présente les problèmes technologiques liés à l'embarquement des systèmes de traitements de signal (en matière de poids, de consommation d'énergie, etc). L'étude date

malheureusement de 1987 mais présente les évolutions technologiques envisagées à cette époque.

Enfin, on peut noter que dès 1976, Sinha présente, dans [106] une description des systèmes satellitaires TDMA. En particulier il introduit les équipements terrestres et embarqués, et donne une première formulation mathématique du problème d'allocation des burst et du TSA.

#### 8.2.4.2 Description globale de systèmes

##### Intelsat

Minoux et Brouder présentent dans [83] un réseau basé sur les systèmes Intelsat. Ils décomposent le problème d'allocation de ressources en 3 problèmes afin d'en faciliter la résolution.

- Affectation des stations de base aux transpondeurs
- Agrégation des différents flots de trafic au niveau de chaque station de base
- Résolution du problème de TSA.

La taille maximale du système qu'ils considèrent est de 19 stations de base et 6 transpondeurs. Il y a deux sortes de trafic : le trafic de données et celui de voix. La résolution du premier problème est statique. La deuxième étape est introduite afin de diminuer le nombre de blocs dans la trame et donc limiter la taille du troisième problème. L'idée est d'agréger les trafics de voix et donc de tenir compte du fait qu'en moyenne un utilisateur n'est actif que 40% du temps dans une conversation téléphonique. La résolution de leur problème pour le TSA est différente légèrement des approches standards car ils utilisent le fait que les sous-blocs obtenus à la partie 2 sont regroupés ensemble avant d'être placés dans la trame. La résolution qu'ils proposent est itérative. La construction d'une matrice de mode se fait en trois étapes :

- détermination du sous-bloc critique, c'est-à-dire celui de plus grande taille,
- détermination d'un ensemble de sous-blocs dont la somme est la plus grande possible, tout en restant inférieure à celle du bloc critique par programmation programmation dynamique,
- heuristique de type glouton pour trouver parmi cet ensemble un ensemble indépendant (c'est-à-dire dont les éléments n'appartiennent pas à la même ligne ou à la même colonne).

Une description complète des systèmes Intelsat se trouve dans [63]. On y trouve en particulier une description de la composition de la trame, des 2 types de rafales (de données et de parole), et la méthode pour agréger les flots au niveau des stations émettrices, les équipements utilisés dans le système.

##### Services RNIS<sup>3</sup> et ATM

Dans [59] les auteurs s'intéressent aux choix technologiques nécessaires dans le cas d'un système satellitaire avec des services RNIS. Le système offre aux utilisateurs un ensemble de

---

<sup>3</sup>RNIS, ou Réseau Numérique à Intégration de Services, ISDN en anglais



débits entre 64 et 1536 kbps. Ils proposent dans un tel système d'utiliser des transpondeurs à sauts de fréquence, similaires à ceux utilisés dans Intelsat. Tout se passe donc comme si l'on avait plusieurs systèmes TDMA synchronisés. Un problème de tels systèmes est l'interférence générée entre les canaux adjacents. En contrepartie, ils permettent d'augmenter la capacité du système et de donner de la flexibilité au réseau. Le déploiement du réseau ainsi décrit devait avoir lieu à la fin de l'année 1991.

Un autre article intéressant est [103]. Il s'agit ici d'un réseau de type VSAT (Very Small Aperture Terminal) pour des services ATM.

### Systeme ACTS

[95] présente un satellite expérimental appelé ACTS (Advanced Communications Technology Satellite). Le système est une combinaison de FDMA et de TDMA et est constitué de 2 spots. Un tel système se compose donc de 2 antennes, chacune visitant périodiquement une série zones terrestres. La durée du cycle de visites est égal à celle de la trame, c'est-à-dire de 1ms. Les slots sont regroupés en burst, eux-même regroupés par zone. Les auteurs calculent ensuite la probabilité de blocage d'appel en fonction de l'algorithme de ré-allocation de ressource et des caractéristiques des appels (durée moyenne, taux d'arrivée, etc). Chaque communication dispose d'un certain nombre de time-slots. L'idée est de comparer les façons possibles de réallouer des times-slots affectée à une connexion qui ne les utilise pas à une autre connexion qui n'en n'a pas assez.

### Autres systèmes

#### **Systeme à 2 types de données ou d'utilisateurs**

[35] présente une analyse de performance pour calculer le taux de perte des paquets, les blocages d'appel et les retards de paquets d'un utilisateur typique dans un environnement regroupant des trafics de voix, données et vidéo.

Dans le lien montant, 2 types d'utilisateurs sont considérés : des utilisateurs de slotted ALOHA et d'autres de DAMA (Demand Assignment Multiple Access). Chaque utilisateur DAMA fait une demande de réservation d'un certain nombre de time-slots (ce nombre peut être différent suivant qu'il s'agit de voix ou de video par exemple). Si il y a collision avec une autre demande, sa requête sera réitérée après un certain temps (aléatoire). Les utilisateurs de slotted Aloha (accès sans connexion) écoutent en permanence le signal descendant pour savoir quelles sont les réservations en cours. L'article présente alors, en fonction de la loi d'arrivée du trafic, la probabilité de blocage en utilisant une chaîne de Markov.

A bord du satellite se trouve un switch TDM. Dans le cas où plusieurs types de trafic sont destinés au même lien descendant on applique des règles de priorité (par exemple la vidéo est prioritaire sur la voix). L'article présente alors également les probabilités de blocage au niveau du switch.

#### **Systeme MC-TDMA**

[84] présente un système complet d'étude de système MC-TDMA. L'allocation est décomposée en différentes étapes. Ainsi, le trafic entre les différentes régions étant connu ainsi que les capacités des transpondeurs, comment associer le trafic aux spots ? (on suppose ici que les spots se recouvrent partiellement). La solution proposée repose sur une formulation en

programmation linéaire.

Le second problème est celui de la détermination de la séquence de switch, correspondant au problème déjà très largement étudié dans la littérature. Les auteurs proposent de reprendre une méthode trouvée dans la littérature puis de réarranger les modes de switch de façon à minimiser les changements de configurations. Pour ce faire, ils considèrent un graphe dans lequel chaque nœud est un mode et chaque arc représente la distance qui sépare les deux modes associés (mesurée par le nombre d'états différents dans le switch). Le problème est alors ramené au classique problème du voyageur de commerce.

Les auteurs décrivent également le problème de l'ordonnancement de rafales. Chaque source peut émettre par rafale, de taille maximale fixée. La source doit minimiser le nombre de rafale qu'elle utilise, c'est pourquoi elle remplit au maximum la rafale disponible en agrégeant éventuellement des demandes. Ainsi, des rafales multi-destinations peuvent être créées. Le problème se ramène à un problème de bin-packing.

Enfin, un réarrangement des rafales est nécessaire afin que deux bursts ne soient pas émis en même temps sur une même porteuse et qu'une source (ou une destination) n'ait pas à émettre (respectivement à recevoir) de burst simultanément sur 2 porteuses différentes. Ils proposent pour ce dernier problème une méthode de branch-and-bound.

## 8.3 Notre problème

Je définis ici le problème que nous nous proposons d'étudier dans cette partie.

### 8.3.1 Formulation du problème

Nous considérons un système géostationnaire MF/TDMA utilisant la bande Ka. Le système se compose d'un certain nombre de spots. Les utilisateurs de chaque spot peuvent utiliser n'importe laquelle des porteuses du système.

On suppose que les différentes porteuses utilisées sont de types c'est-à-dire de bandes passantes différentes. Les utilisateurs du système ont donc des terminaux utilisant des débits différents. Nous considérerons dans la pratique 5 types de porteuses.

Les données du système sont les demandes des opérateurs (en terme de nombre minimal et maximal de porteuses) ainsi que des poids associés aux demandes. Ces poids peuvent par exemple être choisis en fonction de la zone associée à la demande, et donc à la qualité de transmission. Notre algorithme donne alors une allocation équitable en terme de nombre de time-slots pour chaque opérateur, dans chaque spot et pour chaque type de porteuse. Nous proposons en outre un positionnement possible de ces time-slot dans l'espace à deux dimensions constitué de la bande passante disponible et de la durée d'une trame.

Nous voyons que notre problème est plus complexe que les systèmes décrits dans la section précédente dans la mesure où les différentes difficultés sont regroupées : multi-spots, multi-porteuses, multi-type de porteuses et modèle d'interférence plus complexe. A cela s'ajoute des considérations d'équité.

Le problème a été découpé en deux parties, faisant chacune l'objet d'un chapitre de la thèse. Dans la première phase de l'allocation, nous proposons un algorithme calculant, selon le critère d'équité choisi, le nombre de porteuses de chaque type dans chaque spot, pour

chaque opérateur en fonction de sa position géographique, mais sans prendre en compte les phénomènes d'interférences.

Dans une seconde phase, nous proposons une méthode pour calculer le nombre de time-slot ainsi que la disposition des time-slots dans le plan de fréquence de façon à satisfaire l'allocation trouvée dans la première partie.

## 8.3.2 Système physique

### 8.3.2.1 Spots et réutilisation spatiale

La zone géographique couverte par le satellites est divisée en régions hexagonales appelées *spots*, eux-même divisés en *zones*. Un faisceau différent est utilisé pour chaque spot. La subdivision en zone est utile pour prendre en compte les problèmes d'interférences ou d'atténuations. Nous y reviendrons dans la suite.

Chaque spot utilise une certaine bande de fréquence, appelée couleur. Je note  $S(c)$  l'ensemble des spots d'une même couleur  $c$ . En pratique, le nombre de couleurs est inférieur au nombre de spots. Or deux spots ayant la même couleur interfèrent. Il y a donc une affectation statique des couleurs aux spots, que nous ne considérons pas ici. L'affectation a pour but d'espacer au maximum les spots de même couleur. Un exemple d'une telle allocation est présentée dans la figure 8.5 avec ici 4 couleurs dans un système de 14 spots.

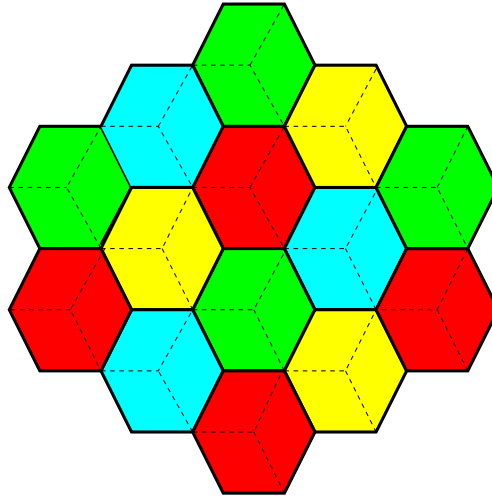


FIG. 8.5 – Exemple d'emprunte de satellite avec 14 spots, 4 couleurs et 3 zones par spots.

Deux spots ayant des couleurs différentes sont clairement associés à des problèmes différents. Dans la suite nous calculerons donc les allocations indépendamment sur chaque groupe de couleurs.

### 8.3.2.2 Zones

La subdivision en zone peut avoir deux rôles différents. Ainsi, celle utilisée dans la première partie peut ou non coïncider à celle de la deuxième partie.

## Atténuation du signal

Nous avons vu qu'un inconvénient de la bande Ka était sa sensibilité à la pluie. Deux utilisateurs se trouvant dans une zone à des endroits différents peuvent être soumis à des conditions atmosphériques différents. D'une façon générale, la localisation géographique des utilisateurs dans un même spot peut avoir des conséquences sur la qualité de son lien. Pour prendre ceci en considération, nous partageons les spots en zones de petites tailles, donc soumises aux mêmes conditions de liens. Ainsi, on peut supposer qu'en tous les points d'une même zone, les opérateurs utilisent le même taux de codage.

### **Interférences**

L'utilisation de couleurs permet évidemment de réduire, mais non de supprimer les interférences. On peut naturellement penser que la position géographique d'un terminal à l'intérieur d'un spot a une influence sur les spots avec qui ils va interférer. Nous introduisons ainsi une seconde division de spots en zones pour permettre de prendre en compte ce phénomène.

On peut remarquer que les interférences sont prises en compte de façon plus fine que dans [40] puisque l'on ne considère plus un système binaire où deux spots interfèrent ou non. A la place, nous considérons que tous les spots interfèrent entre eux, mais qu'une communication peut néanmoins être maintenue si le taux d'interférence est inférieur à un certain seuil.



# Chapitre 9

## Première partie de l'allocation : la planification radio.

Ce chapitre présente la résolution de la première phase de l'allocation. Cette dernière consiste en l'allocation des porteuses à des fournisseurs de service (opérateurs) qui sont situés dans différentes zones de plusieurs spots. L'allocation de porteuses prend en compte les demandes de débit minimum des opérateurs et la charge supplémentaire dûe au taux de codage qui peut varier d'une zone à l'autre à cause des différences de conditions atmosphériques. Le but est de trouver une solution équitable avec des propriétés d'optimalité. La solution est l'entrée de la seconde phase, dite d'allocation de slot, que nous présentons dans le chapitre suivant.

Nous proposons une méthode combinatoire pour obtenir une allocation optimale équitable en temps polynomial en prenant en compte les problèmes pratiques comme des contraintes d'intégrité inhérente au système.

La section suivante introduit les notions et notations relatives à ce problème. Nous détaillons ensuite l'algorithme proposé. Nous présentons enfin quelques résultats numériques.

### 9.1 Introduction et notations

Nous considérons un système satellite géostationnaire dans lequel un manager veut allouer de la bande passante entre différents opérateurs qui opèrent dans des régions géographiques différentes. Une unité de bande passante allouée peut correspondre à différentes quantités de débit. Ceci est dû à différents facteurs, et en particulier aux conditions météorologiques. En effet, pendant les périodes de mauvais temps dans une région, un opérateur local peut avoir besoin d'utiliser une plus grande partie de sa bande passante pour de l'information redondante (un plus fort taux de codage pour la correction d'erreur), et ainsi le débit effectif d'information diminue. Ainsi, l'on maximise le débit total, certains opérateurs ne se verront pas allouer de bande passante si elle peut être allouée à d'autres, se trouvant dans des zones où les conditions météorologiques sont meilleures. Il est alors intéressant de comprendre et de proposer des schémas d'allocations qui sont plus équitables et ne pénalisent pas systématiquement les opérateurs qui souffrent de mauvaises conditions de transmission.

### 9.1.1 Définitions générales

Une des principales difficultés dans la résolution de tels systèmes est que des contraintes d'intégrité peuvent apparaître. Nous supposons en effet que chaque opérateur dans l'ensemble  $O$  des opérateurs peut se voir affecter une ou plusieurs porteuses dans chaque zone dans un ensemble de  $N$  types de porteuses :  $T = \{1, \dots, N\}$ . Soit  $B_t$  la bande passante totale d'une porteuse de type  $t$ . Nous supposons que  $B_1 > \dots > B_N$ .

$B_t$  n'est pas directement proportionnel au débit réel d'information d'une porteuse de type  $t$ . Tout d'abord, comme mentionné précédemment, le débit dépend du taux de codage, qui peut varier d'une région à l'autre à cause des conditions atmosphériques. Deuxièmement, le débit effectif est diminué à cause des overheads utilisés (plus ou moins 10%) pour la signalisation, les marges de fréquences, etc. Ce pourcentage dépend en particulier du type de porteuse. Pour prendre ceci en considération, à chaque porteuse est associée une utilité  $C_t(z, o)$ . L'utilité est la valeur qu'un opérateur  $o$  d'une zone  $z$  est prêt à payer pour une porteuse de type  $t$ . Elle peut être choisie comme une fonction de la quantité de redondance (dans le codage de canal) qui dépend des conditions atmosphériques de chaque zone. L'utilité peut être choisie proportionnellement à la quantité actuelle de débit allouée. Le problème devient ainsi d'allouer équitablement (ou optimalement) le débit.

