

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 - INRIA Sophia-Antipolis

École doctorale : UNSA

Composante universitaire : UNIVERSITÉ NICE SOPHIA-ANTIPOLIS

Titre de la thèse :

Les principes d'équité appliqués aux réseaux de télécommunications

soutenue le 26 Septembre 2003 devant la commission d'examen

M. :	Philippe	NAIN	Président
MM. :	Ravi	MAZUMDAR	Rapporteurs
	Véronique	VÈQUE	
MM. :	Mustapha	BOUHTOU	Examineurs
	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 accordée 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 Kélib, 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 tous les membres du laboratoire OSDP pour leur accueil chaleureux et leur collaboration fructueuse. Toute ma gratitude se porte particulièrement vers Simona et Adrian Vasilache pour leurs conseils avisés et leur soutien pendant cette période de ma thèse.

Les travaux que nous avons effectués sur les réseaux satellitaires n'auraient pu être menés sans la collaboration de Benoit Fabre et d'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 implementé 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 accueillie. Je remercie tous les MISTRALIens et les 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.

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 qui n'a jamais cessé de me soutenir et de 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.

Pourquoi ?

Les opérateurs de téléphones portables (GSM) basent aujourd'hui 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 a été classé meilleur sur X critères". Ainsi, alors qu'on essayait autrefois d'attirer les clients par 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 commerciaux sont désormais souvent basés sur la qualité proposée. Une meilleure couverture, meilleure qualité de son, moins de blocage d'appels, ou l'accès à des services multimédia sont autant de critères de vente.

Considérons par exemple la couverture des réseaux. Le déploiement des différents réseaux 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 puis même dans les montagnes, où les coûts d'installation sont les plus grands et où les signaux radio se propagent mal. Au-delà d'un improbable élan philanthropique 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.

Il semblerait ainsi que la stratégie de marketing 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. On pourrait ajouter à celui-ci de nombreux autres exemples. D'une façon générale, plusieurs stratégies marketing sont possibles pour un opérateur. Soit il décide de se concentrer sur quelques utilisateurs, se trouvant ainsi pieds et poings lié à un petit nombre de clients, soit il tend à toucher un grand nombre d'utilisateurs potentiels, quitte à sacrifier un peu la rentabilité d'un système.

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

Des critères d'équité 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 peuvent être adaptés aux réseaux de télécommunications afin d'y proposer de nouveaux schémas d'allocations de ressources.

Table des matières

1	Introduction	1
1.1	Le contrôle de ressource	2
1.1.1	Pourquoi le contrôle de congestion est-il nécessaire?	2
1.1.1.1	Inefficacité d'un réseau	2
1.1.1.2	Effondrement de congestion	3
1.1.2	Conclusion	4
1.2	Définitions générales	5
1.2.1	Ressource	5
1.2.2	Connexion	5
1.2.3	Allocation faisable	6
1.2.4	Ordre sur les allocations	6
1.2.5	Efficacité	6
1.2.6	Notations	7
1.3	Allocations classiques	7
1.3.1	Maximisation du débit total	8
1.3.1.1	Définition	8
1.3.1.2	Exemple : le réseau linéaire	8
1.3.1.3	Discussion	9
1.3.2	Équité Max-Min	9
1.3.2.1	Définitions	9
1.3.2.2	Exemple du réseau linéaire	11
1.3.2.3	Algorithme de construction	12
1.3.2.4	Implémentation et discussion	13
1.3.3	Équité proportionnelle	15
1.3.3.1	Définition	15
1.3.3.2	Exemple : le réseau linéaire	15
1.3.3.3	Implémentation et discussion	16
1.3.4	Minimisation du délai potentiel	17
1.3.4.1	Définition	17
1.3.4.2	Application au réseau linéaire	17
1.3.5	Liens entre les équités	17
1.3.5.1	Définition	17
1.3.5.2	Exemple du réseau linéaire	18
1.3.6	Autres définitions d'équité	18

1.4	Critère de Nash	19
1.4.1	Quelques mots sur la théorie des jeux	19
1.4.1.1	Jeux non-coopératifs	19
1.4.1.2	Les jeux coopératifs	20
1.4.2	Petit voyage temporel	20
1.4.2.1	Fonctions d'utilité	20
1.4.2.2	Les axiomes	22
1.4.2.3	La résolution	22
1.4.3	L'évolution	23
1.4.4	Reformulation	24
1.4.5	Adaptation pour les réseaux	25
1.4.6	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
1.5.2.1	Algorithmes à taille de fenêtre fixe	29
1.5.2.2	Fenêtres de taille variable	30
1.5.2.3	Indications de congestion	32
1.5.3	Conclusion	34
1.6	Les fonctions d'utilité	34
1.6.1	Les fonctions ON/OFF	35
1.6.2	Applications élastiques	35
1.6.3	Applications à débit variable	36
1.6.4	Conclusion	36
1.7	Comparaison et propriétés des critères d'équité	38
1.7.1	Introduction	38
1.7.2	Impact de la concavité de la fonction d'utilité	38
1.8	La programmation semi-définie positive (SDP)	41
1.8.1	Petits rappels d'algèbre linéaire.	41
1.8.2	De la programmation linéaire à la SDP	44
1.8.2.1	La programmation linéaire	44
1.8.2.2	La programmation conique	45
1.8.2.3	La programmation conique-quadratique	47
1.8.2.4	La programmation Semi-Définie Positive	48
1.8.3	Exemples classiques	51
1.9	Conclusion	52

I Les réseaux terrestres 53

2 Introduction aux réseaux terrestres. 57

2.1	Système de contraintes	57
2.1.1	Notations	57
2.1.2	Demandes des utilisateurs	58

2.1.3	Types de routage	58
2.1.3.1	Le routage fixe	59
2.1.3.2	Le routage fractionnaire	59
2.2	Types d'optimisations	61
2.2.1	L'optimisation «connection-aware»	61
2.2.2	L'optimisation «network-aware»	61
2.2.3	Résumé	62
2.3	Conclusion	62
3	Méthodes analytiques	63
3.1	Exemples de calculs explicites	63
3.1.1	Les allocations équitables dans le réseau linéaire	63
3.1.1.1	Le réseau et les hypothèses	64
3.1.1.2	Solution du problème linéaire	65
3.1.1.3	Limites possibles	65
3.1.2	Analyse asymptotique	67
3.1.2.1	Conclusion	68
3.1.3	Réseau grille	69
3.2	Méthode Lagrangienne	70
3.2.1	Multiplicateurs de Lagrange	70
3.2.2	Problème dual	71
3.3	Conclusion	73
4	Méthode SDP	75
4.1	Philosophie générale de l'approche SDP	75
4.1.1	Contraintes linéaires	76
4.1.2	Relations entre les variables	77
4.2	Les différentes valeurs de α	78
4.2.1	Calculs d'équité pour $\alpha \neq 1$	78
4.2.2	Calcul du NBS (cas $\alpha = 1$)	80
4.3	Un exemple simple de calcul du NBS dans le cas du routage fixé	81
4.4	Expériences	83
4.4.1	Calcul du NBS dans un petit réseau à routage fixé	84
4.4.2	Le réseau COST	85
4.4.2.1	Optimisation «connection-aware» avec routage fixe.	86
4.4.2.2	Optimisation network-aware	87
4.5	Conclusion	89
II	Les réseaux mobiles	95
5	Précédents travaux sur le contrôle de taux dans les réseaux sans fil	99
5.1	L'équité dans les réseaux sans fil.	100
5.2	Choix des débits	101

5.3	Contrôle de puissance	102
5.4	Contrôle joint des débits et des puissances	104
5.5	Conclusion	105
6	Allocation jointe de débit et de puissance : cas général	107
6.1	Introduction	107
6.1.1	Pourquoi une allocation jointe entre débit et puissance?	107
6.1.2	Un réseau, différents liens.	108
6.1.3	La suite...	110
6.2	Intérêt technologique	110
6.2.1	Inconvénients des techniques actuelles	110
6.2.2	But des travaux	111
6.3	Le modèle	111
6.3.1	Allocation des débits	112
6.3.2	Modèle de puissance	113
6.3.3	Lien entre puissance et débit	113
6.3.4	Expression analytique du SIR	114
6.3.5	Résumé : formulation du problème d'équité	115
6.4	Propriétés du système	115
6.4.1	Problème équivalent et concavité de la fonction objectif	116
6.4.1.1	Problème équivalent.	116
6.4.1.2	Concavité de la fonction objectif	116
6.4.2	Propriétés des vecteurs de taux acceptables	117
6.4.2.1	Vecteur de puissance acceptable minimal	117
6.4.2.2	Conditions nécessaires et suffisantes sur les vecteurs de débits et puissances minimales associées	120
6.5	Conclusion	121
7	Application du modèle aux différents liens et approximations possibles	123
7.1	Cas du lien montant	123
7.1.1	Notations et hypothèses.	124
7.1.2	Application des résultats du modèle	125
7.2	Cas du lien descendant	126
7.2.1	Notations et hypothèses	126
7.2.1.1	Puissance d'émission	127
7.2.1.2	Gain de transmission	127
7.2.1.3	Première expression du rapport signal sur bruit	127
7.2.1.4	Valeurs des interférences	128
7.2.1.5	Hypothèses	129
7.2.2	Application des résultats du modèle	129
7.3	Macro-diversité dans le lien descendant	131
7.3.0.1	Notations et hypothèses	132
7.3.1	Application des résultats du modèle	133
7.4	Approximations possibles	134

7.4.1	Premier schéma : approximation de la fonction objectif	134
7.4.2	Second schéma : approximation de l'ensemble des contraintes	135
7.4.3	Conclusion	136
7.5	Quelques résultats numériques et conclusion	136
III Les réseaux satellites		141
8	Introduction aux réseaux satellitaires.	145
8.1	Qu'est-ce qu'un réseau satellitaire?	145
8.1.1	Orbites et emprunts.	146
8.1.2	Altitudes possibles	147
8.1.3	Bandes de fréquence utilisées.	148
8.1.4	Spécificités des systèmes GEO	149
8.1.5	Bilans de liaison	150
8.1.6	Méthodes d'accès	151
8.1.6.1	Systèmes FDMA	152
8.1.6.2	Systèmes TDMA	152
8.1.6.3	Systèmes CDMA	152
8.2	Précédents travaux sur l'allocation de slots dans les systèmes TDMA et MF-TDMA	153
8.2.1	Le problème TSA (Time Slot Allocation)	153
8.2.1.1	Expression matricielle	154
8.2.1.2	Premières méthodes de résolution	155
8.2.1.3	Variantes du problème	156
8.2.1.4	TDM hiérarchiques	158
8.2.1.5	Systèmes à faisceaux de bandes passantes différentes	159
8.2.2	Allocation des rafales	160
8.2.3	Problèmes annexes	161
8.2.4	Descriptions de systèmes complets	162
8.2.4.1	Études générales des systèmes satellitaires.	162
8.2.4.2	Description globale de systèmes	162
8.3	Conclusion et choix du système considéré	165
8.3.1	Système physique	165
8.3.1.1	Spots et réutilisation spatiale	165
8.3.1.2	Zones	166
8.3.2	Formulation du problème et synthèse	167
9	Première partie de l'allocation : la planification radio.	169
9.1	Introduction et notations	169
9.1.1	Définitions générales	170
9.1.2	Inter-Spot Compatibility Conditions (ISCC)	170
9.1.3	Notations (résumé)	171
9.1.4	Critères d'équité	172

9.1.5	Formulation du problème	173
9.2	Algorithme d'allocation proposé	173
9.2.1	Allocation des porteuses aux opérateurs dans un spot à J fixé	173
9.2.2	Obtention du vecteur J minimal	175
9.2.3	Étapes de résolution du problème d'allocation	175
9.2.3.1	Maximisation des bénéficiaires du surplus (pour $\alpha \geq 1$).	176
9.2.3.2	Redistribution équitable des valeurs (toute valeur de α).	176
9.3	Expérimentations	177
9.3.1	Valeurs des paramètres	177
9.3.1.1	Types et utilités des porteuses	177
9.3.1.2	Génération des demandes des opérateurs	177
9.3.2	Résultats numériques	178
9.4	Conclusion	181
10	Deuxième partie de l'allocation : l'allocation des slots.	183
10.1	Le modèle d'interférence.	183
10.2	Le recuit simulé	184
10.2.1	Méthodes heuristiques et recuit simulé	184
10.2.2	Implémentation du recuit simulé	185
10.2.3	Améliorations de l'algorithme	186
10.2.4	Résultats	186
10.2.4.1	Conclusion	187
10.3	Résolution par programmation linéaire	188
10.3.1	Présentation du problème	188
10.3.1.1	Notations	188
10.3.1.2	Modèle d'interférence	188
10.3.2	Définition des familles et notions associées	189
10.3.2.1	Qu'est-ce qu'une famille?	189
10.3.2.2	Définition des super-zones.	191
10.3.2.3	Familles de spots génériques	191
10.3.2.4	Statut des spots	192
10.3.3	Choix et placement des familles.	193
10.3.3.1	Résolution des contraintes de placement sur le canal radio	193
10.3.3.2	Typage des familles.	193
10.3.4	Résultats numériques et conclusion	194
10.4	Conclusion	196
	Conclusion	202
	IV Annexes	203
	A Programme de création des matrices d'entrées.	205
	B Les réseaux satellitaires.	209

B.1	Relation d'ordre sur les allocations	209
B.2	Optimalité de l'algorithme	210
C	Bibliographie	213
D	Publications	223

Chapitre 1

Introduction

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

L'intérêt est 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 besoins des applications temps-réels. Nous désirons d'autre part 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 explique de quelle façon nous proposons d'adapter et d'utiliser les critères d'équités déjà existants dans les réseaux de télécommunications. Nous commençons notre étude en montrant la nécessité d'instaurer des contrôles de ressources dans les télécommunications. Puis, après avoir introduit les notations et définitions utilisées tout au long de la thèse (section 1.2), nous présentons 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 présentons 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 montrons ensuite comment ces fonctions d'utilités peuvent être choisies dans le cas des réseaux de télécommunications (section 1.6), puis expliquons 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 commentons brièvement l'impact de ces fonctions d'utilité sur les allocations obtenues (section 1.7). Nous décrivons enfin une nouvelle méthode de résolution, la programmation Semi-Définie Positive (SDP) (section 1.8), 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 de ressources dans les réseaux.

1.1 Le contrôle de ressource

D'une façon générale, le problème d'allocation de ressources 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 son courrier électronique n'a pas les mêmes besoins qu'une autre utilisant une application audio ou vidéo).

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

On peut très bien imaginer des systèmes dans lesquels chaque source peut émettre en utilisant une quantité quelconque de ressource. 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 considère un petit exemple, inspiré du tutoriel de Le Boudec [20]. Sur le réseau filaire représenté en figure 1.1, deux sources sont en compétition pour émettre. A chaque source S_i correspond un destinataire D_i (noté «Dest» sur le schéma). J'appelle *connexion* une paire source-destination. Les deux connexions du réseau sont alors notées $(S1, D1)$ et $(S2, D2)$. Le média utilisé dans ce réseau est le câble. On suppose que les câbles ont des capacités différentes, conformément à celles indiquées sur la figure.

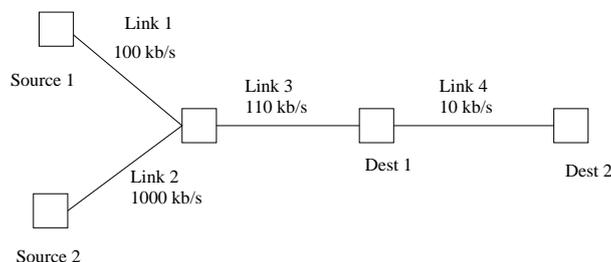


FIG. 1.1 – Petit réseau à 2 connexions.

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

Pour clarifier cela, considérons l'analogie suivante. Les données circulant sur un réseau sont similaires à un fluide. Les liens sont alors des canalisations et les files d'attente, des réservoirs. Enfin l'ordonnanceur est la vanne entre le réservoir et le canal.

Une politique classique d'ordonnancement est appelée **FIFO** (First-In First-Out). Elle signifie que les paquets sont envoyés sur les liens dans leur ordre d'arrivée dans les buffers. Dans la version imagée, la politique FIFO correspond à une vanne située au 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 (représenté en 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 kb/s et $S2$ aura un débit de 1000 kb/s. Nous remarquons que la capacité du lien 3 est inférieure à la somme des débits des sources. Le lien sera donc saturé et les allocations des connexions ($S1, D1$) et ($S2, D2$) seront respectivement 10 kb/s et 100 kb/s. De même, le lien 4 est saturé car sa capacité est inférieure à 100 kb/s donc l'allocation à la connexion ($S2, D2$) sur le lien 4 sera de 10 kb/s. Ainsi, les allocations globales sont de 10 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 allouer un débit global de 10 kb/s alors que le réseau lui permettrait d'obtenir une allocation dix fois plus grande.