Nous supposons qu'il y a un nombre minimal et maximal de porteuses de type  $t$  par zone  $z$  demandée par chaque opérateur  $o$ . On les note respectivement  $D_t^{min}(z, o)$  et  $D_t^{max}(z, o)$ . Comme précédemment, on suppose que les demandes minimales peuvent être respectées.

### 9.1.2 Inter-Spot Compatibility Conditions (ISCC)

Afin de simplifier la seconde partie de l'allocation (cf Chapitre 10), nous étudions la possibilité d'ajouter la contrainte suivante, dite de compatibilité d'inter-spot : ISCC (Inter Spot Compatibility Condition). Elle peut être vue comme :

- (i) imposer la même structure de plan de fréquence sur tous les spots de la même couleur et
- (ii) s'autoriser à remplacer la demande d'un opérateur pour une porteuse donnée  $j$  par une porteuse de type  $t < j$  (c'est-à-dire de plus grande bande passante) en le facturant tout de même pour une porteuse  $j$ .

Si un opérateur ne peut utiliser en pratique plus de bande passante qu'il le demande, alors ceci revient à gaspiller de la bande passante.

Une autre façon de définir l'ISCC est de dire que la position des porteuses dans le plan de fréquence est la même pour tous les spots d'une même couleur. En d'autres termes, supposons que  $\mathcal{O}(f)$  est l'ensemble des opérateurs auxquels sont assignés les porteuses qui contiennent une certaine fréquence  $f$ , et soit  $t$  la porteuse avec la plus grande bande passante parmi  $\mathcal{O}(f)$ . Alors les bandes de fréquence pour tous les autres  $t' \in \mathcal{O}(f)$  doivent être contenues dans la bande de fréquence de  $t$ , et un opérateur ne peut être attribué qu'une seule porteuse dans cette bande. Nous illustrons cette définition par la figure Fig. 9.1. Dans la partie supérieure, les spots 1 et 2 ne respectent pas l'ISCC. La partie inférieure de la figure montre les deux mêmes spots avec l'ISCC. Nous voyons que les porteuses sont alors situées aux mêmes positions. Mais le spot 1 contient alors 4 porteuses de type 3 au lieu de

7, et le spot 2 a 7 porteuses de type 2 au lieu de 8 : la contrainte d'ISCC limite l'utilisation de la bande passante.

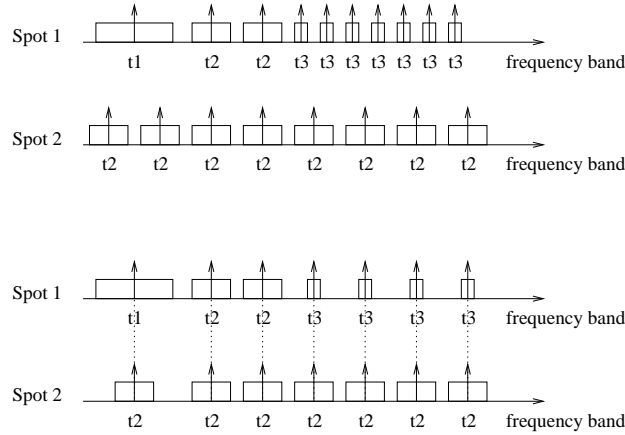


FIG. 9.1 – Deux allocations avec et sans ISCC.

### 9.1.3 Notations (résumé)

Ce tableau résume les notations utilisées dans ce chapitre :

Données du problème :

$o$	un opérateur,
$O$	l'ensemble des opérateurs,
$T$	l'ensemble des types de porteuses ( $ T  = N$ ),
$B_t$	la bande passante d'une porteuse de type $t$ (on suppose $B_1 > \dots > B_n$ ),
$C_t(z, o)$	l'utilité associée à une porteuse de type $t$ pour l'opérateur $o$ dans la zone $z$ ,
$D_t^{\min}(z, o)$ et $D_t^{\max}(z, o)$	les demandes minimales et maximales de l'opérateur $o$ dans la zone $z$ pour des porteuses de type $t$ .

Pour une couleur  $c$  donnée, notre problème est de calculer le vecteur  $J_t, t = 1, \dots, N$  et  $D_t(z, o)$  où  $J_t$  est le nombre de porteuses de type  $t$  dans un spot de couleur  $c$  et  $D_t(z, o)$  est le nombre de porteuses de type  $t$  associées à la zone  $z$  et à l'opérateur  $o$ .

Résultats fournis :

$D_t(z, o)$	Nombre de porteuses de type $t$ attribuées à l'opérateur $o$ dans la zone $z$ ,
$J_t$	nombre de porteuses de chaque type.

### 9.1.4 Critères d'équité

Il y a plusieurs façons possibles de comprendre et implémenter les concepts d'équité dans notre contexte, au niveau géographique. Nous pouvons essayer d'atteindre :

- une équité locale, c'est-à-dire dans chaque zone,
- une équité globale, c'est-à-dire dans l'allocation totale,
- une équité intermédiaire, c'est-à-dire à l'échelle d'une zone géographique plus grande qu'une zone, comme par exemple un spot ou un groupe de spots.



Nous reprenons la formulation générale du critère d'équité (du à [85]). Etant donnée une constante positive  $\alpha \neq 1$ , et une couleur  $c$ , considérons la maximisation de :

$$\frac{1}{1-\alpha} \sum_{o \in O} \left[ \sum_{\substack{s \in S(c), z \in Z(s) \\ t \in T}} (D_t(z, o) - D_t^{min}(z, o)) \right]^{1-\alpha}$$

soumise aux contraintes du problème. Comme la fonction d'utilité est concave et que les contraintes sont linéaires, ceci définit une allocation unique appelée la  $\alpha$ -allocation de débit. C'est en fait une implémentation globale de l'équité.

De la même façon, on peut définir l' $\alpha$ -allocation pondérée de débit comme étant l'unique solution de la maximisation :

$$\frac{1}{1-\alpha} \sum_{o \in O} \left[ \sum_{\substack{s \in S(c), z \in Z(s) \\ t \in T}} C_t(z, o)(D_t(z, o) - D_t^{min}(z, o)) \right]^{1-\alpha}. \quad (9.1)$$

### 9.1.5 Formulation du problème

Nous nous intéressons au critère d'équité obtenu par les solutions de la maximisation de :

$$\sum_{t \in T, s \in S(c), z \in Z(s), o \in O} \frac{\left[ C_t(z, o)(D_t(z, o) - D_t^{min}(z, o)) \right]^{1-\alpha}}{1-\alpha} \quad (9.2)$$

Ceci est en adéquation avec la définition précédente dans le cas de l'optimisation globale. Il l'est également si l'on identifie un sous-opérateur comme étant responsable des demandes d'un type donné pour un opérateur donné dans une zone fixée. Cette équité est en fait appliquée à une échelle locale (c'est-à-dire de chaque zone).

Les contraintes de notre problème peuvent être résumées comme ceci :

$$\begin{aligned} \forall z \in Z(c), \forall o \in O, \forall t \in T, & \quad D_t^{min}(z, o) \leq D_t(z, o) \leq D_t^{max}(z, o), \\ \forall s \in S(c), \forall t \in T, & \quad \sum_{i \in T, s.t. i \leq t} J_i \geq \sum_{z \in Z(s), o \in O} D_t(z, o),^1 \\ \forall t \in T, & \quad J_t \geq 0, \\ & \quad \sum_{i \in T} J_i B_i \leq B. \end{aligned}$$

---

<sup>1</sup>Ceci provient du fait que nous pouvons affecter à un opérateur une porteuse de plus grand type que ce qu'il requiert.

## 9.2 Notre algorithme d'allocation

Cette partie s'intéresse à la résolution du problème formulé dans 9.1.5. L'algorithme se compose de deux phases. Nous supposons dans la première que le vecteur  $J$  est connu, c'est-à-dire que le nombre de porteuses de chaque type est connu et fixé pour tous les spots et déterminons comment les répartir entre les opérateurs. Dans la seconde phase, nous montrons comment déterminer le nombre de porteuses de chaque type. Finalement, nous montrons comment ces deux composantes sont combinées pour obtenir une allocation équitable optimale.

### 9.2.1 Allocation des porteuses aux opérateurs dans un spot à $J$ fixé

A l'intérieur d'un spot  $s$ , soit  $p$  le nombre total de demandes :

$$p = \sum_{z \in Z(s), o \in O, t \in T} D_t^{max}(z, o).$$

Nous définissons alors la notion de *requête*. Si un opérateur a une demande maximale de  $D_t^{(max)}(z, o)$  porteuses de type  $t$  dans la zone  $o$ , nous considérons alors que cela représente  $D_t^{(max)}(z, o)$  requêtes. Notre algorithme doit satisfaire au moins  $D_t^{(min)}(z, o)$  d'entre elles.

L'ensemble des requêtes est  $R$ . On a par définition  $|R| = p$ . Pour une requête  $i$ , on note  $t_i$  son type (où  $t_i \in \{1, \dots, N\}$ ), et on lui associe une *valeur*, notée  $v_i$ . Nous expliquerons en Section 9.2.3 comment choisir les valeurs  $v_i$  des requêtes afin d'obtenir le critère d'équité souhaité. Les requêtes sont triées par valeur décroissante ( $v_1 \geq \dots \geq v_p$ ).

Nous présentons un algorithme qui permet de trouver **Accept**, un sous ensemble de l'ensemble de requêtes  $\{1, \dots, p\}$ , qui satisfait les contraintes des porteuses disponibles par type (c'est-à-dire le nombre donné par  $J_i$ ,  $i = 1, \dots, N$ ) et maximise la valeur  $V = \sum_{i \in \text{Accept}} v_i$ . Une preuve de son optimalité se trouve en Annexe B.2.

L'algorithme est présenté à la figure 9.2. L'idée est la suivante : nous ajoutons les requêtes une par une, en commençant par celles de plus grandes valeurs. La variable  $k$  correspond à la valeur de type à laquelle une requête doit être supérieure pour pouvoir être acceptée. Plus précisément, quand toutes les requêtes disponibles de type  $1, 2, \dots, j$  ont été allouées (c'est-à-dire que le nombre de requêtes de types  $1, 2, \dots, j$  acceptées a atteint  $J_1 + \dots + J_j$ ) alors  $k = j$ . Pour chaque nouvelle requête, on vérifie si son type est plus grand que  $k$ . Si c'est le cas, elle est acceptée et nous mettons à jour la valeur de  $k$ . Sinon, elle est refusée.

La sortie de l'algorithme est un vecteur "accept" de  $p$  éléments booléens. Donc l'ensemble **Accept** contient les requêtes dont la valeur correspondante dans le vecteur "accept" vaut 1. On peut alors en déduire, par simple comptage, les valeurs des  $D_t(z, o)$  pour les différents opérateurs.

### 9.2.2 Obtention du vecteur $J$ minimal

Dans chaque spot  $s$  nous définissons

$$J_t^s = \sum_{o \in O} \sum_{z \in Z(s)} D_t^{min}(z, o), \text{ et } J^s = (J_1^s, \dots, J_N^s)$$

```

k := min{i such that Ji > 0} - 1
for i := 1 to p
  begin
    if ti > k then
      begin
        accept[i] := 1
        for l := k + 1 to N
          if  $\sum_{m \leq i \text{ and } t_m \leq l} \text{accept}[m] = \sum_{n \leq l} J_n$ 
            then k := l;
          end
        else accept[i] := 0;
      end
    end;
  end;

```

FIG. 9.2 – Un algorithme d'acceptation des requêtes pour maximiser la somme de leur valeur.

$J^s$  correspond au nombre minimal de porteuses de chaque type nécessaire dans le spot  $s$  pour satisfaire les demandes minimales des utilisateurs.

On pose alors  $J^* = \max(\{J^s\}_s \text{ spot})$ .<sup>2</sup> Si  $J$  satisfait les contraintes de capacité du système alors toutes les demandes minimales peuvent être satisfaites. Sinon, le système est non faisable.

D'un point de vue pratique, cette étape peut se faire en utilisant l'algorithme décrit à la section 9.2.1 et en fixant la valeur 1 aux  $D_t^{\min}(z, o)$  premières requêtes de chaque opérateur dans chaque zone et pour chaque type et 0 aux autres.

### 9.2.3 Etapes de résolutions de la solution générale

Afin de déterminer le vecteur global  $J = (J_1, \dots, J_N) \in \mathbf{N}^N$ , nous considérons maintenant la procédure globale d'allocation pour une couleur donnée.

Pour chaque vecteur  $J$  admissible (c'est-à-dire chaque vecteur  $J \geq J^*$ ), nous trouvons dans chaque spot l'allocation optimale et équitable des porteuses pour chaque opérateur en utilisant l'algorithme de la Section 9.2.1. Nous évaluons alors la qualité de l'allocation

<sup>2</sup>Nous pouvons considérer un ordre partiel parmi les allocations : on dit qu'une allocation  $J'$  est plus grande qu'une autre allocation  $J$  (et nous écrivons  $J' \succeq J$ ) si et seulement si

$$\begin{array}{rcl}
J'_1 & \geq & J_1, \\
J'_2 + J'_1 & \geq & J_2 + J_1, \\
\cdots & & \cdots \\
\sum_{t=1}^N J'_t & \geq & \sum_{t=1}^N J_t.
\end{array}$$

Nous avons la propriété suivante (la preuve se trouve dans B.1) : étant donnée deux allocations  $J$  et  $J'$ , il y a une unique allocation  $J^* = \max(J, J')$  telle que  $J^* \succeq J$ ,  $J^* \succeq J'$ , et pour toute allocation  $J''$  telle que  $J'' \succeq J$  et  $J'' \succeq J'$ , nous avons  $J'' \succeq J^*$ .

obtenue avec  $J$  en utilisant notre critère global d'équité. Le vecteur  $J$  ayant la meilleure valeur est sélectionné.

L'obtention de l'allocation optimale des porteuses pour chaque opérateur se fait en une ou deux étapes, selon la valeur de  $\alpha$ . Les deux paragraphes suivants détaillent ces étapes.

### 9.2.3.1 Maximization des bénéficiaires du surplus (pour $\alpha = 1$ ).

L'équation 9.2 montre que, pour  $\alpha \geq 1$ , si il existe  $t, z, o$  tels que  $D_t(z, o) = D_t^{min}(z, o)$ , alors le critère d'optimisation donné par 9.2 devient infini. C'est pourquoi nous souhaitons minimiser le degré de nullité du pôle  $\prod_{z,t,o} D_t(z, o) - D_t^{min}(z, o)$  en maximisant le nombre d'opérateurs à qui l'on peut allouer strictement davantage que leur demandes minimales.

Pour cela, nous utilisons l'algorithme de la section 9.2.1, pour chaque opérateur, dans chaque zone et chaque type de porteuse, nous fixons les valeurs des requêtes à :

- 2 pour  $D_t^{min}(z, o)$  requêtes
- 1 pour une requête
- 0 pour les  $D_t^{max}(z, o) - D_t^{min}(z, o) - 1$  autres requêtes

L'analyse de l'ensemble **Accept** alors obtenu permet de trouver les opérateurs à qui l'on va pouvoir accorder plus de porteuses que le minimum qu'ils requierent (il s'agit des opérateurs dont l'une des requêtes de valeur 1 a été acceptée). Nous restreignons alors notre problème uniquement à ces opérateurs.