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

1.1.1.2 Effondrement de congestion

Plus dramatiques encore sont les situations d'effondrement de congestion :

Définition 1.1. *L'effondrement de congestion est une situation dans laquelle 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 pour elle 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, la situation inverse se produit, et alors que la source cherche à améliorer son débit, elle augmente la congestion globale du système et voit sa propre allocation décroître. On entre alors dans une situation en spirale. L'exemple suivant illustre une telle situation sur un petit exemple.

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

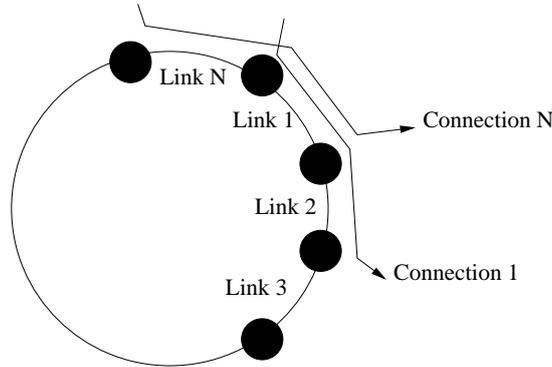


FIG. 1.2 – Réseau cyclique.

On peut alors montrer [20] que pour de grandes valeurs de λ , l'allocation de la connexion n au niveau du deuxième lien qu'elle traverse est :

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

Nous sommes ainsi en présence d'une situation paradoxale dans laquelle plus les demandes des connexions augmentent (i.e. plus λ croît), plus le débit que chaque connexion obtient est faible. Le réseau est alors de moins en moins efficacement 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.

D'autres scénarios menant à des situations d'effondrement sont présentées dans [38].

1.1.2 Conclusion

Dans les premières années 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 [96], essentiellement dû aux connexions qui retransmettaient inutilement des paquets qui étaient en transit ou avaient déjà été reçus par leurs destinataires.

Les contrôles de congestion ont pour but d'éviter ce genre de scénario catastrophique en limitant aux sources leur taux de transmission en fonction de l'état du réseau. Les solutions classiques sont les protocoles dits de *bout-en-bout* qui ont une vue globale de la route suivie par 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 (reposant sur ATM ou 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 diffèrent d'un protocole à l'autre, ce qui induit donc des différences dans les allocations qu'ils engendrent. 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 enfin que l'introduction de notions d'équité dans les réseaux est de plus en plus prise 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 énonçons 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

Là où les ressources qui sont partagées entre les utilisateurs peuvent être de natures différentes. Nous utilisons le terme de «ressource» dans toute cette partie afin de garder la généralité des formulations d'équité. En le remplaçant par «bande passante» on obtient la formulation correspondant aux réseaux terrestres. Dans le cas des réseaux mobiles où la ressource partagée est un canal radio, on affecte 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 multimédia. 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 différents types 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 satellitaires, etc.
- Les contraintes des terminaux : dans les réseaux mobiles par exemple à chaque mobile est associée une puissance d'émission maximale. Dans les réseaux satellitaires, chaque terminal ne peut utiliser qu'un nombre fini de types de porteuses.
- Les contraintes de l'utilisateur : comme nous le verrons dans la suite, à chaque utilisateur est associée une demande minimale et maximale.

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

1.2.4 Ordre sur les allocations

Nous définissons ici des relations d'ordre entre les allocations afin de pouvoir par la suite les comparer, notamment en terme d'équité et d'efficacité. Soit \vec{x} une allocation, c'est-à-dire un vecteur qui associe à chaque connexion une quantité de ressource. Supposons qu'il existe une relation d'ordre sur les éléments de \vec{x} . On peut alors définir plusieurs relations d'ordre sur les allocations [16] et montrer aisément qu'elles sont réflexives, anti-symétriques et transitives. On note :

$$\begin{aligned} \vec{x} \leq \vec{y} &\Leftrightarrow \forall n \in 1, \dots, N, x_n \leq y_n \\ \vec{x} \leq_{lg} \vec{y} &\Leftrightarrow \begin{cases} \text{Soit } \exists n, n > 0, \forall j \in 1, \dots, n-1, x_j = y_j \text{ et } x_n < y_n, \\ \text{Soit } \vec{x} = \vec{y} \text{ (ordre lexicographique)} \end{cases} \\ \vec{x} \preceq \vec{y} &\Leftrightarrow \sigma(\vec{x}) \leq_{lg} \sigma(\vec{y}) \text{ avec } \sigma(\vec{x}) \text{ et } \sigma(\vec{y}) \text{ les versions ordonnées de } \vec{x} \text{ et de } \vec{y}^1. \end{aligned}$$

1.2.5 Efficacité

L'efficacité (ou l'optimalité) 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 Pareto-optimaux forme un sous ensemble de dimension $N - 1$. On peut comprendre qu'un opérateur veuille

¹ $\sigma(\vec{x})$ est la version ordonnée de \vec{x} si pour tout n , $\sigma(x)_n$ est le n -ième plus petit élément de \vec{x} .

profiter d'une utilisation optimale du réseau. Le problème auquel nous nous intéressons est de savoir quelle allocation choisir dans cet hyper-ensemble.

1.2.6 Notations

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

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

Nous présentons dans la section suivante les allocations classiques rencontrées dans les réseaux. Nous illustrons chacune d'elle sur un exemple simple tiré d'une situation de réseau terrestre. Bien évidemment, ces critères ne sont cependant pas limités à ces types de réseaux.

1.3 Allocations classiques

Nous décrivons dans cette section les optimisations de partage de ressources dans les réseaux rencontrés fréquemment dans la littérature. Une brève description des allocations correspondantes se trouve dans [86].

Chaque formulation d'optimisation 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 : si la connexion n a un poids K , $K \in \mathbb{N}$, $K > 1$, elle peut être remplacée par K connexions n_1, \dots, n_K de poids unitaire utilisant les mêmes ressources que la connexion n de sorte que $x_n = \sum_{k=1}^K x_{n_k}$. 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.

Nous illustrons chaque allocation présentée 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 l'indice des connexions.

Le réseau se compose de N liens sur lequel circulent $N + 1$ connexions, numérotées comme suit :

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

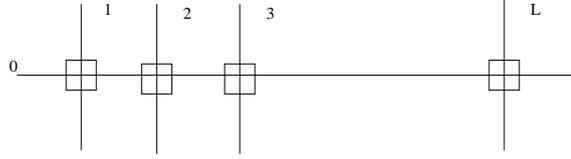


FIG. 1.3 – Réseau linéaire

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 pour des raisons évidentes : dans le cas où la tarification se fait proportionnellement au débit de la connexion cette optimisation maximise le revenu de l'opérateur du réseau.

Cependant, un gros inconvénient de ce critère est qu'il peut conduire à des allocations pour lesquelles plusieurs connexions se voient refuser l'accès au réseau, comme le montre l'exemple suivant. Pour cette raison il n'est pas considéré comme équitable.

1.3.1.2 Exemple : le réseau linéaire

Les contraintes du système associé à la Fig. 1.3 sont les contraintes de capacité :

$$\forall l \in \llbracket 1, N \rrbracket, \quad x_0 + x_l \leq c_l \quad (\text{en notant } c_l \text{ la capacité du lien } l).$$

Le revenu global du système s'écrit alors :

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

On voit aisément que pour cet exemple, on peut (en l'absence de contraintes supplémentaires) saturer chaque lien et obtenir : $\forall l \in \llbracket 1; N \rrbracket, x_0 + x_l = c_l$. Le revenu total est alors

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

Ainsi, dès lors que $N > 1$, la maximisation du revenu total implique que x_0 soit minimal, c'est-à-dire $x_0 = 0$.

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

1.3.1.3 Discussion

On trouve fréquemment dans la littérature une variante de ce critère, sous le nom de *social welfare*. 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 vidéo. Les algorithmes proposés visent alors la maximisation de la somme des utilités des utilisateurs. Par exemple, [60] propose un tel algorithme.

Un exemple d'extension possible pour l'allocation à des sessions multicast est donné dans [61]. 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 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 ne soit pas définie de manière unique, comme on peut le remarquer dans l'exemple de la figure 1.3 de [20], et sur l'exemple de la Fig. 1.4 du présent manuscrit.

A l'opposé de ce critère de maximisation globale, on trouve l'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 au point max-min est évidemment inférieure à celle que l'on vient d'étudier. On voit donc dès à présent que la recherche d'une allocation correspond bien souvent au choix d'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 nous décrivons 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 [20].

Définition intuitive

Intuitivement, l'allocation max-min est la plus égalitaire possible dans la mesure où c'est celle dont les valeurs des allocations aux différents utilisateurs 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 le traversant 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_n le poids d'une connexion n , alors l'allocation au lien congestionné sera de $c_l w(n) / \sum_{j \in n_l} w(j)$.)

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 sur 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 partagée également entre les autres.

Pour clarifier cela, considérons l'exemple du petit réseau représenté en Fig. 1.4. 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 2 et 4 peuvent également utiliser 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 égalitaire possible étant données les contraintes du système. Elle induit un revenu de $5c/3$.

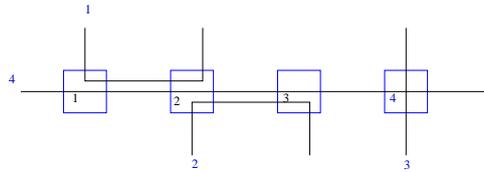


FIG. 1.4 – Petit réseau

L'optimisation globale aurait donné :

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

D'où un revenu total de $2c$. Remarquons que dans cet exemple, l'allocation globalement optimale 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 si :

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

On a alors la proposition suivante (dont la preuve est donnée dans [20]) :

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 l'exemple précédent, pour l'allocation max-min, 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 également remarquer 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 noter un fait intéressant : dans ce réseau, contrairement au réseau linéaire, il n'est pas possible de saturer simultanément tous les liens. On peut alors interpréter 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. Appelons la \vec{x}_{Mm} . C'est un vecteur de taille N . J'appelle «plus petit bénéficiaire» la connexion dont l'allocation est la plus faible. Dans l'exemple de la figure 1.4, pour l'allocation max-min les plus petits 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 \vec{x}'_{Mm} , de dimension $N - 1$ est également max-min dans l'ensemble des allocations de taille $N - 1$ (associées au mêmes contraintes du système), 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, l'allocation max-min est maximale pour la relation d'ordre \preceq (telle que défini à la section 1.2.4, 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 n se fasse au détriment de celle d'une connexion m qui était déjà défavorisée par rapport à la première connexion (i.e. $x_m \leq x_n$).

De façon plus formelle, si \vec{x} est l'allocation max-min et \vec{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 :

$$\forall n \in \llbracket 0, N \rrbracket, x_n = c_n - x_0.$$

Soit maintenant l_{\min} l'indice du lien de plus petite capacité. Pour un x_0 donné, c'est donc l'indice du x_l minimal. L'équité max-min implique alors que $x_0 = x_{l_{\min}} = c_{l_{\min}} - x_0$. Ainsi, $x_0 = C_{l_{\min}}/2$ et 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

L'équité Max-min peut être obtenue par la procédure de remplissage suivante [13, 20] :

1. A l'instant 0, toutes les connexions ont une bande passante allouée nulle.
2. On augmente les allocations de toutes les connexions avec le même taux
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 le traversant, mais on continue la procédure de remplissage pour les connexions n'utilisant que des liens non saturés.

On peut aisément comprendre que cette procédure de remplissage tend vers l'allocation max-min. Cependant, sa mise en pratique n'est pas sans inconvénients. Supposons qu'une fois la procédure effectuée, une nouvelle connexion arrive sur le réseau. Il ne suffit pas d'informer les seules 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 $A \rightsquigarrow B$ si elles ont au moins un lien en commun. La relation \rightsquigarrow 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 :

- $A \rightsquigarrow A_1$
- $\forall i \in \llbracket 1, k-1 \rrbracket, A_i \rightsquigarrow A_{i+1}$
- $A_k \rightsquigarrow B$.

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 $A \rightleftharpoons B$

Clarifions les choses sur un nouvel exemple. Dans la figure 1.5, supposons que la connexion 4 (en bleu) se joint au réseau. Elle partage avec la connexion 1 un lien en commun, et elles sont donc 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.

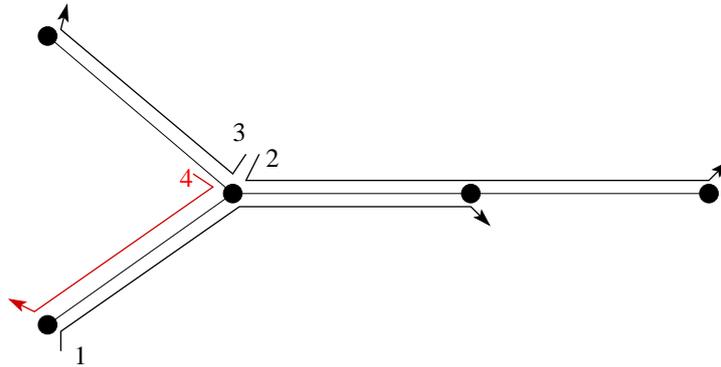


FIG. 1.5 – Limites de l'algorithme de remplissage

En revanche, cet 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 l'allocation 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 [41]. 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.

Une étude intéressante dans le cas du multicast est présentée dans [113]. En effet, adapter les définition d'équité max-min à de tels systèmes 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 moins performante dans la mesure où un utilisateur disposant de moins de bande passante va pénaliser l'ensemble des receveurs. Cette version provoque également le non-remplissage (non-saturation) de liens. Dans cette version, la définition de l'équité max-min est la même que précédemment en les transposant les propriétés du niveau de « connexion » à celui de « session »². En particulier, une connexion ne possède plus alors nécessairement un goulot d'étranglement. En revanche, cela devient vrai au niveau des sessions. 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.

²Une session multicast regroupe l'ensemble des connexions entre la source et les multiples récepteurs.

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. Le multicast en couches repose sur une architecture en arbre, dont la racine est la source et chaque receveur constitue une feuille. Chaque receveur peut alors joindre ou se séparer d'un nœud de l'arbre au cours de la transmission, comme illustré dans la figure 1.6. Cependant des résultats surprenants 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 [113].

De plus, de gros problèmes se posent dans la pratique. Dans les implémentations actuelles, 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. Au lieu de cela, 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. Le problème survient lorsqu'un utilisateur se déconnecte. Supposons que

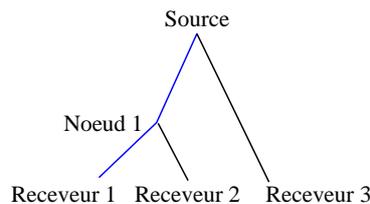


FIG. 1.6 – Réseau multicast en couche

sur l'exemple de la figure 1.6, 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 l'intérêt de la méthode, par rapport à l'approche unicast, pour laquelle cette allocation aurait été de $x_1 + x_2$. Le problème survient 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, ce que les auteurs de [113] appellent la redondance. Ils étudient alors l'impact de la redondance sur l'équité et montrent 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 directement généralisables aux autres utilités.

Bien qu'il soit Pareto optimal, il a été reproché à l'équité max-min de trop favoriser les connexions coûteuses (en termes de nombre de liens) et de ne pas utiliser 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, avec l'allocation max-min, la moitié de la bande passante

disponible, soit autant qu'une «petite» connexion. Une compromis entre l'optimisation globale et l'équité max-min a donc été introduite : l'équité proportionnelle.

1.3.3 Équité 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 [20].

Première définition

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

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

On peut montrer que dans le cas d'un nombre fini de connexions et de liens, ce critère mène à une solution unique [20].

Deuxième définition

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

$$\sum_{n \in \mathcal{N}} \frac{y_n - x_n}{x_n} \leq 0$$

1.3.3.2 Exemple : le réseau linéaire

Reprenons la première définition de l'équité proportionnelle. Comme dans les cas précédents, on peut saturer les liens, le critère étant Pareto optimal, et s'écrit :

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

On obtient immédiatement en dérivant la formule :

$$x_0 = \left(\sum_{n \in \mathcal{N}} \frac{1}{c_n - x_0} \right)^{-1}.$$

Dans le cas de liens identiques de capacité c , 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 d'obtenir une allocation pour x_0 qui soit strictement inférieure aux autres connexions. De plus, $x_0 \rightarrow_{\mathcal{L} \rightarrow \infty} 0$.