### 9.2.3.2 Redistribution équitable des valeurs.

Nous utilisons l'algorithme de la Section 9.2.1 dans lequel l'on choisit judicieusement les variables "valeurs" en fonction du paramètre d'équité choisi.

Pour chaque opérateur (dans chaque zone et pour chaque type), on fixe les valeurs des requêtes comme suit :

- On fixe la valeur de  $D_t^{min}(z, o)$  requêtes à  $+\infty$  (dans la pratique, n'importe quelle borne supérieure  $\Lambda$  suffit). On garantit ainsi que le système solution respectera les demandes minimales des opérateurs.
- Si  $\alpha \geq 1$ , nous fixons la valeur d'une requête supplémentaire à  $\Lambda$  afin de s'assurer que leur allocation sera strictement supérieure à leur demande minimale.
- Nous fixons alors la valeur de chaque demande non-allouée  $j$  ( $D_t^{min} + 2 \leq j \leq D_t^{max}$  si  $\alpha \geq 1$ ,  $D_t^{min} + 1 \leq j \leq D_t^{max}$  sinon) à :

$$\begin{cases} \log \left( \frac{j - D_t^{min}(s, o)}{j - 1 - D_t^{min}(s, o)} \right) & \text{si } \alpha = 1^3 \\ C_t(s, o)^{1-\alpha} [(j - D_t^{min}(s, o))^{1-\alpha} - (j - 1 - D_t^{min}(s, o))^{1-\alpha}] & \text{sinon.} \end{cases}$$

<sup>3</sup>On peut noter que la valeur  $v$  des requêtes est alors indépendante de  $C$ . Ceci est dû au fait qu'il n'y a pas "d'équité proportionnelle pondérée" avec le concept d'équité que nous utilisons ici, cf Section 9.1.4.

## 9.3 Expérimentations

### 9.3.1 Le réseau considéré

On considère un exemple de réseau satellitaire contenant 32 spots et 4 couleurs. Le nombre de spots par couleur vaut donc 8.

#### 9.3.1.1 Types et utilité des porteuses

Chaque spot est couvert par un canal MF-TDMA (dont la bande passante est de 36 MHz). Il peut être utilisé avec 5 types de porteuses, de bande passante 6000, 3000, 1500, 750, et 187.5 kHz. L'espacement est de 50%, ce qui signifie que la bande passante peut accepter un maximum de 4 porteuses du premier type, 8 du second, 16 du troisième, 32 du quatrième et 128 de la dernière, si toute la bande passante est utilisée pour un seul type de porteuse. Chaque spot est divisé en 3 zones, et les valeurs  $C_t$  d'un type de porteuse dans l'un des spots est donné par :

$C_t(z, o)$	Zone z=0	Zone z=1	Zone z=2
Type t=1	4920.000000	4373.328613	3280.000000
Type t=2	2460.000000	2186.664307	1640.000000
Type t=3	1230.000000	1093.332153	820.000000
Type t=4	615.000000	546.666077	410.000000
Type t=5	153.750000	136.666519	102.500000

Ces valeurs peuvent représenter plusieurs estimations des types de porteuses par les opérateurs. En particulier, l'évanouissement du signal radio peut demander une augmentation du codage de Viterbi de 3/4 à 1/2 ce qui change le débit final et le service fourni par l'opérateur.

#### 9.3.1.2 Génération des demandes des opérateurs

Nous supposons qu'il y a 40 opérateurs et que chacun peut utiliser au plus 3 spots. Il peut y avoir au plus deux opérateurs par spot et chaque opérateur dans un spot peut utiliser 1, 2 ou 3 zones. Une même zone peut être allouée à plus d'un opérateur. Pour générer une telle demande, nous choisissons 3 spots différents parmi les 32 pour chaque opérateur. Si l'un de ses spots est déjà utilisé par deux opérateurs, nous annulons simplement la requête du nouvel opérateur. A l'intérieur de chaque spot sélectionné, l'opérateur choisit au hasard s'il utilisera chaque type de porteuse avec une probabilité 1/2. Si il n'utilise finalement aucune porteuse, sa demande sera retirée pour ce spot. Nous fixons le nombre maximum de porteuses à la valeur maximale possible (c'est-à-dire de façon à saturer le spot si l'opérateur était seul). Le nombre minimum de porteuses est choisi aléatoirement de façon à occuper entre 20% et 50% de la bande passante totale divisée par le nombre de porteuses utilisées par cet opérateur. Nous supposons dans l'exemple numérique qu'une demande d'un opérateur  $o$  dans une zone  $z$  est donné par trois valeurs : un type de porteuse  $t$  ainsi que le nombre minimum  $d_{\min}$  et maximum  $d_{\max}$  de porteuses désirées. Ces valeurs sont supposées avoir l'interprétation suivante : l'opérateur demande exactement  $D_t^{\min}(z, o)$  porteuses de type  $t$  et espère avoir au plus  $d_{\max} - d_{\min}$  porteuses de type  $\min(t + 1, N)$ . En d'autres termes,  $D_{\min(t+1, N)}^{\min} = 0$ ,  $D_{\min(t+1, N)}^{\max} = d_{\max} - d_{\min}$ , et pour tout  $s > \min(t + 1, N)$ ,  $D_s^{\min} = D_s^{\max} = 0$ . Par exemple,

si un opérateur demande un minimum de 10 porteuses de type 1 et un maximum de 20 porteuses, il recevra les porteuses de type 1 demandées (au nombre de 10) et il recevra au plus 10 porteuses au maximum de type 2 (ces porteuses seront éventuellement de type inférieur compte tenu de la contrainte ISCC).

### 9.3.2 Résultats numériques

Les résultats sont donnés sur les figures 9.3 et 9.4 pour différentes valeurs de  $\alpha$ . Pour chaque valeur du critère d'équité, nous traçons la bande passante allouée à chaque opérateur dans les différents spots par types de porteuses croissant<sup>4</sup>. Les allocations de chaque opérateur sont représentées dans une couleur spécifique, en couleur pleine pour les allocations minimales et en hachurée pour le surplus octroyé. Comme expliqué plus tôt (cf. sous-section 9.3.1.2), les allocations supplémentaires seront de type immédiatement inférieur aux demandes minimales.

Dans chaque spot et pour chaque type, la hauteur de la barre est proportionnelle à la bande passante. Ainsi, une porteuse supplémentaire de type 2 augmentera davantage la taille de la barre qu'une porteuse additionnelle de type 3.

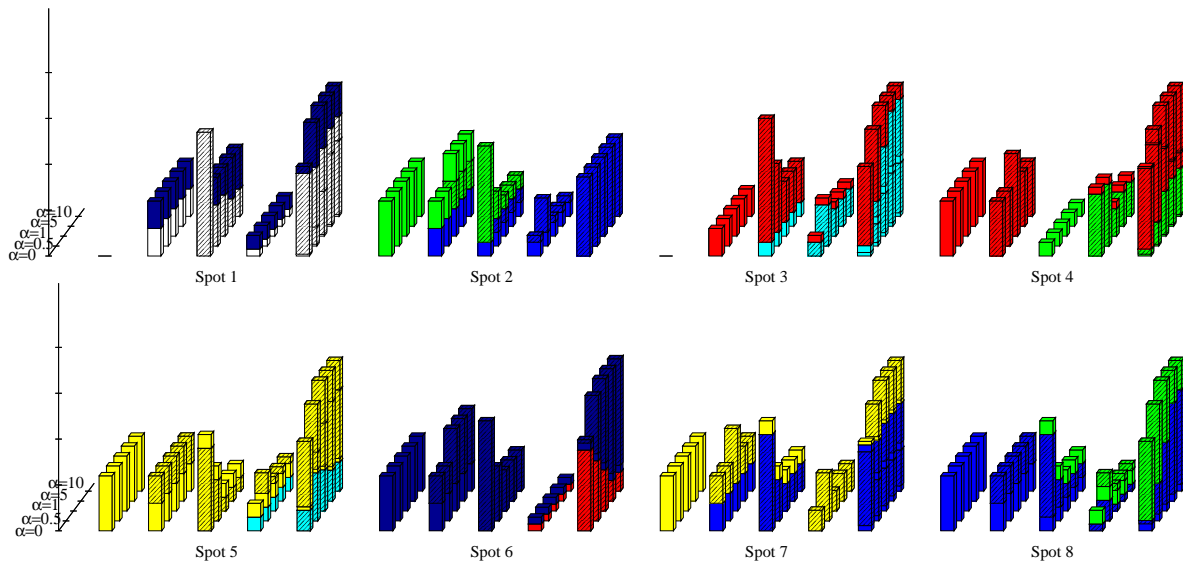


FIG. 9.3 – Allocation équitable avec ISCC

Les expériences montrent clairement que, lorsque  $\alpha$  augmente, les porteuses de plus grands types sont favorisées au détriment de celles de plus grande capacité. Aussi lorsque  $\alpha$  augmente on remarque (figure 9.5), l'efficacité est réduite ce qui est le prix logique d'une plus grande équité. Une remarque intéressante peut également être tirée de 9.5. Nous remarquons que la capacité du système perdue avec les contraintes ISCC reste inférieure à 10%. Cela nous conforte dans l'idée que le coût de ISCC reste raisonnable.

<sup>4</sup>En fait nous représentons la barre au niveau du type correspondant à celui qui sera facturé à l'opérateur. Comme nous l'avons précédemment vu, il se peut, à cause de l'ISCC qu'une ou plusieurs porteuses allouée soit en fait de type inférieur, c'est-à-dire de capacité supérieure à celle demandée.

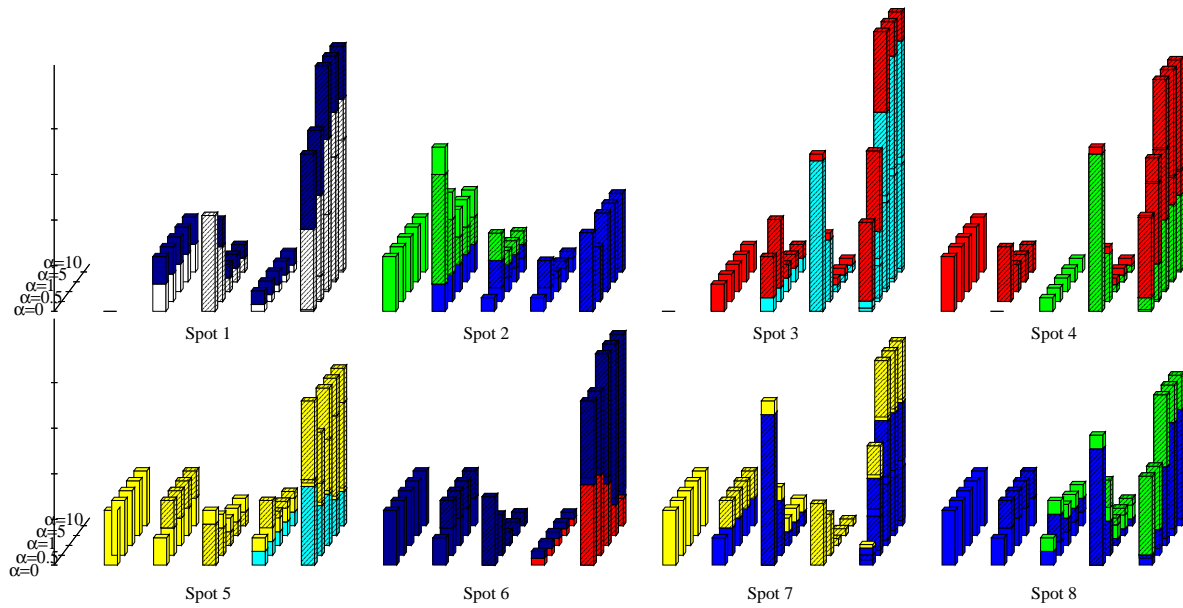


FIG. 9.4 – Allocation équitale sans ISCC

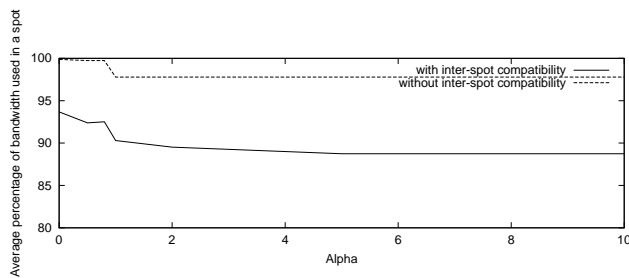


FIG. 9.5 – Efficacité de l'allocation de la bande passante avec et sans l'ISCC

## 9.4 Conclusion

Nous avons dans ce chapitre adapté les notions d'équité au problème d'allocation de ressources dans les réseaux satellitaires. Une difficulté majeure liée à ces réseaux est l'introduction de contraintes d'intégrité. En effet, on ne peut allouer une quantité quelconque de bande passante à un opérateur mais l'on doit déterminer un nombre entier de porteuses d'un certain type, c'est-à-dire d'une certaine bande passante prédéfinie.

L'approche est en cela novatrice que les méthodes jusqu'ici présentées ne prenaient en compte qu'un ou deux types de porteuses. De plus, la méthode peut être utilisée sur des systèmes relativement grands sans que le temps de calcul devienne prohibitif. Enfin, cette étude est, à notre connaissance, la première introduisant des notions d'équité. Le critère général d'équité utilisé ici est une variante de l'équité pondérée générale introduite par [85].

Nous nous intéressons dans le chapitre suivant, à l'allocation des slots aux opérateurs. Je présente rapidement deux algorithmes que nous avons développés pour résoudre le problème, car ils complètent l'étude du système et font suite à l'étude effectuée dans ce chapitre. Cependant, ils sont fortement résumés car non directement liés à des notions d'équité.

# Chapitre 10

## Deuxième partie de l'allocation : l'allocation des slots.

Nous proposons deux méthodes pour résoudre le problème d'allocation des slots. La première est basé sur un recuit simulé et la seconde sur la programmation linéaire. Je les mentionne ici afin de compléter la solution proposée du problème d'allocation de ressource dans le cas des réseaux satellitaires. Une version détaillée des deux solutions se trouve dans [112] pour l'approche par recuit simulé et [4] pour la programmation linéaire.

Après une courte introduction sur le modèle d'interférence considéré je présente l'idée générale utilisée dans les deux approches.

### 10.1 Le modèle d'interférence.

La bande de fréquence est potentiellement réutilisée d'un spot à l'autre. Cette réutilisation n'est cependant possible que si le niveau d'interférence généré par l'ensemble des spots utilisant simultanément la ressource radio est acceptable.

Chaque zone génère une certaine interférence sur tous les autres spots  $I(s, z)$  (si  $s$  appartient au spot  $z$  alors  $I(s, z) = 0$ ). Chaque zone a également un gain  $G(z)$ . Si  $spot(z)$  est le spot contenant  $z$ , alors pour tout instant  $t$  et fréquence  $f$  la zone  $z$  peut être active si

$$\frac{G(z)}{\sum_{z' \text{ active en } t, f} I(spot(z), z')} \geq \sigma \quad (10.1)$$

Dans les tests effectués par la suite, la valeur du gain en décibels est prise aléatoirement entre deux bornes. Le seuil d'interférence  $\sigma$  est une donnée du problème. Si  $\sigma$  est très faible (par exemple 10dB, ce qui n'est pas très réaliste), alors tous les spots peuvent être actifs en toute paire  $(t, f)$ . Plus le seuil est élevé, moins de spots peuvent être actifs simultanément et dans la même bande de fréquence.