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

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

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

On remarque alors que le revenu total du réseau vaut, pour l'allocation équitale proportionnelle : $R_{prop} = (1 + N^2) \frac{C}{N + 1}$. On remarque que $R_{opt\ global} \geq R_{\text{éq prop}} \geq R_{\text{max-min}}$. Ce résultat se généralise : l'allocation proportionnelle équitale 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 proportionnelle. Une allocation est dite équitale proportionnellement pondérée et notée $\vec{x} \in \mathcal{X}^{wpf}$, si pour toute autre allocation $\vec{y} \in \mathcal{X}$, la somme des changements proportionnels pondérée par la famille w_i est négative ou nulle ([64, 65] (voir également [16, 85, 94]) : $\sum_{n \in \mathcal{N}} w_n \frac{y_n - x_n}{x_n} \leq 0$.

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

$$x^{wpf} = \max_{\vec{x} \in \mathcal{X}} \sum_{x \in \mathcal{N}} w_n \ln(x_n) \quad \text{ou} \quad x^{wpf} = \max_{\vec{x} \in \mathcal{X}} \prod_{n \in \mathcal{N}} x_n^{w_n}. \quad (1.1)$$

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 équitales sous certaines conditions [65]. 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 [86], est :

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

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 l'on a : $\forall n \in \llbracket 0; N \rrbracket, x_n = c_n - x_0$. Le critère s'écrit alors :

$$\min \frac{1}{x_0} + \sum_{n \in \mathcal{N}} \frac{1}{c_n - x_0}.$$

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 montré [94] que tous ces critères pouvaient être écrits sous la forme d'une seule fonction d'optimisation. Considérons l'allocation \vec{x}_α solution de :

$$\vec{x}_\alpha = \max_{\vec{x} \in \mathcal{X}} \sum_{n \in \mathcal{N}} \frac{x_n^\alpha}{1 - \alpha} \text{ avec } \alpha \geq 0, \alpha \neq 1. \quad (1.2)$$

Alors, pour des valeurs particulières de α , \vec{x}_α correspond à des allocations connues :

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

Dans [20], Le Boudec présente une preuve élégante du cas max-min, en utilisant une famille plus générale que celle de notre formulation. Dans [16], les auteurs s'intéressent au cas d'un système où le processus d'arrivée et de départ des connexions est dynamique et 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

Observons la valeur de n_0 lorsque α croît de 0 à l'infini : elle croît continûment de 0 à $c/2$, conformément à la représentation de la figure 1.7

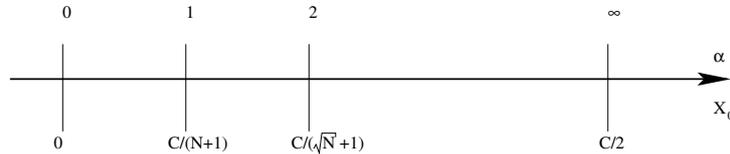


FIG. 1.7 – Allocation de la connexion 0 en fonction de α

On voit bien ici l'influence du paramètre α . Plus α 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 définie comme étant la solution d'une maximisation de fonction H [20] telle que :

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

avec :

- u une fonction d'utilité liée 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 vu par l'administrateur du réseau, et permet de prendre en compte 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.

Notons que le concept d'équité de Nash a déjà été utilisé (sous une forme simplifiée) dans des réseaux de topologie générale pour permettre le partage équitable de ressources [128, 87]. 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 Quelques 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 [83, 104, 112, 121].

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 d'agents, appelés joueurs participent à un jeu, à prendre ici dans un sens large de situations d'interactions entre des acteurs aux intérêts conflictuels. Chacun peut entreprendre un certain nombre d'actions et l'on s'intéresse aux éventuels équilibres qui sont issus de leurs comportements. Enfin, 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 choisit la stratégie maximisant sa propre utilité, indépendamment des intérêts des autres joueurs. Un exemple de telle situation est celle de la concurrence entre vendeurs. On suppose que chaque client peut acheter des biens à des vendeurs indépendants. Chaque marchand est libre de modifier ses prix au fil du temps, de façon à maximiser son revenu, en fonction de ses coûts de production et des prix pratiqués par les autres vendeurs. Dans ce genre de jeux, les équilibres éventuellement obtenus sont appelés *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 :*

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

Une perturbation désigne un état obtenu par le changement de stratégie d'un seul des joueurs. Soit \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 $\forall i, \forall Y, (Y_i, X^{[-i]}) \in \widehat{X}, u_i(Y_i, X^{[-i]}) \leq u_i(X)$.

Un équilibre de Nash est malheureusement 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, l'optimisation «égoïste», ou non-coopérative, consiste à avancer au maximum une fois le feu passé au vert. Ce faisant, chacun minimise son temps de passage. Cependant, chaque conducteur qui restera bloqué au milieu de l'intersection empêchera à son tour le passage des automobilistes circulant dans l'autre direction au prochain changement de feu, provoquant des inter-blocages. 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 complexe que dans le cas de N voitures, si $N - 1$ conducteurs adoptent ce comportement peu courtois, le N ème 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* (*marchandeurs*). Le résultat d'un marchandage porte le nom d'*équité*.

C'est à eux que nous nous intéresserons 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. Nous restreignons notre présente étude sur le critère d'équité le plus célèbre, celui de Nash.

1.4.2 Petit voyage temporel

Le texte fondateur du NBS est [97]. 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'utilité

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, $u(\alpha) > u(\beta)$ signifie que l'objet α est d'avantage désiré que l'objet désigné par y . Aussi, la fonction d'utilité, telle que désignée par Nash est linéaire : si $0 \leq p \leq 1$, $u[p\alpha + (1 - p)\beta] = pu(\alpha) + u(\beta)$. 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 correspondant 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, c'est-à-dire les issues possibles du marchandage.

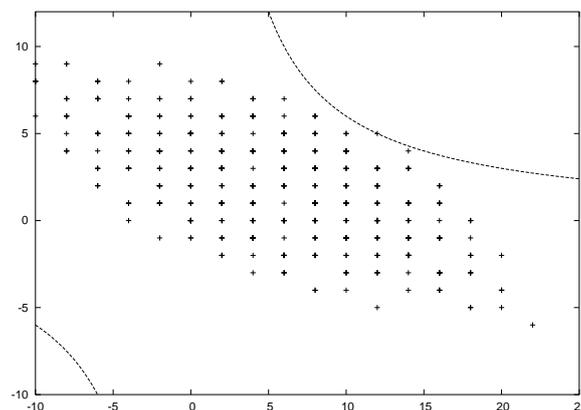


FIG. 1.8 – Utilités de Bill et Jack

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

Soit $c(S)$ le point solution du marchandage. Selon Nash, pour être équitable, il doit vérifier les conditions suivantes :

(Axiome 1) Si α 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. Notons que cette condition correspond à l'optimalité de Pareto appliquées aux utilités des joueurs.

La deuxième condition énoncée par Nash est la suivante :

(Axiome 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 :

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

Cette condition, connue sous le nom de «indépendance aux possibilités non pertinentes» est plus difficile à comprendre. Soit T l'ensemble des marchandages possibles. Si deux individus sont 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 $c(t)$. En d'autres termes, l'équilibre ne dépend pas de solutions potentielles qui seront plus tard rejetées.

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 côtés et qui contient le nuage de points et lui appliquer les hypothèses (Cond 2) et (Cond 3). Le point solution sera alors de coordonnées $(a/2, a/2)$. Cette méthode est illustrée par la figure 1.9. Dans l'exemple de Bill et Jack, le nuage de points a (après transformation) a l'allure de la figure 1.10.