## 10.2 Le recuit simulé

Nous décrivons ici une solution heuristique basée sur le recuit simulé. Après une brève introduction de l'heuristique (sous-section 10.2.1), nous montrons comment l'adapter à notre problème (sous-section 10.2.2). Nous fournissons ensuite quelques idées simples permettant d'améliorer la performance du recuit simulé (sous-section 10.2.3). Des expériences numériques sont proposées pour tester nos améliorations (sous-section 10.2.4).

Rappelons tout d'abord que nous nous intéressons à un système MF-TDMA. Dans ce système, un slot ne peut être alloué simultanément à plus d'une zone d'une cellule. Les cellules interfèrent entre elles, et un slot peut être alloué à un opérateur dans une zone donnée si l'interférence qu'il subit est inférieure à un certain seuil. L'allocation de slot est statique mais peut être changée une fois par heure (à cause des changements de demandes, d'une part et des changements de conditions de transmission d'autre part). Notre but est de maximiser débit d'information utile, en prenant en compte le taux de codage d'un slot, qui peut varier d'une zone à l'autre. En castant le problème en une coloration de graphe on peut montrer qu'il est NP-complet de maximiser le débit utile. C'est pourquoi nous proposons donc une approche heuristique basée sur le recuit simulé [81, 64] et analysons ses performances.

Comme notre nouvelle allocation doit être générée chaque heure, les solutions sont nécessaires en moins de quelques minutes. Nous avons travaillé avec huit spots, trois zones par spot, quatre couleurs et six types de porteuses. Le nombre total de porteuses des différents types par spot était généré par hasard, et le nombre total de porteuses était autour de 100 par spot.

### 10.2.1 Présentation générale du recuit simulé

L'article [64] présente une belle introduction des méthodes heuristiques et en particulier du recuit simulé. L'inconvénient d'algorithmes basés sur des heuristiques est qu'ils sont plutôt spécifiques d'un problème. Ainsi il n'y a ainsi aucune garantie qu'une heuristique efficace pour un système le serait pour un autre.

Les problèmes comme celui du voyageur de commerce<sup>1</sup> sont NP-complets, c'est-à-dire qu'il n'existe pas d'algorithme pouvant les résoudre en un temps polynomial. Le but des heuristiques est alors de trouver une (bonne) approximation de la solution en un temps polynomial. Il existe deux stratégies principales d'heuristiques : le "divide and conquer" (dans lequel on divise successivement le problème en sous-problèmes) et les méthodes itératives. Le problème d'une méthode dans laquelle on améliorerait progressivement la solution est que l'algorithme peut alors converger vers un optimum local et non global. Une possibilité est alors d'utiliser l'algorithme à partir de plusieurs configurations initiales et de ne retenir que le meilleur résultat. Le recuit simulé propose une autre approche.

Le recuit simulé est né dans les années 50 des travaux de Metropolis. L'idée est née d'une analogie avec la physique, plus précisément de l'étude de la façon dont les atomes

---

<sup>1</sup>Le problème du voyageur de commerce peut être formulé comme ceci. Étant donné une liste de  $N$  villes et un moyen de calculer le coût de déplacement entre 2 villes, le problème est de décrire l'itinéraire du voyageur, de sorte qu'il passe par toutes les villes et retourne finalement à son point de départ, tout en minimisant le coût total.

s'agrègent (en moyenne) lors des très faibles températures. Par exemple, pour obtenir un cristal parfait, il faut abaisser la température très progressivement et rester longtemps dans une température proche du point de cristallisation. Si au contraire le corps est refroidi brutalement, le cristal n'est pas parfait, c'est-à-dire que les atomes ne sont pas dans leur état le plus stable. C'est l'idée du recuit simulé.

Pour éviter de converger vers un minimum local, on accepte, à chaque itération des états qui ne sont pas plus proche de l'optimum que les états précédents avec une certaine probabilité. La fonction de probabilité utilisée représente la "température" du système. A mesure que le processus se déroule, on abaisse la température, c'est-à-dire que l'on diminue cette probabilité. On comprends alors que le choix de cette fonction, c'est-à-dire le choix du schéma de refroidissement est crucial pour assurer à la fois une convergence vers le minimum absolu et de façon aussi rapide que possible.

Hajek énonce ([44]) des conditions que doivent remplir les schémas de refroidissement pour assurer une convergence vers le minimum absolu.

### 10.2.2 Recuit simulé efficace

A chaque étape du recuit simulé, un spot, une zone, une porteuse et un slot sont choisis aléatoirement. Nous appelons ce choix un candidat à l'allocation. Si le slot n'est pas encore alloué alors nous n'acceptons le candidat si les conditions précédente d'interférences sont satisfaites. Si le slot choisi a déjà été alloué à une autre zone alors on peut :

1. maintenir l'allocation existante,
2. utiliser le nouveau candidat à la place de celui existant, ou
3. abandonner l'allocation existante sans accepter l'autre allocation de ce slot.

Pour décider que faire à chaque étape, nous introduisons une mesure de performance  $\mathcal{G}$  qui est la somme du nombre de slots alloué pondérée par le taux de codage. Elle correspond au débit utile du système. C'est notre fonction à maximiser dans le recuit simulé. La probabilité avec laquelle on maintient l'allocation courante dans le cas où le nouveau candidat est acceptable (en terme de quantité d'interférences) s'écrit alors :

$$P(T) = \exp\left(-\frac{|\Delta\mathcal{G}|}{T}\right).$$

Nous avons implémenté et présenté dans [112] une méthode de recuit simulé hiérarchique, le second critère étant la minimisation de l'interférence créée.

L'algorithme s'arrête si l'une des conditions suivantes est vérifiée :

- Une limite de temps est dépassée,
- Une limite sur le nombre d'itérations est dépassée,
- Le gain  $\mathcal{G}$  est resté inchangé pendant un certain nombre d'itérations consécutives.

Nous avons testé à la fois des taux linéaires et logarithmiques pour le refroidissement de  $T$ . On peut noter que le refroidissement linéaire est le plus souvent utilisé, car il est beaucoup plus rapide que le logarithmique. Mais la convergence asymptotique à un optimum global n'est plus garantie ([44]).

### 10.2.3 Améliorations de l'algorithme

Nous améliorons d'avantage les performances du système par des actions non directement liées au recuit simulé décrit à la section précédente. Il s'agit en fait de méthodes élémentaires mais permettant d'améliorer grandement les performance par rapport à une méthode brutale, sans pour autant augmenter la complexité de l'algorithme ou de l'implémentation.

Ainsi, dans la moitié des spots d'une couleur donnée, nous avons placé les porteuses de plus grand type (c'est-à-dire de plus grande bande passante) dans le sens croissant de la bande passante et dans l'autre moitié dans le sens décroissant. Ceci nous permet d'avoir moins d'interférences entre les porteuses de grande bande passante.

Une seconde amélioration a été d'effectuer le recuit simulé en deux étapes plutôt qu'une : à chaque phase est allouée la moitié du temps limite dont nous disposions initialement pour la simulation totale. A la première étape, le recuit simulé se restreint à l'allocation des slots pour les porteuses ayant les plus grandes bandes passantes seulement. Nous effectuons alors un second recuit simulé pour allouer le reste des porteuses.

### 10.2.4 Résultats

Nous avons réalisé nos expériences pendant 1min, 4min et 10min sur un ordinateur DELL INSPIRON 2500 avec un processeur Pentium III 1GHz et 256 MB de SDRAM. L'algorithme a été programmé dans un programme Java d'environ 1500 lignes.

Nous avons tout d'abord fait tourner un recuit simulé très simple, sans contraintes sur les interférences de façon à obtenir une borne supérieure des performances qui peuvent être atteintes. Nous faisons alors différentes simulations avec le recuit simulé, chaque fois avec une autre limite d'interférence permise et traçons le débit efficace obtenu, normalisé par celui obtenu avec la borne supérieure. Les résultats sont présentés sur la figure 10.1. La meilleure performance est obtenue avec la version optimisée. Nous la comparons à quatre autres schémas. Trois d'entre eux n'ont qu'un seul recuit simulé (consistant à l'optimisation de toutes les porteuses en une seule étape). Certains utilisent le refroidissement logarithmique et d'autres linéaires.

La figure 10.1 montre que notre recuit simulé optimisé améliore les solutions d'environ 50 % pour les faibles seuils d'interférences considérées (c'est-à-dire  $-30$  dB) et 600 % pour les forts seuils d'interférences, ( $+50$  dB).

#### 10.2.4.1 Conclusions

Nous avons proposé une approche de recuit simulé pour le problème d'allocation dans le lien montant des slots dans un système géostationnaire multi-faisceaux. Nous avons proposé une approche d'optimisation hiérarchique qui essaye non seulement de maximiser le débit mais préfère également les allocations qui génèrent le minimum d'interférences. Plusieurs améliorations ont été implémentées dans une approche de recuit simulé en deux étapes que nous avons comparée numériquement à d'autres approches. Nous avons montré qu'un gain substantiel est obtenu.

En fait, le recuit simulé est une amélioration des travaux entrepris par Alcatel. Certes, le gain est substantiel par rapport à une affectation aléatoire notamment, mais nous avons

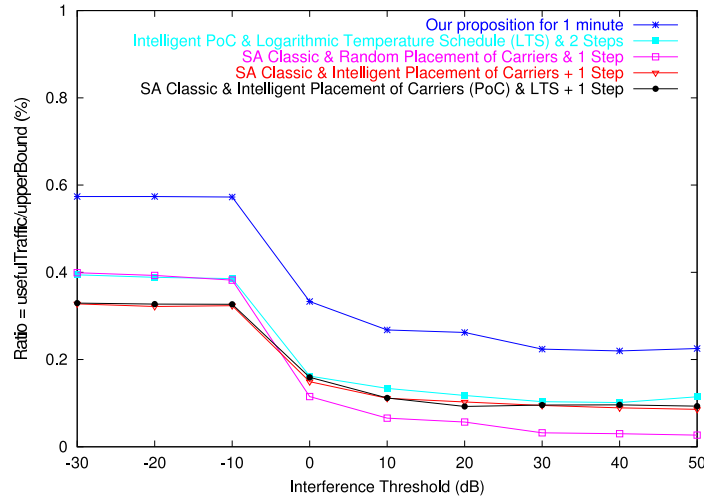


FIG. 10.1 – Comparaison pour une minute

développé une seconde technique, bien plus performante car mieux adaptée au problème. Je l'évoque dans la sous-section suivante. Je ne présente pas de comparaison des deux méthodes, les différences de performances étant trop importantes, notamment en ce qui concerne la taille du problème que l'on peut considérer (en terme de nombre de spots et de types de porteuses).

## 10.3 Résolution par programmation linéaire

Nous avons proposé une deuxième méthode de résolution d'allocation des slots ([4]). Elle est basée sur une programmation linéaire. Je la mentionne ici pour compléter l'étude de l'allocation de ressources dans les réseaux satellitaires et en explique le principe mais pas l'implémentation complète.

Nous présentons dans une première sous-section les notations utilisées ici. Notre algorithme se compose de deux phases. Dans la première un ensemble de *familles* est généré. La seconde phase de l'algorithme consiste à allouer ces familles dans le plan de fréquence. Ces deux étapes font l'objet de sous-sections indépendantes. Enfin, nous terminons ce chapitre par quelques résultats numériques et une brève conclusion.

### 10.3.1 Présentation du problème

#### 10.3.1.1 Notations

On notera  $N$  le nombre de spots de chaque couleur et  $B$  la taille de la bande passante. En pratique, on s'intéresse à  $N \leq 32$ . Chaque zone de chaque spot est notée sous la forme : *spot.zone*.

Les terminaux ont des capacités variables d'émission. Ces capacités sont indiquées par la notion de type. Un type correspond à la bande de fréquence qu'utilise un terminal. L'intervalle de temps effectif d'un time-slot dépend du type utilisé, de sorte que le volume de données transmises lors d'un time-slot est une constante du problème. Ainsi, pour tout type

$k$ , le produit de la largeur de bande de la porteuse  $t_k^b$  multiplié par la durée de transmission  $t_k^t$  est constant pour tous les types et est noté  $\Delta$ .

$$\forall k = 1, \dots, \tau, \Delta = t_k^b t_k^t.$$

Chaque spot a une demande spécifique à satisfaire, exprimée en nombre de time-slots (appelés nombre de porteuses dans la première partie) par types de terminaux et par zone.

$$\forall z_i : \forall k, d_{z_i}^{t_k} := \text{demande de } z_i \text{ en type } t_k$$

### 10.3.1.2 Modèle d'interférence

En fait, plus une antenne est directive, plus petits sont les lobes secondaires, autrement dit plus faible est la puissance du signal interférent. N'ayant pas à notre portée de simulation du niveau du signal interférent dans toutes les zones géographiques, nous adoptons le raisonnement suivant : les zones voisines sont celles qui génèrent le plus d'interférence l'une sur l'autre ; toutefois, les zones lointaines génèrent également des interférences, d'intensité plus faible. Nous distinguons ainsi l'interférence de voisinage, générée sur les (six) spots les plus proches, notée  $S_1(z)$  et l'interférence globale, générée par l'ensemble des spots notée  $S_2(z)$ . Enfin, la valeur finale de l'interférence générée sur la zone  $z$ , est prise comme étant un barycentre à poids fixe des deux valeurs :

$$S_1(z) = \sum_{z' \text{ active dans spot voisin}} I(\text{spot}(z), z'), \quad S_2(z) = \sum_{z' \text{ active}} I(\text{spot}(z), z')$$

$$I(z) = \gamma S_1(z) + (1 - \gamma) S_2(z).$$

Le critère d'allocation (10.1) devient alors

$$\frac{G(z)}{I(z)} \geq \sigma. \quad (10.2)$$

## 10.3.2 Utilisation des familles

### 10.3.2.1 Qu'est-ce qu'une famille ?

Pour bien comprendre la modélisation, considérons un système simple comme représenté sur la figure 10.2. Il se compose d'un seul type de terminal, comme en GSM.

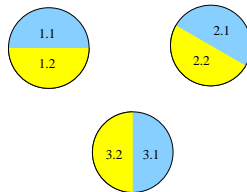


FIG. 10.2 – Petit exemple sur 3 spots.

Le canal radio se découpe en "time-slots" où un certain nombre de zones peuvent émettre simultanément. La première question est de savoir quelles zones peuvent alors

émettre (on qualifiera de telles zones d'*allumées*) dans un time-slot donné? Dans notre exemple, chaque spot est découpé en deux zones qui ont une interférence spécifique sur les autres spots. Par exemple, on peut considérer le tableau 10.1 qui précise, pour chaque zone, son gain ainsi que les interférences avec elle qu'elle génère sur les autres spots.

Zone	Gain	I(.,Spot 1)	I(.,Spot 2)	I(.,Spot 3)
1.1	4	-	5	3
1.2	6	-	5	7
2.1	3	4	-	2
2.2	8	7	-	10
3.1	5	3	7	-
3.2	5	7	3	-

TAB. 10.1 – Tableau des gains et interférences des zones de l'exemple.

Concentrons-nous maintenant sur la zone 1.1. Dans quelle mesure cette zone peut-elle émettre? Cela dépend des états des zones qui interfèrent. En effet, suivant que les zones 2.1, 2.2, 3.1 et 3.2 sont allumées ou non, son rapport C/I va être différent. La table 10.2 résume les 9 cas possibles. On peut effectuer les calculs de façon similaire pour les 5 autres zones ([4]).

C/I (Zone 1.1)	Zone 2.1 allumée	Zone 2.2 allumée	Spot 2 éteint
Zone 3.1 allumée	0.57	0.40	1.33
Zone 3.2 allumée	0.36	0.29	0.57
Spot 3 éteint	1.00	0.57	-

TAB. 10.2 – Détails des interférences possibles pour la zone 1.1.

En fixant le seuil d'admissibilité à 0.30, on peut voir que toutes les combinaisons possibles de spots avec deux zones allumées dans deux spots différents sont possibles, et que trois combinaisons avec trois spots actifs sont possibles. Nous montrons ces dernières sur la figure 10.3.

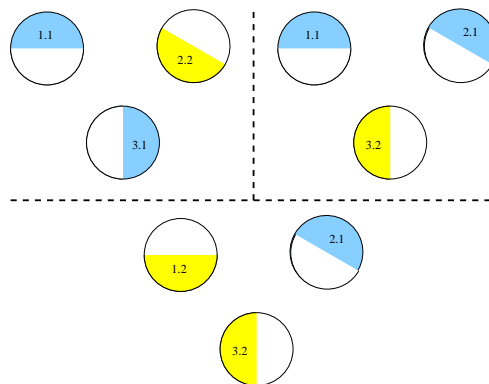


FIG. 10.3 – 3 configurations d'émission possibles pour un seuil à 0.30.

Nous appellerons dans la suite *famille* une configuration admissible.

Pour une meilleure résolution du problème, d'autres notions pourront être utiles. Par exemple, notons que les familles 2 et 3 (de la figure 10.3) diffèrent seulement par le spot 1. En fait, il est correct de dire qu'il est possible d'allumer simultanément *tout* le spot 1 avec les zones 2.1 et 3.2. Nous allons ainsi créer des configurations qui en résument plusieurs autres en laissant de la flexibilité sur la zone que l'on allume au sein d'un spot. Nous utiliserons également quelquefois des familles qui ont "trop" de zones allumées dans le sens où certaines resteront inutilisées pendant l'utilisation effective du système.

**10.3.2.2 Définition des super-zones.**

Le nombre de familles possible est en pratique très grand pour que l'on puisse les décrire de manière exhaustive. Par conséquent nous avons introduit la notion de super-zone au sein d'un spot. Une super-zone contient plusieurs zones d'un même spot. Si une famille contient une super-zone, cela indique que chacune des zones qu'elle contient peut émettre avec les autres zones en respectant les conditions d'interférences. Les paramètres d'une super-zone  $\zeta$  regroupant deux zones  $z$  et  $z'$  d'un même spot sont les suivants :

$$\begin{cases} G(\zeta) &= \min(G(z), G(z')) & \text{et} \\ I(s, \zeta) &= \max(I(s, z), I(s, z')) & \text{pour tout spot } s. \end{cases}$$

Notre méthode de résolution consiste donc, dans un premier temps, à générer des familles  $F_i$  constituées de zones ou super-zones de spots distincts.

**10.3.2.3 Familles de spots génériques**

Vu que l'interférence de voisinage est décisive dans le calcul de  $I(z)$  dans l'équation (10.2), nous répartissons géographiquement les spots qui seront éteints en  $(t, f)$ . On peut noter qu'à part les spots à la limite de la couverture du satellite, tout spot est voisin de 6 autres spots. Notre but est donc de répartir les spots éteints de façon à éviter les situations où plusieurs spots voisins sont éteints.

Si  $\sigma$  est plutôt petit, il suffit de s'assurer que tout spot allumé ait exactement un spot voisin éteint. Ce schéma, représenté dans la figure 10.4, est désigné sous le code 6/7 puisque 6 spots sur 7 sont allumés. Il y a 7 familles 6/7 distinctes.

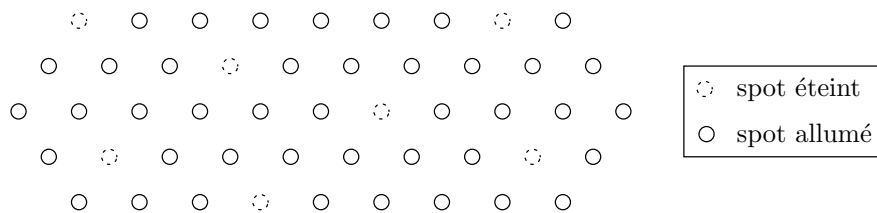


FIG. 10.4 – Illustration d'une famille 6/7

Quand le seuil d'interférence  $\sigma$  est plutôt élevé, il arrive que l'on ne puisse pas satisfaire le critère (10.2) pour toutes les zones allumées dans le schéma 6/7. Nous introduisons donc le schéma 5/7 dans lequel 5 spots sur 7 sont allumés. Le schéma 5/7 est représenté dans la

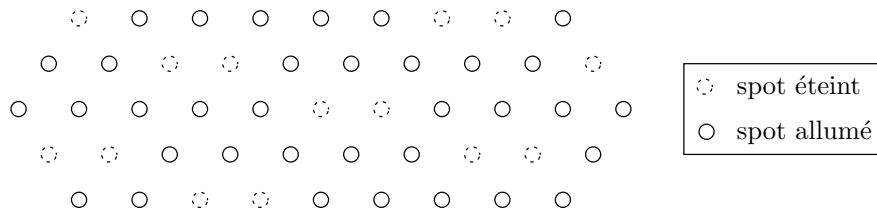


FIG. 10.5 – Illustration d’une famille 5/7

figure 10.5. Les spots allumés forment une famille 5/7 et il y a, au total,  $7 \times 3 = 21$  familles 5/7 distinctes.

Il est également possible de générer des familles 4/7 (il y a en a  $7 \times 5 = 35$  au total), ce qui serait utile dans le cas où le seuil d’interférence  $\sigma$  est très élevé.

### 10.3.2.4 Statut des spots

Dans la sous-section précédente, nous avons distingué entre un spot allumé ou actif et un spot éteint. Or, tout spot comporte un certain nombre de zones géographiques (typiquement 2 ou 3 zones). Il existe donc plusieurs possibilités pour chaque spot allumé. Comme le niveau de puissance du signal utile dépend de la zone, ainsi que les niveaux d’interférences générées par les zones des autres spots, il se peut qu’une zone d’un spot ne vérifie pas le critère d’allocation (10.2), alors qu’une autre zone du même spot le vérifie. Nous introduisons la notion de *statut* d’un spot qui prend valeur dans l’intervalle  $[0, 2^{nbZones(s)} - 1]$ , où  $nbZones(s)$  désigne le nombre de zones dans le spot  $s$ . Pour un spot contenant 3 zones, son statut peut prendre 7 valeurs selon les zones allumées simultanément.

L’introduction du statut d’un spot donne plus de liberté et de souplesse lors de l’allocation de ressources. Ainsi, si un spot ayant 3 zones a un statut 7, ceci veut dire que nous pouvons allouer les ressources à n’importe laquelle des 3 zones.

Une autre conséquence de cette notion de statut de spot est relative au calcul du critère d’allocation (10.2). Au lieu d’avoir à vérifier ce critère zone par zone, nous pouvons le faire au niveau des spots, ce qui diminue le nombre de fois où ce calcul doit être fait.

Pratiquement, une famille  $F_i$  est représentée par un vecteur, dont la taille est égale au nombre de spots, et dont les éléments sont les statuts des spots (c’est-à-dire un chiffre de 0 à  $2^{nbZones(s)} - 1$ ). Comme nous voulons maximiser le nombre de zones allumées, nous commençons par générer les 7 familles 6/7 dans lesquelles tout spot allumé  $s$  a le statut  $2^{nbZones(s)} - 1$ , les spots éteints ayant le statut 0. Nous testons successivement la validité de ces familles et les séparons en deux ensembles : d’un côté les familles 6/7 valides et de l’autre côté les familles 6/7 non valides. Nous pouvons faire de même avec les familles 5/7, 4/7, etc., selon le seuil d’interférence  $\sigma$  considéré. Pour qu’une famille non valide devienne valide, faut éteindre certaines de ses zones allumées. Nous proposons une heuristique pour choisir les zones à éteindre.



### 10.3.3 Choix et placement des familles.

Notre algorithme génère donc dans un premier temps un ensemble de familles, c'est-à-dire de zones qui peuvent émettre dans une même aire temps/fréquence. Une seconde phase de l'algorithme consiste à placer ces familles dans le plan de fréquence. Ceci est effectué par un programme linéaire.

#### 10.3.3.1 Résolution des contraintes de placement sur le canal radio

Lorsque l'on planifie l'utilisation d'un terminal de type  $t_k$  pendant un time-slot, on utilise schématiquement un rectangle d'une certaine aire dans l'intervalle temps-fréquence. La figure 10.6 illustre l'utilisation dans un même spot de deux terminaux de types différents.

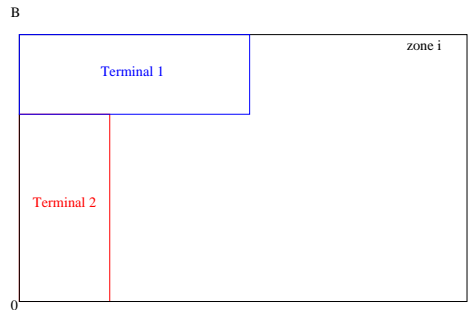


FIG. 10.6 – Utilisation de terminaux de types différents.

On utilise alors un algorithme de placement qui permet, pour les données du problème, de résoudre le problème du placement en sacrifiant moins de 1% de la bande passante.

#### 10.3.3.2 Typage des familles.

La méthode proposée consiste, au lieu d'allouer des time-slots d'un certain type dans un slot, à allouer des familles typées simultanément dans l'ensemble des spots.

Pour une famille  $F_i$  on peut choisir un type de transmission  $t_k$  qui sera diffusé sur tous les spots concernés. On place cette transmission au même endroit du plan de fréquence pour tous les spots concernés. De cette manière, le niveau d'interférence, à cet endroit de la bande est respecté. Sur une largeur de bande voisine, on peut utiliser une autre famille qui servira à satisfaire un autre (ou le même) type de demande.

Une famille peut éventuellement recevoir plusieurs types suivant les spots. Elles sont alors dites *multiple*. Si  $t_k$  est le type de la famille de plus large bande et  $t_{k'}$  celui de largeur de bande la plus étroite, alors la famille utilisera l'aire  $(\max(t_k^t), \max(t_{k'}^b))$ . Sur chaque spot, la famille se composera de time-slot de type différents, mais de même nombre.

On utilise alors un programme linéaire pour trouver les familles adéquates. Il reste ensuite à leur attribuer un type.

On remarque qu'étant donné une famille non-typée  $F_i$ , il existe  $\tau^N$  manières différentes de typer cette famille, soit bien plus qu'il n'est raisonnable de mettre dans notre programme linéaire ! C'est pourquoi on utilise le concept de *génération de colonnes*.

### 10.3.4 Résultats numériques

Nous avons testé différents types de configurations entre 8 et 32 spots. La demande a été générée suivant des exemples fournis par ALCATEL et les interférences ainsi que les gains, ont été générés aléatoirement suivant une loi uniforme en dB, selon les spécifications données par un document ALCATEL. L'interférence a été choisie, pour 85% de sa valeur, comme une interférence de voisinage (les six spots voisins) et pour 15%, comme une interférence globale.

La sortie "classique" de nos programmes est un plan d'allocation temps/fréquence, comme celui indiqué sur la figure 10.7. Le temps est représenté en abscisse et la fréquence en ordonnée. Les times-slots sont notés sous la forme "Type ij" dans la légende,  $i$  et  $j$  représentant respectivement le numéro de plus faible et de plus fort type. Ainsi un time-slot de type "13" pourra contenir  $n$  porteuses de type 1, 2 ou 3 selon les spots avec  $n = t_3^t/t_1^t$ .

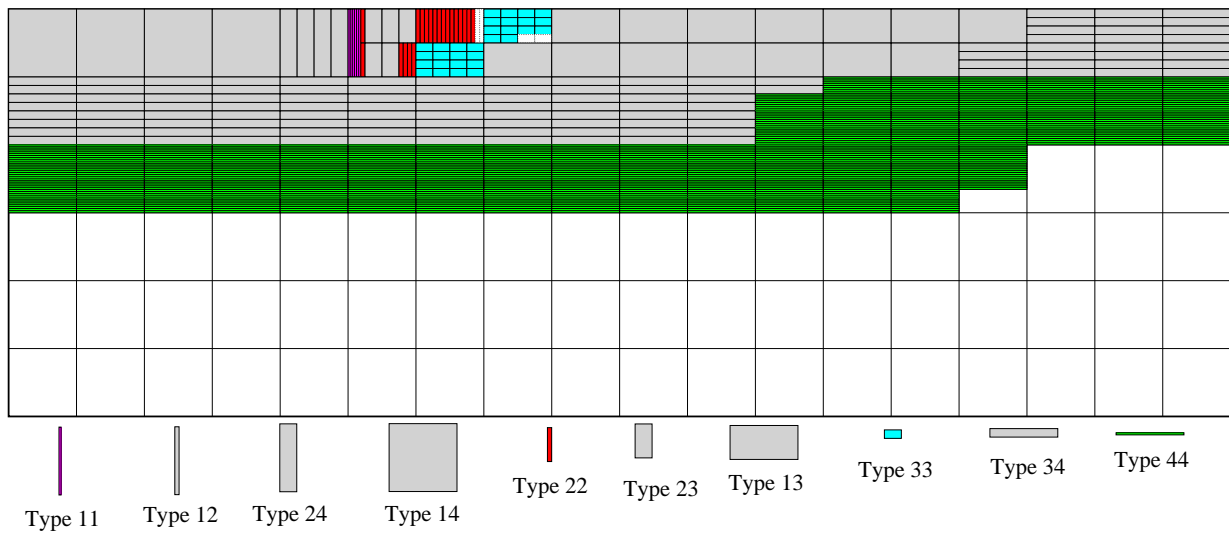


FIG. 10.7 – Un exemple de résultat d'allocation de ressources.

Si l'on considère des configurations à 8 spots par couleur, pour des temps de calcul de l'ordre de la minute nos programmes peuvent trouver la solution flottante optimale. Ce résultat est particulièrement satisfaisant, car il permet d'analyser avec précision les causes d'interférences au niveau des spots.

Pour une configuration de 32 spots, nous proposons un fonctionnement non-optimal partant d'un nombre de familles limité. Le nombre de familles augmente le temps de résolution tout en améliorant la solution. Il est donc possible de trouver un compromis pour l'élaboration d'une solution en temps réel.

### 10.3.5 Conclusion

Notre solution au problème des questions d'interférence vient de la constatation simple que pour tout couple temps-fréquence  $(t, f)$ , il existe une famille  $Z$  de zones simultanément actives, qui vérifient les conditions d'interférences. Il est alors assez naturel d'utiliser le concept d'émissions non-concurrentes de manière similaire à ce qui se fait pour le coloriage

de graphes [43]. Dans le cas du coloriage de graphe, on s'appuie sur un ensemble de familles de sommets indépendants. Ici, on utilise des familles de zones pouvant communiquer simultanément.

Comme nous l'avons vu dans la partie introductive, les approches classiques ne considèrent pas de problème d'interférence couplées à un choix multiple de type de porteuses. Le fait de les considérer simultanément améliore grandement l'utilisation faite des ressources, c'est-à-dire ici du plan constitué de la gamme de fréquences disponibles sur la durée de la trame.

Une approche beaucoup plus simple serait de diviser la bande passante totale en un certain nombre de porteuses (typées), et d'appliquer un algorithme de remplissage sur ces porteuses. Cela reviendrait à utiliser une superposition de système TDMA.