On peut montrer [97] que cette approche correspond en fait à maximiser le produit $u_1 u_2$ 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 l'exemple considéré

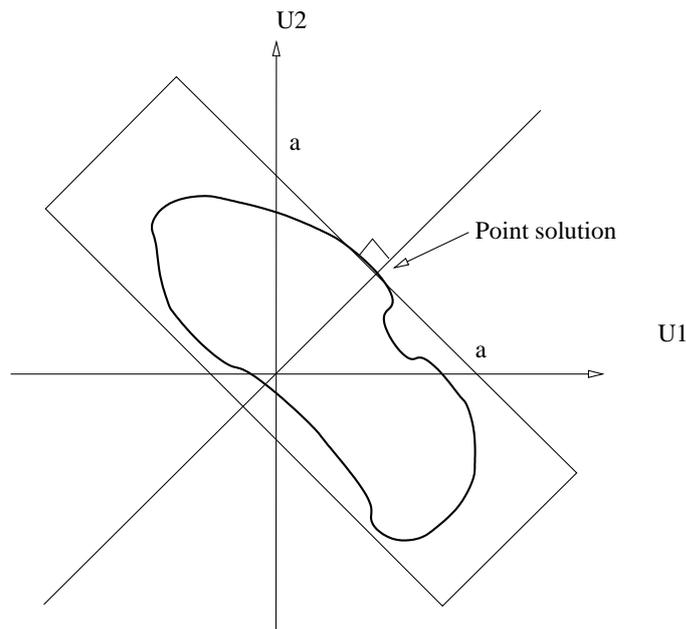


FIG. 1.9 – Normalisation des fonctions d'utilités

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 cède son stylo, son jouet et son couteau.

1.4.3 L'évolution

La formulation de Nash a été légèrement étendue 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, soit :

- 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, à un couple (S, d) associe l'utilité du joueur i : $g_i : (S, d) \mapsto u_i$.

Si je note S' et d' les images de S et d par une transformation affine ϕ ³, alors l'image par toute fonction g_i de (S', d') sera l'image par la même transformation affine de $g_i(S, d)$: $g_i(\phi(S), \phi(d)) = \phi(g_i(S, d))$. 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 $\max_{u_i, i \in \{1, 2\}} \prod_i (u_i - d_i), \forall i \in \{1, 2\}, u_i \geq d_i$.

³c'est-à-dire que j'introduis deux paires (a_1, b_1) et (a_2, b_2) telles que $S' = \{(a_i u_i + b_i)_{i \in \{1, 2\}} : (u_i)_{1 \leq i \leq 2} \in S\}$ et $d' = (a_i d_i + b_i)_{1 \leq i \leq 2}$.

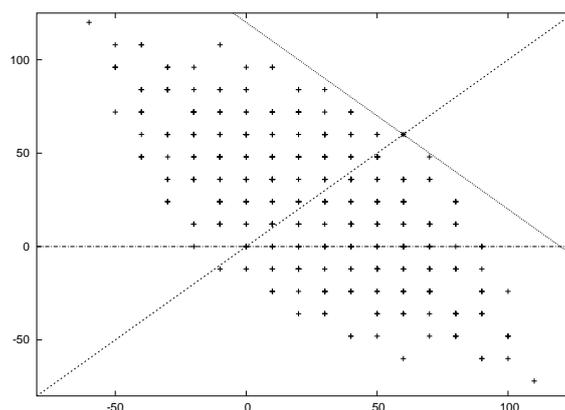


FIG. 1.10 – Utilités de Bill et Jack «normalisées»

En 1999, Muthoo reprend dans le premier chapitre de son livre [95] le concept de Nash. Cet économiste s'interroge sur le bien fondé des axiomes et présente deux exemples, le premier relevant du contrôle du crime, le second 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 (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, 95].

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 ([116]). Ils ont donné au problème sa forme actuellement utilisée.

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

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

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 $U = \{u \in \mathbb{R}^N, \exists x \in X, u \leq f(x)\}$ on obtient la formulation équivalente d'optimisation (théorème 1.1).

1.4.5 Adaptation pour les réseaux

En 1991, [87] 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 [128] une adaptation des formules de Ştefănescu pour l'allocation de bande passante dans les réseaux. Leur idée est la suivante. Soit :

- N 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 ($Y \subset \mathbb{R}^N$).
- f_n la fonction d'utilité de l'utilisateur n
- y_n^0 l'utilité minimale requise par la connexion i et MR_n l'allocation associée :

$$y_n^0 = f_n(MR_n).$$

- $\mathcal{G} = \{(Y, y^0), Y \subset \mathbb{R}^N\}$ est l'ensemble des solutions éventuelles.

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 : $S(Y, y^0) \in Y^0 := \{y \in Y \mid y \geq y^0\}$.*
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 $(\phi(y), \phi(y^0)) = \phi(S(Y, y^0))$*
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 trouvée sur un domaine plus restrictif. Plus précisément, si $V \subset U$, $(V, u^0) \in \mathcal{G}$ and $S(U, u^0) \in V$ alors $S(U, u^0) = S(V, u^0)$.*

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

Soit $\mathcal{X}_0 := \{x \in \mathcal{X} \mid f(x) \geq y^0\}$. Nous avons alors le théorème suivant :

Théorème 1.1. [128, Thm. 2.1, Thm 2.2] et [116]. Supposons les fonctions d'utilité f_n 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 $J = \{j \in \{1, \dots, N\} \mid \exists x \in X, \text{ s.t. } f_j(x) > y_j^0\}$. Supposons de plus que les fonctions $\{f_j\}_{j \in J}$ sont injectives. Alors, il existe une unique NBP ainsi qu'une unique NBS x qui vérifie $f_j(x) > y_j^0, j \in J$, et c'est l'unique solution du problème P_J :

$$(P_J) \quad \max \prod_{j \in J} (f_j(x) - y_j^0), \quad x \in \mathcal{X}_0. \quad (1.3)$$

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

$$(P'_J) \quad \max \sum_{j \in J} \ln(f_j(x) - y_j^0), \quad x \in \mathcal{X}_0. \quad (1.4)$$

Remarque

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 Résolution de Yaiche, Mazumdar et Rosenberg

Dans [87], les auteurs s'intéressent au cas où les fonctions f sont égales à la fonction identité et proposent des solutions centralisées et décentralisées pour résoudre le NBS.

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? Le 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 :

$$(C) \quad \begin{cases} \max_{x_i} \ln(x_i - MR_i) - \alpha_i x_i \\ x_i \geq MR_i \\ x_i \leq PR_i \end{cases} \quad (R) \quad \begin{cases} \max_x \sum_i \alpha_i x_i \\ x_i \geq MR_i \\ x_i \leq PR_i \\ (Ax)_l \leq C_l \end{cases} \quad (1.5)$$

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 [65] dans le cas où le but était de maximiser la somme des utilités des utilisateurs du réseau. Ainsi, c'est l'introduction des prix α_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 [29] les auteurs présentent un algorithme qui converge en un nombre fini d'itérations vers l'exakte allocation max-min équitable. Le principe est le suivant : 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 ,
- Γ_l l'ensemble des connexions dont le goulot d'étranglement est autre que l ,
- g_l le nombre de connexions de Γ_l qui utilisent le lien l .

Nous avons précédemment vu qu'une caractérisation d'une allocation max-min était que chaque source 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 Γ_l .

D'autres algorithmes de calculs explicites des taux, étudiés dans le contexte de l'ABR, ont été repris dans [8] et [52]. La principale difficulté est de trouver un compromis entre équité, stabilité, robustesse, vitesse de convergence et utilisation des liens. De plus, 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 développés 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 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'attentes appelés AQM (Active Queue Management) tels que DropTail et Red [81].

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 source devra patienter d'avantage avant de pouvoir émettre un nouveau paquet. Au niveau macroscopique, cela équivaut 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 [58] eu l'idée 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 x_n variant 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 élimine de telles fluctuations et permet ainsi l'établissement d'un régime stationnaire. Suivant la politique implémentée à chaque nœud, on peut alors réaliser différents types d'équité. (Les démonstrations sont disponibles dans [86].)

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, 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 $\max \sum_n B_n \cdot x_n - \frac{1}{2} \cdot (x_n)^2 T_n$ 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 nœuds utilisent une politique définie dans [86] 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 Massoulié et Roberts [86]. A chaque nœud, 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 [94] ont été étudiées des politiques possibles pour le cas de l'équité avec un paramètre α quelconque.

Fenêtres par liens.

E. Hahne [47] proposa une méthode basée sur une fenêtre de taille fixe dont l'originalité est d'être utilisée sur chaque lien et non de bout en bout. Ainsi chaque connexion 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 disponible quand vient son tour, alors

l'ordonnanceur passe à la connexion suivante. 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 variable

Cette méthode est de loin la plus répandue. C'est notamment celle implémentée dans le protocole TCP. L'idée y est de faire varier la taille de la fenêtre de chaque connexion. Ainsi, si la taille de la fenêtre augmente, 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 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 le nombre de flot actifs change.