Dans notre approche, nous utilisons donc au mieux les capacités d'un système MF-TDMA. Malheureusement, dans les équipements mobiles actuels, il n'est pas encore possible pour un terminal de pouvoir émettre dans un time-slot quelconque. Cependant, nous pouvons espérer que les systèmes futurs permettront de telles libertés, qui permettent, de façon évidente, une bien meilleure utilisation du spectre.

# Conclusion

Nous avons proposé une méthode de résolution de l'allocation équitable et optimale de porteuses à des opérateurs par une méthode originale en deux étapes. Les résultats offrent une très bonne utilisation des ressources dans les systèmes TDMA.

La méthode est plus générale que celles que nous avons pu trouver dans la littérature car elle réunit plusieurs difficultés, jusque là considérées séparément. Tout d'abord l'on considère que les spots sont divisés en zones, chacune ayant des propriétés d'interférences distinctes. De plus, les spots n'interfèrent plus deux à deux mais de façon globale. Les terminaux peuvent utiliser un certain nombre de types différents, correspondant à des quantités de bande passante distinctes. Enfin, la taille des systèmes peut être assez large, puisque nos simulations, incluant jusqu'à 32 spots par couleurs sont les plus grands systèmes considérés parmi ceux des articles mentionnés précédemment. Enfin, bien évidemment, l'approche est novatrice dans la mesure où elle intègre des notions d'équité.

Nous pouvons noter ici que nous n'avons pas pu utiliser notre concept général d'équité associé aux fonctions d'utilité mais uniquement la version pondérée. En effet, les allocations sont adressées à des opérateurs et non des utilisateurs finaux. Ainsi, ils ne sont pas associés à une mais un ensemble d'applications.

L'intérêt de ce type de réseaux était également de fournir un exemple typique de cas où des contraintes d'intégrité apparaissent. De par la simplicité relative du problème (car les fonctions d'utilité n'intervenaient pas), nous avons pu fournir un algorithme original et spécifique, sans nécessiter l'introduction d'un solveur général tel que SDP.



# Conclusion



## Conclusion

Les travaux présentés dans cette thèse portent sur l'adaptation de concepts d'équité aux réseaux mobiles. L'idée fondamentale a été d'appliquer une formulation globale de l'équité non pas à la ressource que l'on cherche à allouer mais à une fonction de celle-ci. Cette fonction représente la satisfaction que l'utilisateur tire de l'accès à la ressource. Cette idée est née de l'étude des travaux faits par John Nash dans les années 50 dans le cadre de la théorie des jeux coopératifs. En effet, Nash s'est intéressé à des problèmes de marchandage, c'est-à-dire des situations où deux joueurs possédant des biens de nature différentes collaborent afin de trouver un accord. Chacun associe une valeur relative à chaque objet de la transaction.

La formulation générale de l'équité que nous utilisons est celle introduite par Mo et Walrand dans [85]. Elle est définie sous la forme d'un problème d'optimisation contenant un paramètre  $\alpha$ . Les concepts classiques d'équité, tels que l'équité proportionnelle et l'équité max-min correspondent aux équilibres obtenus pour des valeurs particulières du paramètre d'équité.

Nous n'avons pas détaillé ici les façons de déterminer les fonctions d'utilité. On trouve dans la littérature de nombreux travaux présentant les qualités de perceptions liées à l'utilisation de différents types d'applications temps réels, tels que la vidéo ou la téléphonie notamment. Nous ne nous sommes intéressés ici uniquement à l'allure de ces fonctions et avons proposé de les approcher par des fonctions quadratiques paramétrisables.

Nous avons considéré trois types de réseaux. En effet, a priori l'allocation équitable se justifie dans tout réseau dans lequel évoluent des applications temps-réel et dans lequel on peut allouer une quantité variable ressource. Or nous pouvons voir que jusqu'à présent seules quelques maigres tentatives d'intégration de l'équité avaient vu le jour, principalement dans les réseaux terrestres. De plus, nous avons pu voir que les problèmes d'optimisation correspondants aux allocations dans les différents réseaux sont de nature très diverse. En particulier, nous avons vu que pour les réseaux utilisant des communications radio, le système de contraintes est nettement plus complexe que dans les réseaux filaires.

Les réseaux que nous avons considérés sont les réseaux filaires, les réseaux de téléphonie mobile de troisième génération (de type UMTS) et enfin un système satellitaire géostationnaire MF-TDMA. Le premier exemple est le plus aisé à résoudre car le système de contraintes se compose d'une suite d'inégalités linéaires. Nous avons vu que dans le cadre des réseaux mobiles la difficulté du problème vient de l'existence de contraintes jointes entre puissance et débit. Enfin, dans les réseaux satellitaires la difficulté naît de l'existence de contraintes d'intégrité (en effet, on suppose que l'on ne peut allouer n'importe quelle quantité de débit mais seulement un certain nombre de porteuses d'un certain type).

Nous avons proposé une méthode de résolution pour les trois systèmes étudiés. Dans les deux premiers nous avons utilisé un programme mathématique appelé SDP (Semi-Definite Programming) qui peut être vu comme une extension de la programmation linéaire. J'ai expliqué en détail la façon dont elle a été implémentée dans le cadre des réseaux terrestres. Dans l'étude des systèmes cellulaires, j'ai présenté une étude mathématique du système de contraintes et donné une nouvelle formulation du problème dans lequel les contraintes



s'exprime simplement. L'implémentation avec la SDP qui en découle est alors très similaire à celle utilisée dans les réseaux terrestres. Enfin, dans le cadre des communications par satellites, bien qu'il aurait été possible de proposer de nouveau une solution basée sur la SDP (car celle-ci permet d'intégrer des problèmes aux contraintes entières), nous avons proposé un algorithme adapté et nous avons montré qu'il était de complexité polynômiale.

Les possibilités d'extensions de ces travaux sont multiples. Un axe majeur serait de considérer des méthodes décentralisées permettant d'obtenir des équilibres correspondants aux différents critères d'équité, en particulier dans les réseaux terrestres. En effet, dans les réseaux mobiles et satellitaires l'allocation est gérée de façon centralisée. Un autre axe serait de considérer les problèmes posés par l'introduction de connexions multicast. Enfin, on peut noter que d'autres concepts d'équité ont été développés dans le cadre des jeux coopératifs, comme par exemple la solution proposée par Thomson ([1]). Il serait intéressant de pouvoir les comparer et d'en étudier plus avant les propriétés et notamment leur comportement face aux différents paramètres des fonctions d'utilités.

# Annexe A

## Programme de création des matrices d'entrées.

Le programme donné ci-dessous est un exemple de ce que l'on peut utiliser pour créer les matrices d'entrées au solveur csdp. Dans le cas présenté ici, la matrice de trafic ainsi que tous les arguments du problème font partie du programme. Il est évidemment possible de le modifier afin de les transformer en arguments.

Cet exemple est codé en langage C pour le cas où  $\alpha = 1$ . Il s'agit du programme utilisé pour résoudre le problème présenté dans la section 4.4.1.

```
main(int argc, char *argv[]) {

/* Problem parameters */
// Network topology
int n = 11, liens = 10;
int mat[10][11]= {
    {0,1,0,0,0,0,0,1,0,0,0},
    {0,1,0,0,0,0,0,1,1,0,0},
    {0,1,0,0,0,0,0,0,1,0,0},
    {0,0,0,0,0,0,0,0,1,0,0},
    {1,0,0,1,0,0,0,0,0,1,0},
    {1,0,0,1,1,0,0,0,0,1,0},
    {1,0,0,0,1,0,0,0,0,1,1},
    {0,0,1,0,1,0,1,0,0,1,1},
    {0,0,1,0,0,0,0,0,0,0,1},
    {0,0,0,0,0,1,1,0,0,1,0}};

// Connections parameters (supposed to be identical)
double T = 3;
double PR = 80;
double MR = 10;
double fPR = 200;
double cap = 100;

// linkage with the initial parameters
double alpha = (2*fPR)/(T*(PR-MR))-1;
double a = T*(1-alpha) / (2*(PR-MR)) ;
double b = (PR - alpha*MR) / (1-alpha);
double c = T*(PR-MR) / (2*(1-alpha));
```

```

int i, j, nblocs;
int prof=0;
int p=0; /*p is the number of additional variables*/
int compteur=0;
int vraiN=1;
int temp, temp2, compt, compt2;

// computation of vraiN, p and prof.
// prof and vraiN are such that for all t < prof, 2^t < vraiN
while (vraiN<n) {
    prof++; p+=vraiN; vraiN*=2;
}
p+=vraiN;
printf("* SDP program for proportional fairness\n");
printf("%d = mdim\n",p+vraiN); /* total number of variables */

// total number of blocs : vraiN (for the xi) + (vraiN-1) (linkage of the "fi")
//                               + liens (links capacities) + vraiN (constraints MR)
//                               + (vraiN) (constraints MR)
// Remark : p=2*vraiN-1
nblocs = p+2*vraiN+liens;
printf("%d = nblocs\n",nblocs);

// Line representing the different block sizes.
printf("{2}");
for (i=1; i<p;i++) printf(", 2");
for (i=1; i<=2*vraiN+liens; i++) printf(", 1");
printf("}\n");

// vector L
for (i=1;i<=p+vraiN-1;i++) printf("0 ");
printf("-1\n");

/* A0 */
for (i=1;i<=vraiN;i++) {
    printf("0 %d 2 2 -1\n", i);
    printf("0 %d 1 1 %g\n", i, -c/a);
    printf("0 %d 1 2 %g\n", i, b);
}
for (i=1;i<=liens;i++) {
    printf("0 %d 1 1 %g\n", p+i, -cap);
}

/*A1 to AvraiN */
for (i = 1; i<=vraiN; i++) {
    printf("%d %d 1 2 1\n", i, i); /*link between xi and fi*/
    printf("0 %d 1 1 %g\n", i, i+liens+p, MR); /* x > 0 */
    printf("%d %d 1 1 1\n", i, i+liens+p); /* x > 0 */
}
for (i=1; i<=liens; i++)
    for (j=1; j<=n; j++)
        if (mat[i-1][j-1]==1)
            printf("%d %d 1 1 %d\n",j,p+i,-mat[i-1][j-1]);

```

```

/*AvraiN+1 to AvraiN*2 */
temp=vraiN+1;temp2=1;
for (i=vraiN+1; i<=2*vraiN; i++) {
    printf("%d %d 1 1 %g\n", i, i-vraiN, -1/a);
}
for (i=vraiN+1; i<=2*vraiN; i++) {
    printf("%d %d %d %d 1\n", i, temp, temp2, temp2);
    if (temp2==1) {temp2=2;}
    else {temp2=1;temp+=1;}
}

compt = vraiN/2;
compt2 = vraiN+vraiN;

while ( compt >1) {
    for (i=compt2+1; i<=compt2+compt; i++) {
        printf("%d %d 1 2 1\n", i, i-vraiN);
        printf("%d %d %d %d 1\n", i, temp, temp2, temp2);
        if (temp2==1) {temp2=2;}
        else {temp2=1;temp+=1;}
    }

    compt2+=compt;
    compt = compt/2;
}

/*last variable */
printf("%d %d 1 2 1\n", vraiN+p, p);

/* max constraints */
for (i=1;i<=vraiN;i++) {
    printf("0 %d 1 1 %g\n", p+liens+vraiN+i,-PR);
    printf("%d %d 1 1 -1\n",i,p+liens+vraiN+i);
}
}

```



# Annexe B

## Les réseaux satellitaires.

### B.1 Relation d'ordre parmi les allocations

On dit qu'une allocation  $J'$  est plus grande qu'une autre allocation  $J$  (et nous écrivons  $J' \succeq J$ ) si et seulement si

$$\begin{aligned} J'_1 &\geq J_1, \\ J'_2 + J'_1 &\geq J_2 + J_1, \\ &\vdots \\ &\vdots \\ \sum_{t=1}^N J'_t &\geq \sum_{t=1}^N J_t. \end{aligned}$$

La relation  $\succeq$  est un ordre partiel. Nous avons la propriété suivante de la relation d'ordre :

**Proposition B.1** *Soient deux allocations  $J$  et  $J'$ . Il existe une unique allocation  $J^* = \max(J, J')$  telle que  $J^* \succeq J$ ,  $J^* \succeq J'$ , et telle que pour toute allocation  $J''$  telle que  $J'' \succeq J$  et  $J'' \succeq J'$ , nous avons  $J'' \succeq J^*$ .*

**Proof** Posons

$$\begin{aligned} J_1^* &= \max\{J_1, J'_1\} \\ &\vdots \\ J_i^* &= \max \left\{ \sum_{k=1}^{k=i} J_k, \sum_{k=1}^{k=i} J'_k \right\} - \sum_{k=1}^{k=i-1} J_k^* \\ &\vdots \\ J_N^* &= \max \left\{ \sum_{k=1}^{k=N} J_k, \sum_{k=1}^{k=N} J'_k \right\} - \sum_{k=1}^{k=N-1} J_k^* \end{aligned}$$

Alors  $J^*$  satisfait les propriétés requises. ■

L'interprétation de la proposition précédente dans notre contexte est la suivante. Soient données des contraintes minimales  $J$  et  $J'$  des taux de deux opérateurs. Une allocation  $J''$ , candidate pour être une allocation équitale (faisable), doit satisfaire  $J'' \succeq J^*$ . Ceci s'étend de façon évidente pour un nombre quelconque d'opérateurs. La valeur  $J^*$ , qui peut

être déterminée de façon explicite, comme fait dans la preuve de la proposition B.1, peut désormais servir de point de départ pour les algorithmes d'allocations équitables : nous déterminons tout d'abord l'élément  $J^*$  et trouvons ensuite la meilleure allocation  $J''$  telle que  $J'' \succeq J^*$ .

Malheureusement, la bande passante allouée à notre service est limitée. Il y a donc une contrainte supplémentaire sur le nombre maximum de porteuses de chaque type. La contrainte est de la forme :

$$\sum_{i=1}^N B_i J_i \leq C \quad (\text{B.1})$$

avec  $C$  la bande passante totale et  $B_i$  la bande passante d'une  $i$ -porteuse.

**Remarque B.1** Soit  $\mathcal{B}$  l'ensemble des allocations acceptables par rapport à la bande passante. Évidemment, si  $J \succeq J'$  et  $J \in \mathcal{B}$  alors  $J' \in \mathcal{B}$ . Cependant, il est possible que pour un  $J \in \mathcal{B}$  et  $J' \in \mathcal{B}$  nous avons  $\max(J, J') \notin \mathcal{B}$ . En d'autres termes, une allocation peut être localement mais pas globalement acceptable.

Dans notre algorithme d'allocation nous considérerons les vecteurs  $J$  qui satisfont les contraintes des utilisateurs et qui sont maximaux dans l'ensemble des vecteurs acceptables (c'est-à-dire qui satisfont  $J' \in \mathcal{B}$  et tels que  $\forall J'' \succeq J', J'' \notin \mathcal{B}$ ).

## B.2 Optimalité de l'algorithme

Le but de cette sous-section est de montrer que notre algorithme trouve une allocation optimale de porteuses pour un vecteur fixé donné  $J$ .

Dans le contexte général, nous considérons un ensemble fini  $I$  d'éléments aux valeurs positives. Soit  $v_i$  la valeur associée à l'élément  $i$ .

**Définition B.1** Soient  $A$  et  $B$  deux sous-ensembles disjoints de  $I$ . Soit alors  $\text{Union}(A, B, k)$  un sous-ensemble de  $H$  contenant au plus  $k$  éléments de  $A \cup B$  tels que  $\sum_{i \in H} v_i$  est maximum.