Famille AIMD

Une famille populaire de ces taux correspond aux systèmes AIMD (Additive Increase Multiplicative Decrease). Chiu et Raj [32] ont étudiés plusieurs familles de paramètres et les ont comparés suivant plusieurs critères, notamment en matière d'efficacité, de rapidité de convergence et 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 [32], ils concluent que la politique AIMD est optimale. Ce raisonnement a été également repris dans [20].

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 , soit $f(t)$ la taille de la fenêtre. On a alors :

$$f(t+1) = u(y(t))f(t) + v(y(t)), \text{ avec } (u, v) : \{0, 1\} \rightarrow R$$

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 et c'est celle qui est utilisée dans le protocole TCP. Kelly, Maulloo et Tan ont montré dans [65] 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 [125], qui considère des RTT hétérogènes. Les auteurs montrent alors que la fonction d'utilité maximisée par le réseau est

$$\sum_{n \in N} \frac{1}{\tau_n} \log \frac{x_n}{v(0) + u(1)x_n}$$

soumis aux contraintes du réseau. τ_n est le RTT de la connexion n .

Une description précise des équilibres obtenus par les différentes versions et implémentations de TCP se trouve dans [81]. 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 est principalement déterminé par le protocole TCP utilisé tandis que le l'AQM joue le rôle de stabilisateur. 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 [34] Crowcroft et Oeschlin proposent une variante de TCP. Leur but est de changer l'agressivité du protocole, 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 p est utilisé pour représenter l'agressivité d'une connexion. Une p -connexion correspond alors à p 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 α afin d'obtenir un partage satisfaisant la version pondérée de l'équité considérée.

Algorithmes aléatoires

L'idée de tels algorithmes, proposés dans [86], est d'associer un processus stochastique aux tailles des fenêtres. Dans ces modèles, on suppose que chaque source peut instantanément savoir si une augmentation de sa taille de fenêtre 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 m deux possibilités. Soit l'on passe à l'état $m + 1$ avec la probabilité b_m (naissance), soit l'on retourne à l'état $m - 1$ avec la probabilité d_m (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.

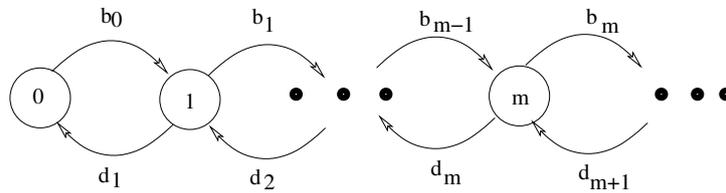


FIG. 1.11 – Processus de naissance et de mort

Ainsi, si l'on choisit des suites b_m et d_m constantes (i.e. $\forall m, b_m = b, d_m = d$) alors, a mesure b/d croit, la distribution des taux se concentre sur les allocations qui maximisent le débit total $\sum_{\mathcal{N}} x_n$.

Lorsque $b_m = (m + 1)^a$ et $d_m = (m - 1)^a$ avec un paramètre a quelconque, alors quand a devient grand, la distribution stationnaire se concentre sur les allocations qui maximisent la somme des logarithmes des taux : c'est l'équité proportionnelle.

Si, pour deux paramètres A et M , on choisit b_m et d_m de sorte que $\frac{b_{m-1}}{d_m} = \exp(A^{M-m})$ alors on peut montrer que si M est suffisamment grand la distribution π converge vers l'allocation max-min quand A tend vers une valeur infinie.

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

1.5.2.3 Indications de congestion

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 un paquet 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, la variante Vegas 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 également présente dans les protocoles tels que l'ECN (*Early Congestion Notification*). L'idée de l'ECN est de marquer les paquets aux niveau des routeurs. Un paquet est marqué s'il traverse un lien congestionné, permettant ainsi à la source d'adapter son débit en conséquence 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 [109, 38]. Deux techniques sont principalement utilisées sur l'ECN pour déterminer comment choisir les paquets à marquer : le Random Exponential Marking (REM, [9]) et le Self-normalized Additive Marking (SAM, [2]). L'ECN a par exemple été utilisé dans [71] 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 [60]. Le principe ici est que chaque source recalcule à chaque itération le débit qu'elle va utiliser en fonction du nombre de liens congestionnés rencontrés. L'idée est proche de celle des algorithmes à base de fenêtre de taille variables, à la différence que l'on tient compte du nombre de liens congestionnés traversés, ce qui offre 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 ces informations).

Dans [60], les auteurs visent la maximisation de la somme des utilités des connexions. Leur algorithme peut alors être modifié de sorte à obtenir une allocation équitable quelconque. L'extension au cas du multicast se trouve dans [61].

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 chaque source adapte son débit (par l'intermédiaire d'une fenêtre de congestion) en fonction de l'état du réseau. Afin d'optimiser l'utilisation des ressources du réseau, on tente désormais d'utiliser des mesures de congestion qui permettent de réagir avant la perte de paquets dans les buffers.

Un axe de recherche est l'obtention de protocoles qui convergent vers un certain débit en évitant les solutions oscillatoires, telles que celles observées dans la version Reno de TCP (dus à la modification à chaque RTT la taille des fenêtres des connexions). 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 aujourd'hui utilisés 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 par l'intermédiaire des fonctions d'utilités.

Il s'avère en effet que 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 λ 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 m . En général, l'idée est d'alors de modifier les priorités implémentées aux niveaux des buffers pour favoriser aux paquets marqués comme étant d'une certaine classe. S'il n'existe pas (pour l'instant) de classes de services dans l'Internet, la situation pourrait changer dans les évolutions futures. Ainsi, dans l'architecture de Internet2 [118] les auteurs proposent un service, «QBone Premium Service» offrant une minimisation des pertes, des délais et de la gigue⁴. On remarque que ce service est

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

principalement intéressant pour les utilisateurs d'applications multimédia 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 [114]. Pour des résultats quantitatifs, sur les communications audio par exemple, le lecteur pourra se référer à [59, 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.

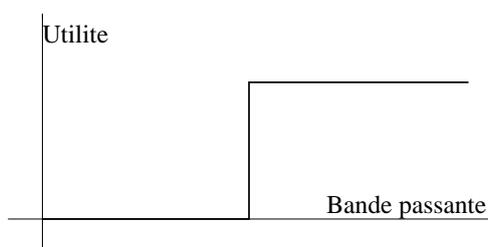


FIG. 1.12 – Fonction ON/OFF

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 leur fonction d'utilité est donnée la figure 1.13.

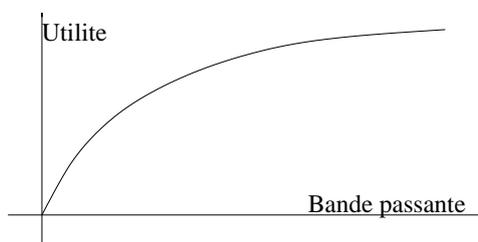


FIG. 1.13 – Application élastiques

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 à une valeur minimale de débit dont à besoin la connexion, que l'on 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. L'allure de leur fonction d'utilité est donné en figure 1.14.

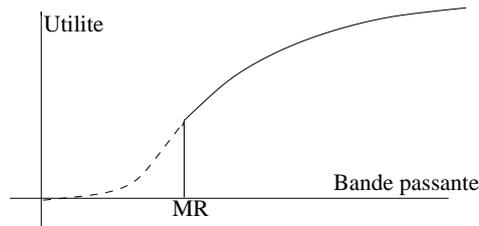


FIG. 1.14 – Application à débit variables

1.6.4 Conclusion

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

$$\max_{n \in \mathcal{N}} \frac{[f_n(x_n)]^{1-\alpha}}{1-\alpha} \quad (1.6)$$

sous les contraintes du système.

Les utilités des applications élastiques et à débits variables 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 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. (en d'autres termes, la solution NBS est invariante par translation.)

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.

Approximation parabolique.