Notons que si  $I = \{1, \dots, p\}$  et  $v_1 \geq \dots \geq v_p$ ,  $H$  se compose des  $k$  plus petits indices de  $A \cup B$ .

**Définition B.2** Soit  $A$  et  $B$  deux sous-ensembles de  $I$ . On dit que  $A$  domine  $B$  ssi  $|A| \geq |B|$  et il existe une injection  $\varphi$  de  $B$  à  $A$  telle que

$$\forall i \in B \quad v_i \leq v_{\varphi(i)}.$$

**Proposition B.2** Soit  $A, B, C$  et  $D$  trois sous-ensembles de  $I$  tels que

- $A$  domine  $B$ ,
- $C$  domine  $D$ ,
- $A \cup B$  et  $C \cup D$  sont disjoints.

Alors, pour tout  $k \in \mathbb{N}$ ,  $\text{Union}(A, C, k)$  domine  $\text{Union}(B, D, k)$ .

**Démonstration :** Soit  $\varphi_1$  un mapping dominant de  $B$  vers  $A$ , et  $\varphi_2$  un de  $D$  vers  $C$ .

On définit  $\psi$  de  $B \cup D$  vers  $A \cup C$  comme suit :

$$\begin{aligned} \psi(v) &\mapsto \varphi_1(v) && \text{si } v \in B, \\ \psi(v) &\mapsto \varphi_2(v) && \text{si } v \in D. \end{aligned}$$

De façon évidente, la restriction de  $\psi$  à  $Union(B, D, k)$  génère un ensemble d'au plus  $k$  éléments dans  $A \cup C$ , qui est par définition dominé par  $Union(A, C, k)$ . ■

Nous considérons maintenant que à chaque élément  $i$  de  $I$  est associé un type  $t_i$  et appelons

$$A_p = \{i \in A : t_i = p\}.$$

Alors notre problème d'allocation consiste à trouver un ensemble  $A$  tel que

$$\forall k \in \{1, \dots, N\} \quad |A_1 \cup \dots \cup A_k| \leq J_1 + \dots + J_k.$$

**Théorème B.1** Soit  $A^{(k)}$  construite comme suit :

$$\begin{aligned} A^{(0)} &= \emptyset \\ A^{(k+1)} &= Union(A^{(k)}, I_{k+1}, J_1 + \dots + J_{k+1}) \end{aligned}$$

alors  $A^{(N)}$  est une solution de notre problème d'allocation qui domine toutes les autres solutions.

**Démonstration :** Evidemment,  $A^{(N)}$  est une solution de notre problème d'allocation.

Soit  $B$  une autre solution. Nous montrons par récurrence que  $A^{(k)}$  domine  $B_1 \cup \dots \cup B_k$ .

En fait, on a  $B_1 \cup \dots \cup B_{k+1} = Union(B_1 \cup \dots \cup B_k, B_{k+1}, J_1 + \dots + J_{k+1})$ . Clairement  $I_{k+1}$  domine  $B_{k+1}$ , et par récurrence  $A^{(k)}$  domine  $B_1 \cup \dots \cup B_k$ . D'où le résultat par la proposition B.2. ■





# Bibliographie

- [1] Traffic management specification, version 4.0. Technical Report AF-TM-0056.000, ATM Forum Traffic Management Working Group, April 1996.
- [2] Cost 239. Ultra high capacity optical transmission networks, final report of Cost project 239. *ISBN 953-184-013-X*, 1999.
- [3] Micah Adler, Jin-Yi Cai, Jonathan K. Shapiro, and Don Towsley. Estimation of congestion price using probabilistic packet marking. *preprint*, 2002.
- [4] Sara Alouf, Jérôme Galtier, Jean-François Lalande, and Corinne Touati. Un algorithme de programmation linéaire pour la résolution de ressources. Technical report, INRIA, 2003. (to be published).
- [5] Tansu Alpcan, Tamer Başar, Rayadurgam Srikant, and Eitan Altman. CDMA uplink power control as a noncooperative game. In *40th IEEE Conference on Decision and Control*, pages 197–202, Orlando, Florida, December 4-7 2001.
- [6] Eitan Altman. Capacity of multi-service CDMA cellular networks with best-effort applications. *Mobicom*, 2002.
- [7] Eitan Altman, Afonso Ferreira, and Jérôme Galtier. *Les réseaux satellitaires de télécommunication*. SNEL, 1999.
- [8] Gunnar Andersson, Lars Engebretsen, and Johan Hästad. A new way to use semidefinite programming with applications to linear equations mod  $p$ . *Journal of Algorithms*, 39 :162–204, February 2002.
- [9] Ambalavanar Arulambalam and Xiaoqiang Chen. Allocating fair rates for available bit rate service in ATM networks. *IEEE Communications Magazine*, pages 92–100, 1996.
- [10] Sanjeeva Athuraliya, Victor H. Li, Steven H. Low, and Qinghe Yin. Rem : Active queue management. *IEEE Network*, 15(3) :48–53, 2001. Extended version in Proceedings of ITC17, Salvador, Brazil, September 2001.
- [11] Deepak Ayyagari and Anthony Ephremides. Power control for link quality protection in cellular DS-CDMA networks with integrated (packet and circuit) services. *Mobicom '99, Seattle Washington, USA*, pages 96–102, 1999.
- [12] Andrea Baiocchi, Francesca Cuomo, and Cristina Martello. Optimizing the radio resource utilization of multiaccess systems with a traffic-transmission quality adaptive packet scheduling. *Computer Networks*, 38 :225–246, 2002.
- [13] Randall Berry. *Power and Delay Trade-offs in Fading Channels*. PhD thesis, Massachusetts Institute of Technology, June 2000. Available at : <http://www.ece.nwu.edu/~rberry/pubs.html>.

- [14] Dimitri Bertsekas and Robert Gallager. *Data Networks*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
- [15] Vaduvur Bharghavan, Songwu Lu, and Thyagarajan Nandagopal. Fair queuing in wireless networks, issues and applications. *IEEE Personal Communications*, 6(1), Feb 1999.
- [16] Vineet Birmani. Master of science. Master's thesis, University of Maryland, 1999.
- [17] Thomas Bonald and Laurent Massouillé. Impact of fairness on internet performance. *Sigmetrics*, 2001.
- [18] G. Bongiovanni, Don Coppersmith, and C. K. Wong. An optimum time slot assignment algorithm for an SS/TDMA system with variable number of transponders. *IEEE Trans. Communications*, pages 721–726, May 1981.
- [19] Maurizio A. Bonuccelli. A fast time slot assignment algorithm for TDM hierarchical switching systems. *IEEE Trans. Communications*, COM-37 :870–874, 1989.
- [20] Maurizio A. Bonuccelli, Inder Gopal, and C. K. Wong. Incremental time-slot assignment in SS/TDMA satellite systems. *IEEE Transactions on Communications*, 39 :1147–1156, Jul 1991.
- [21] Jean-Yves Le Boudec. *Rate Adaptation, Congestion Control and Fairness : A Tutorial*. Ecole Polytechnique Federale de Lausanne, Dec 2000.
- [22] Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear matrix inequalities in system and control theory*. Springer, 1994.
- [23] Josepk P. Burke and James R. Zeidler. Data throughput in a multi-rate CDMA reverse link : Comparing optimal spatial combining vs. maximal ratio combining. In *Proc. of IEEE Globecom*, volume Vol. 1, pages 700–705, San Antonio, November 2001.
- [24] S. Joseph Campanella and Roger J. Colby. Network control for multibeam TDMA and SS/TDMA. *IEEE Journal on Selected Areas in Communications*, 01 :174–187, Jan 1983.
- [25] Xi-Ren Cao, Hong-Xia Shen, Rodolfo Milito, and Patrica Wirth. Internet pricing with a game theoretical approach : Concepts and examples. *IEEE/ACM Transactions on Networking*, 10(2) :208–216, Apr 2002.
- [26] Gregory W. Cermak. Subjective quality of speech over packet networks as a function of packet loss, delay and delay varitation. *International Journal of Speech Technology*, 5(1) :65–84, Jan 2002. ISSN 1381-2416.
- [27] Suresh Chalasani and Anujan Varma. Efficient time-slot assignment algorithms for SS/TDMA systems with variable-bandwidth beams. *IEEE Transactions on Communications*, 42(2/3/4) :1359–1370, Feb 1994.
- [28] Anna Charny, David Clark, and Raj Jain. Congestion control with explicit rate indication. In *Proceedings of IEEE International Conference on Communications*, pages 1954–1963, 1995.
- [29] Wen-Tsuen Chen, Pi-Rong Sheu, and Jiunn-Hwa Yu. Time slot assignment in TDM multicast switching systems. *IEEE Trans. Communications*, COM-42 :149–165, Jan 1994.

- [30] Michael Cheung and Jon W. Mark. Resource allocation in wireless networks based on joint packet/call levels qos constraints. *Proc. IEEE GLOBECOM*, pages 271–275, Nov 2000.
- [31] Dah-Ming Chiu and Raj Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN systems*, 17 :1–14, 1989.
- [32] Jaeweon Cho and Daehyoung Hong. Downlink throughput and fairness analysis in multi-rate CDMA systems. *Proceedings of IEEE VTC*, pages 1001–1005, 2001.
- [33] Jon Crowcroft and Philippe Oechslin. Differentiated end-to-end internet services using a weighted proportional fair sharing tcp. *ACM SIGCOMM Computer Communication Review*, 28(3) :53–69, July 1998.
- [34] Anton Ştefănescu and Maria V. Ştefănescu. The arbitrated solution for multi-objective convex programming. *Rev. Roum. Math. Pure Applicat.*, 29(7) :593–598, 1984.
- [35] Ahmed K. Elhakeem, Stephan Bohm, Mohamed Hachicha Tho Le-Ngoc, and H. T. Mouftah. Analysis of a new multiaccess/switching technique for multibeam satellites in a prioritized isdn environment. *IEEE Journal Selected Areas in Comm.*, 10 :378–390, Feb 1992.
- [36] K. Y. Eng and Anthony S. Acampora. Fundamental conditions governing TDM switching assignments in terrestrial and satellite networks. *IEEE Transactions on Communications*, 35 :755–761, Jul 1987.
- [37] Mirko Ferracioli, Velio Tralli, and Roberto Veerdone. Channel adaptive scheduling for a wideband TDD/TCDMA wireless system under heterogeneous traffic conditions. *Computer Networks*, 38 :207–223, 2002.
- [38] Sally Floyd. Congestion control principles. Internet draft RFC 2914, oct 1999.
- [39] Nobuo Funabiki and Yoshiyasu Takefuji. A parallel algorithm for time-slot assignment problems in TDM hierarchical switching systems. *IEEE Trans. Communications*, COM-42 :2890–2898, Oct. 1994.
- [40] Inder S. Gopal, Maurizio A. Bonuccelli, and Chung K. Wong. Scheduling in multibeam satellites with interfering zones. *IEEE Trans. Communications*, COM-31 :941–951, Aug. 1983.
- [41] Inder S. Gopal, Don Coppersmith, and Chung K. Wong. Minimizing packet waiting time in a multibeam satellite system. *IEEE Trans. Communications*, COM-30 :305–316, Feb 1982.
- [42] Inder S. Gopal and Chung K. Wong. Minimizing the number of switchings in an SS/TDMA system. *IEEE Trans. Communications*, COM-33 :497–501, June 1985.
- [43] Martin Grötschel, László Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1 :169–197, 1981.
- [44] Bruce Hajek. Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2), May 1988.
- [45] Bassam Hashem and Halim Yanikomeroglu. Power control for code-division multiple access cellular systems. In *Proc. International Congress on Dynamics and Control of Control Systems (DYCONS'99)*, Ottawa, Canada, 5-6 August 1999.

- [46] Christoph Helmberg. *Semidefinite Programming for Combinatorial Optimization*. ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustraße 7, D-14195 Berlin, Germany, October 2000. Habilitationsschrift, <http://www.zib.de/hemlberg>.
- [47] Thomas H. Henderson, Emile Sahouria, Steven McCanne, and Randy H. Katz. On improving the fairness of TCP congestion avoidance. *IEEE Globecom conference, Sydney*, 1998.
- [48] Enrique J. Hernandez-Valencia and Lbachir Benmohamed. Rate control algorithms for the atm abr service. *European Trans. Telecom*, 8 :7–20, 1997.
- [49] Kimmo Hiltunen and R. De Bernardi. WCDMA downlink capacity estimation. *VTC'2000*, 992–996, 2000.
- [50] Harri Holma and Antti Toskala. WCDMA for UMTS. *Revised Edition, J. Wiley & Sons*, 2001.
- [51] T. Inukai. An efficient SS/TDMA time slot assignment algorithm. *IEEE Trans. Communications*, COM-27 :1449–1455, Oct. 1979.
- [52] Bijan Jabbari and David McDysan. Performance of demand assignment TDMA and multicarrier TDMA satellite networks. *IEEE Journal on Selected Areas in Communications*, 10 :478–486, Feb. 1992.
- [53] Van Jacobson and Michael J. Karels. Congestion avoidance and control. *ACM Computer Communication Review*, 18 :314–329, 1988. Proceedings of the Sigcomm '88 Symposium in Stanford, CA, August, 1988.
- [54] Jan Janssen, Danny De Vieeschauwer, Maarten Büchli, and Guido H. Petit. Assessing voice quality in packet based telephony. *IEEE Internet Computing*, 6(3) :48–56, 2002.
- [55] John F. Nash Jr. The bargaining problem. *Econometrica*, 18 :155–162, 1950.
- [56] Koushik Kar, Saswati Sarkar, and Leandros Tassiulas. Optimization based rate control for multirate multicast sessions. In *IEEE INFOCOM'01*, pages 123–132, Anchorage, Alaska, April 2001.
- [57] Koushik Kar, Saswati Sarkar, and Leandros Tassiulas. A simple rate control algorithm for maximizing total user utility. In *Infocom*, Anchorage, Alaska, April 2001.
- [58] Shuzo Kato, Takemi Arita, and Kozo Morita. Onboard digital signal processing technologies for present and future TDMA and SCPC systems. *IEEE Journal on Selected Areas in Communications*, 05 :685–700, May 1987.
- [59] Shuzo Kato, Masahiro Morikura, and Shuji Kubota et al. A TDMA satellite communication system for isdn services. *IEEE Journal Selected Areas in Comm.*, 10 :456–464, Feb. 1992.
- [60] Frank P. Kelly. Charging and rate control for elastic traffic. *European Trans. on Telecom.*, 8 :33–37, 1998.
- [61] Frank P. Kelly, Aman Maulloo, and David Tan. Rate control for communication networks : shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49, 1998.

- [62] Seong-Lyun Kim, Zvi Rosberg, and Jens Zander. Combined power control and transmission selection in cellular networks. *IEEE Vehicular Technology Conference*, Fall 1999.
- [63] C. King, P. Trusty, J. Jankowski, R. Duesing, and P. Roach. INTELSAT TDMA/DSI burst time plan development. *Int. J. Satellite Commun.*, 3(1-2) :35–43, 1985.
- [64] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598) :671–679, 1983.
- [65] Jon Kleinberg and Michel Goemans. The lovasz theta function and a semi-definite programming relaxation of vertex cover. *SIAM J. Discrete Math*, 11, 1998.
- [66] Srisankar Kunniyur and Rayadurgam Srikant. A time scale decomposition approach to adaptive ECN marking. In *Infocom*, pages 1330–1339, 2001.
- [67] Jaana Laiho and Achim Wacker. Radio network planning process and methods for WCDMA. *Ann. Telecommun.*, 56(5-6), 2001.
- [68] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. Preliminary version appeared as Report PNA-R0210, CWI, Amsterdam, April 2002, December, 17 2002.
- [69] Rangsan Leelahakriengkrai and Rajeev Agrawal. Scheduling in multimedia CDMA wireless networks. Technical Report Technical Report ECE-99-3, ECE Dept., University of Wisconsin - Madison, July 1999.
- [70] J. L. Lewandowski, J. W. S. Liu, and C. L. Liu. SS/TDMA time slot assignment with restricted switching modes. *IEEE Trans. Communications*, COM-31 :149–154, Jan. 1983.
- [71] Ellen L.Hahne. Round-robin scheduling for max-min fairness in data networks. *IEEE Journal on Selected Area in Communications*, 9(7) :1024–1039, Sep 1991.
- [72] Steven H. Low, Larry L. Peterson, and Limin Wang. Understanding TCP vegas : a duality model. In *SIGMETRICS/Performance*, pages 226–235, 2001.
- [73] Steven H. Low and Rayadurgam Srikant. A mathematical framework for designing a low-loss, low-delay internet. *Network and Spatial Economics*, 2003 (to appear).
- [74] Songwu Lu, Vaduvur Bharghavan, and Rayadurgam Srikant. Fair scheduling in wireless packet networks. *IEEE/ACM Trans. on Networking*, 7(4) :473–489, 1999. Egalement ACM SIGCOMM Aug 1997.
- [75] Duncan Luce and Howard Raiffa. *Games and Decisions*. New York\* : Wiley, 1957.
- [76] Jon W. Mark and Shihua Zhu. Power control and rate allocation in multirate wideband CDMA systems. *Proc. IEEE WCNC*, pages 168–172, Sept 2000.
- [77] Laurent Massouillé and James W. Roberts. Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems*, 2000.
- [78] Laurent Massoulié and James Roberts. Bandwidth sharing : objectives and algorithms. In *IEEE INFOCOM*, volume 3, pages 1395–1403, 1999.
- [79] Ravi Mazumdar, Lorne G. Mason, and Christos Douligeris. Fairness in network optimal flow control : optimality of product forms. *IEEE Trans. on Comm.*, 39 :775–782, 1991.

- [80] Jeremy W. McLean and Hugo J. Woerdeman. Spectral factorization and sums of squares representations via semidefinite programming. *SIAM J. Matrix Anal. Appl.*, 23(3) :646–655, May 2001.
- [81] N. Metropolis, A.W. Rosenbluth, M.N.Rosenbluth, A.H. Teller, and E.Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), June 1953.
- [82] Gyorgy Miklós and Sándor Molnár. Fair allocation of elastic traffic for a wireless base station. *IEEE Globecom*, pages 1673–1678, Dec 1999. Rio de Janeiro.
- [83] Michel Minoux and Christian Brouder. Models and algorithms for optimal traffic assignment in SS/TDMA switching systems. *Int. J. Satellite Commun.*, 5 :33–47, 1987.
- [84] Takeshi Mizuike, Yasuhiko Ito, Lan N. Nguyen, and Eijiro Maeda. Computer-aided planning of SS/TDMA network operation. *IEEE Journal on Selected Areas in Communications*, 09 :37–47, Jan 1991.
- [85] Jeonghoon Mo and Jean Walrand. Fair end-to-end window-based congestion control. In *SPIE '98, International Symposium on Voice, Video and Data Communications*, 1998.
- [86] Abhinay Muthoo. *BARGAINING THEORY WITH APPLICATIONS*. Cambridge University Press, 1999.
- [87] John Nagle. Congestion control in ip/tcp internetworks. Internet draft RFC 896, Jan 1984.
- [88] Arkadi Nemirovski. What can be expressed via conic quadratic and semidefinite programming? *Lecture notes*, 1998.
- [89] Arkadi Nemirovski. Five lectures on modern convex optimization. CORE Summer School on Modern Convex Optimization, August 2002.
- [90] Aleksandar Neskovic, Natasa Naskovic, and George Paunovic. Modern approaches in modeling of mobile radio systems propagation environment. *IEEE Communications surveys*, pages 2–12, 2000.
- [91] Seong-Jun Oh and Kimberly M. Wasserman. Adaptive resource allocation in power constrained CDMA mobile networks. *Proc. IEEE WCNC*, pages 510–514, September 1999.
- [92] Seong-Jun Oh and Kimberly M. Wasserman. Optimality of greedy power control and variable spreading gain in multi-class CDMA mobile networks. *Mobicom'99, Seattle Washington USA*, pages 102–112, 1999.
- [93] Seong-Jun J. Oh and Kymberly M. Wasserman. Dynamic spreading gain control in multiservice CDMA networks. *IEEE J. Selected Area in Comm.*, 17(5) :918–927, May 1999.
- [94] Guillermo Owen. *Game theory*. San Diego, CA : Academic, 1995.
- [95] Larry C. Palmer and Lawrence W. White. Demand assignment in the acts lbr system. *IEEE Trans. Communications*, COM-38 :684 – 692, May 1990.

- [96] Jung M. Park, Uday R. Savagaonkar, and Edwin K.P. Chong et al. Efficient resource allocation for qos channels in MF-TDMA satellite systems. In *Proceedings of the 2001 IEEE Military Communications Conference*, pages Section U19.6, pp. 1–5, October 2000.
- [97] Michael Pióro, Gábor Malicskó, Gábor Fodor, and TOmasz Szymanski. Optimal link capacity dimensioning in proportionally fair networks. In E. Gregori et al., editor, *LNCS 2345*, pages 277–288, Networking 2002.
- [98] Sudhir Ramakrishna and Jack M. Holtzman. A scheme for throughput maximization in a dual-class CDMA system. *IEEE Journal Selected Areas in Comm.*, 16 :830–844, August 1998.
- [99] K. Ramakrishnan and S. Floyd. A proposal to add explicit congestion notification (ECN) to ip. Internet draft RFC 2481, Jan. 1999.
- [100] Motakuri V. Ramana. An exact duality theory for semi-definite programming and its complexity implications. *Math. Programming*, 1996.
- [101] R. Ramaswamy and P. Dhar. Comments on an efficient SS/TDMA time slot assignment algorithm. *IEEE Tran. Communications*, COM-32 :1061–1065, September 1984.
- [102] Alvin E. Roth. *Axiomatic Models of Bargaining*. New York\* : Wiley, 1979.
- [103] Tolga Örs, Zhili Sun, and Barry G. Evans. A meshed vsat satellite architecture using an on-board atm switch. In *IEEE International Performance Computing and Communication Conference*, pages 208–214, Tempe/Phoenix, Arizona, 5-7 February 1997. ISBN :0-78-033874-X.
- [104] Dan Rubenstein, James F. Kurose, and Donald F. Towsley. The impact of multicast layering on network fairness. In *SIGCOMM*, pages 27–38, 1999. Also Technical Report 99-08, available at <http://www-net.cs.umass.edu/papers/papers.html>.
- [105] Scott Shenker. Fundamental design issues for the future internet. *IEEE J. on Selected Areas in Comm.*, 1995.
- [106] A. K. Sinha. A model for TDMA burst assignment and scheduling. *COMSAT Tech.Rev.*, 6, Fall 1976.
- [107] Chi W. Sung and Wing S. Wong. Power control and rate management for wireless multimedia CDMA systems. *IEEE Trans. on Communications*, 49(7) :1215–1226, July 2001.
- [108] Benjamin Teitelbaum, Susan Hares, Larry Dunn, and al. Internet2 qbone – building a testbed for differentiated services. *IEEE Network*, 13(5) :8–16, Sept/Oct 1999.
- [109] Yiu K. Tham. Burst assignment for satellite-switched and earth-station frequency-hopping TDMA networks. *Proc. IEE*, 137 part I(4) :247–255, Aug. 1990.
- [110] Yiu K. Tham. On fast algorithms for TDM switching assignments in terrestrial and satellite networks. *IEEE Transactions on Communications*, 43(8) :2399–2404, August 1995.
- [111] William Thomson. Nash’s bargaining solution and utilitarian choices rules. *Econometrica*, 49(2) :535–538, 1981.



- [112] Tuna Toker, Eitan Altman, Jérôme Galtier, and Corinne Touati. Slot allocation in a TDMA satellite system : Simulated annealing approach. In *AIAA International Communication Satellite Systems Conference and Exhibit (ICSSC)*, Yokohama, Japan., 2003.
- [113] Corinne Touati, Eitan Altman, and Jérôme Galtier. Fair power and transmission rate control in wireless networks. In *Globecom 2002*, 2002.
- [114] Milan Vojnovic, Jean-Yves Le Boudec, and Catherine Boutremans. Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times. In *INFOCOM (3)*, pages 1303–1312, 2000.
- [115] Liang Xu, Xuemin Shen, and Jon W. Mark. Performance analysis of adaptive rate and power control for data service in DS-CDMA systems. *Proc. IEEE GLOBECOM*, pages 627–631, November 2001.
- [116] Masami Yabusaki and Shigefusa Suzuki. Approximate performance analysis and simulation study for variable-channel-per-burst ss-tdma. *IEEE Trans. Communications*, COM-38 :318–326, March 1990.
- [117] Haïkel Yaïche, Ravi R. Mazumdar, and Catherine Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5) :667–677, 2000.
- [118] Roy D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7) :1341–1348, Sept 1995.
- [119] Kwan L. Yeung. Efficient time slot assignment algorithms for TDM hierarchical and nonhierarchical switching systems. *IEEE Trans. Communication*, COM-49 :351–359, February 2001.
- [120] James Yoh, Charles C. Wang, and Gary W. Goo. Survey of ka-band satellites for wideband communications. *IEEE Military Communications Conference*, 18 :120–125, 1999.
- [121] Dongmei Zhao, Xuemin Shen, and Jon W. Mark. Quality-of-service support by power and rate allocation in WCDMA systems. *Proc. IEEE GLOBECOM*, pages 604–608, Nov 2001.

# Table des figures

1.1	Petit réseau à 2 connexions. . . . .	4
1.2	Réseau cyclique. . . . .	6
1.3	Réseau linéaire . . . . .	9
1.4	Petit réseau . . . . .	11
1.5	Un autre petit exemple . . . . .	14
1.6	Exemple de réseau multicast en couche . . . . .	15
1.7	Allocation de la connexion 0 en fonction de $\alpha$ . . . . .	19
1.8	Utilités de Bill et Jack . . . . .	22
1.9	Normalisation des fonctions d'utilités . . . . .	23
1.10	Utilités de Bill et Jack "normalisées" . . . . .	23
1.11	Processus de naissance et de mort . . . . .	32
1.12	Fonction ON/OFF . . . . .	34
1.13	Application élastiques . . . . .	35
1.14	Application à débit variables . . . . .	35
1.15	Fonction d'utilité quadratique . . . . .	36
1.16	Fonctions d'utilités des 2 connexions concurrentes. . . . .	39
1.17	NBS de deux connexions se partageant un lien. . . . .	39
2.1	Les trois schémas d'optimisation. . . . .	59
3.1	Un réseau linéaire . . . . .	62
3.2	Un réseau grille. . . . .	67
4.1	Exemple simple de calcul du NBS. . . . .	79
4.2	Petit réseau à routage fixé. . . . .	81
4.3	Premier réseau . . . . .	82
4.4	Réseau COST : liens. . . . .	83
4.5	Allocation de bande passante dans les réseaux COST. . . . .	84
4.6	réseau COST : solution. . . . .	84
4.7	Allocation de bande passante pour le réseau COST network, $\alpha = 0.5$ . . . . .	86
4.8	Allocation de bande passante pour les réseaux COST, $\alpha = 5$ . . . . .	86
6.1	Un exemple de cellule avec 3 secteurs . . . . .	106
6.2	Un exemple de petit réseau sans fil avec 2 cellules . . . . .	107
6.3	Algorithme de contrôle d'admission. . . . .	109
7.1	Allocation de la bande passante pour $\alpha = 0$ . . . . .	134

7.2	Allocation de la bande passante pour $\alpha = 1$ .	134
8.1	Allure générale d'une ellipse.	142
8.2	Un switch TDM classique	149
8.3	Un exemple de système TDMA hiérarchique	153
8.4	SS/TDMA avec faisceau de taille variable	154
8.5	Exemple d'emprunte de satellite avec 14 spots, 4 couleurs et 3 zones par spots.	160
9.1	Deux allocations avec et sans ISCC.	165
9.2	Un algorithme d'acceptation des requêtes pour maximiser la somme de leur valeur.	168
9.3	Allocation équitable avec ISCC	171
9.4	Allocation équitable sans ISCC	172
9.5	Efficacité de l'allocation de la bande passante avec et sans l'ISCC	172
10.1	Comparaison pour une minute	177
10.2	Petit exemple sur 3 spots.	178
10.3	3 configurations d'émission possibles pour un seuil à 0.30.	179
10.4	Illustration d'une famille 6/7	180
10.5	Illustration d'une famille 5/7	181
10.6	Utilisation de terminaux de types différents.	182
10.7	Un exemple de résultat d'allocation de ressources.	183





# Publications

## Rapports de recherche :

- C. Touati, E. Altman, J. Galtier, "On Fairness in Bandwidth Allocations", INRIA Research report RR 4269, sept 2001.
- C. Touati, E. Altman, J. Galtier, "Fair Bandwidth Allocation between Providers in a Geostationary Satellite Network", INRIA Research Report RR 4421, march 2002.
- S. Alouf, J. Galtier, J.F. Lalande, C. Touati, "Un algorithme de programmation linéaire pour la résolution de ressources", INRIA Research Report (to appear).

## Articles :

- C. Touati, E. Altman, J. Galtier, "Fair power and transmission rate control in wireless networks", proceedings of IEEE Globecom 2002, November 2002, Taipei, Taiwan.
- C. Touati, E. Altman, J. Galtier, "Utility Based Fair Bandwidth Allocation", proceedings of the IASTED International Conference on Networks, Parallel and Distributed Processing and Applications (NPDPA 2002), October 2002, Tsukuba, Japan.
- C. Touati, E. Altman, J. Galtier, "Nash Bargaining Allocations with Applications to Mobile and Satellite Communications", Tenth International Symposium on Dynamic Games and Applications Workshop "Networking games & resource allocation", 12-15 July 2002, Petrozavodsk, Russia. An extended version is under revision in Computer Networks journal (Elsevier).
- C. Touati, E. Altman, J. Galtier, "Semi-Definite Programming Approach for Bandwidth Allocation and Routing in Networks", Tenth International Symposium on Dynamic Games and Applications Workshop "Networking games & resource allocation", 12-15 July 2002, Petrozavodsk, Russia. An extended version of the talk will be published in the special issue of the serial Game Theory and Applications, vol 9.
- C. Touati, E. Altman, J. Galtier, "Radio Planning in Multibeam Geostationary Satellite Networks", AIAA International Communication Satellite Systems Conference and Exhibit (ICSSC 2003), Yokohama, Japan.
- T. Toker, E. Altman, J. Galtier, C. Touati, I. Buret, B. Fabre, C. Guiraud, "Slot Allocation in a TDMA Satellite System : Simulated Annealing Approach", AIAA International Communication Satellite Systems Conference and Exhibit (ICSSC 2003), Yokohama, Japan.
- C. Touati, E. Altman, J. Galtier, "Semi-Definite Programming for Bandwidth Allocation in Communication Networks", to appear in the proceedings of ITC'18, September 2003, Berlin, Germany.
- E. Altman, B. Fourestie, J. Galtier, JM Kelif, C. Touati, "Stratégies d'allocations de débits et d'admission d'appels pour les mobiles d'un réseau CDMA ou W-CDMA (in a patent registration process)