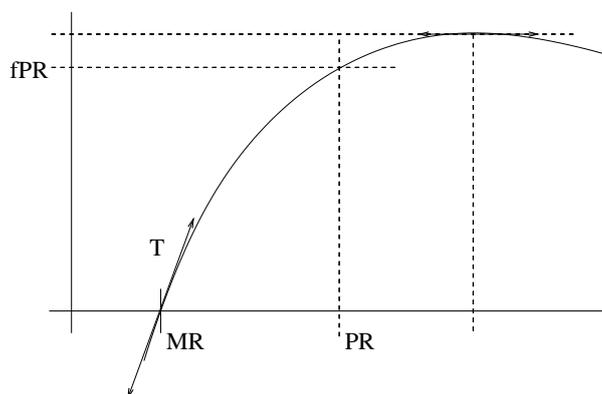


FIG. 1.15 – Fonction d'utilité quadratique

La forme générale d'une telle courbe est :

$$f_i(x_i) = c_i - a_i(x_i - b_i)^2. \quad (1.7)$$

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, β comme étant :

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

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

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

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

Les valeurs correspondantes pour a_i , b_i et c_i peuvent être obtenues par :

$$a_i = T_i \frac{1 - \beta_i}{PR_i - MR_i}, \quad b_i = \frac{PR_i - (2\beta_i - 1)MR_i}{2(1 - \beta_i)}, \quad c_i = \frac{T_i PR_i - MR_i}{4(1 - \beta_i)} \quad (1.8)$$

1.7 Comparaison et propriétés des critères d'équité

1.7.1 Introduction

Le but de cette thèse 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.

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 sous forme d'un problème d'optimisation mais comme une série d'axiomes jugés intéressants pour définir l'équité.

On sait que le paramètre α , 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 la formulation aux fonctions d'utilité, l'intérêt du paramètre α reste inchangé, même s'il n'est pas possible de quantifier son influence de façon analytique, sauf évidemment pour des très simples réseaux.

L'application 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 de pouvoir mieux approcher des fonctions réelles, tout en permettant d'étudier simplement l'influence de quelques paramètres sur les équilibres obtenus. En particulier, nous nous intéressons dans cette section à l'influence de la concavité sur 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 $[MR, PR]$ et 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 super-gradients de f et g : si $\hat{f}(x)$ est le plus grand super-gradient 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 \vec{x} est plus équitable que \vec{y} dans le sens du NBS si $\prod_{n \in N} f_n(x_n) \geq \prod_{n \in N} f_n(y_n)$.

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 (quelle que soit la valeur du paramètre α et en particulier selon la définition originelle de l'équité proportionnelle), et égale à $c/2$.

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 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 ϵ et que ϵ tend vers zéro. Nous concluons qu'il existe un ϵ_0 tel que pour tout $\epsilon < \epsilon_0$, $g(x + \epsilon)f(x - \epsilon) > g(x)f(x)$. Ainsi, l'équité augmente (strictement) en transférant une quantité ϵ_0 vers la connexion ayant l'utilité la moins concave.

En augmentant encore le montant transféré, nous atteignons finalement un maximum local (puisque la fonction est continue sur un intervalle compact). C'est en fait un maximum global car (1.4) 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. Nous supposons que pour les deux connexions $PR > c$. Leurs fonctions d'utilités respectives sont f et g comme représentées sur la figure 1.16.

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

Alors

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

(En $x = 1$, f n'est pas différentiable mais ses super-gradients à ce point sont l'ensemble $[1/3, 1]$). Ainsi f est plus concave que g .

Le NBS est l'argument 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 $\begin{cases} h(x) = 6x(c - x) \\ k(x) = 2(x + 2)(c - x). \end{cases}$

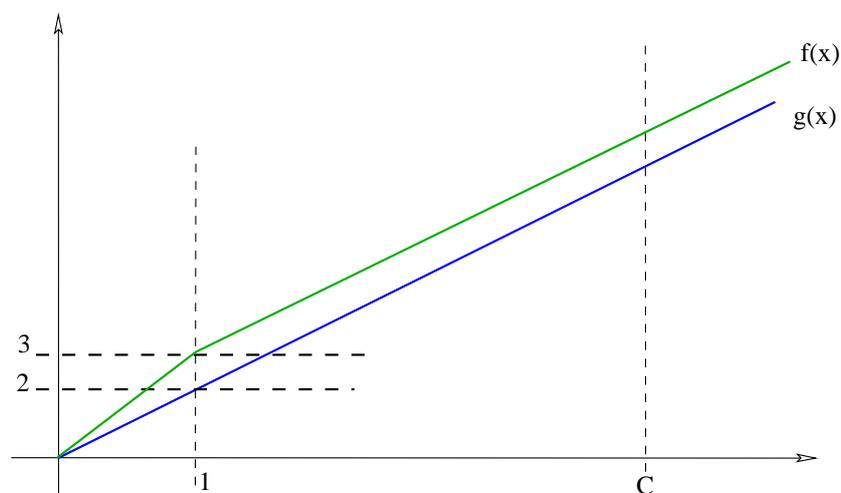


FIG. 1.16 – Fonctions d'utilités des 2 connexions concurrentes.

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 $(x_1^*, x_2^*) = (c/2 - 1, c/2 + 1)$.

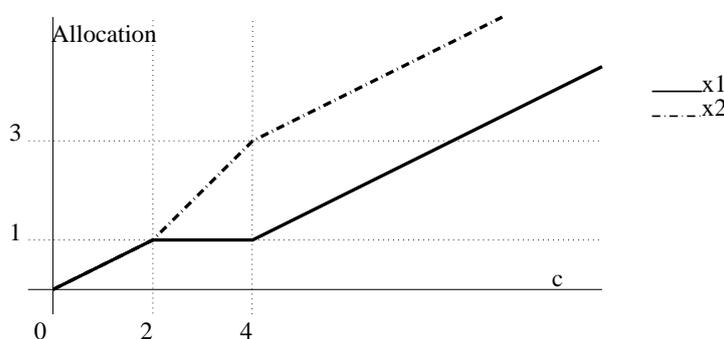


FIG. 1.17 – NBS de deux connexions se partageant un lien.

Nous voyons dans cet exemple que la connexion dont la fonction est 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 existe 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 qui serait indépendante des capacités. En revanche, les concepts d'équité relatifs aux utilités 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 solveurs SDP dans le domaine public⁵. SDP résout 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. Plus de détails sur la SDP sont disponibles dans [21].

Dans cette section, nous verrons comment la programmation SDP est issue de la programmation linéaire (section 1.8.2), et quelques exemples classiques de résolutions par la SDP (section 1.8.3). Pour comprendre le reste de cette section, il est cependant nécessaire de connaître quelques résultats élémentaires d'algèbre linéaire que nous rappelons dans la première sous-section.

1.8.1 Petits rappels d'algèbre linéaire.

Nous rappelons dans cette section quelques résultats élémentaires d'algèbre linéaire qui s'avèreront utiles lors de la résolution par la SDP des problèmes d'allocation équitables de ressources dans les réseaux de télécommunications. Le lecteur familiarisé avec ce domaine pourra donc aisément ignorer cette section.

Des résultats plus complets sont disponibles dans de nombreux ouvrages mathématiques, comme par dans le premier chapitre du livre de Helmberg [50].

Formes quadratiques

Définition 1.8 (Formes bilinéaires). Une fonction $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ est dite bilinéaire si :

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

Soit e_i le i -ème vecteur de la base canonique de \mathbb{R}^n . Définissons la matrice A par : $A = [\phi(e_i, e_j)]_{1 \leq i, j \leq n}$. Nous avons alors la relation suivante :

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

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

Définition 1.9 (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 être écrite sous la forme $q(x) = x^t A x$. Si l'on note x le vecteur $x = (x_1, \dots, x_n)^t$, alors $q(x)$ est un polynôme 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 :

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

Définition 1.10. (formes quadratiques positives) On dit qu'une fonction quadratique $q : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $q_1 - q_2 \succeq 0$.

Définition 1.11. On dit qu'une fonction quadratique positive $q : \mathbb{R}^n \rightarrow \mathbb{R}$ est définie (et on écrit $q \succ 0$) si :

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

Définition 1.12. Une fonction quadratique semi-définie est soit positive soit négative.

Matrices semi-définies positives

Définition 1.13 (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 \mathbb{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.

Définition 1.14 (Matrices symétriques). Une matrice symétrique A est égale à sa transposée⁶ : $A = A^T$.

On note S_m l'ensemble des matrices symétriques réelles.

Proposition 1.4 (Matrices symétriques réelles). Une matrice symétrique de taille m a toutes ses valeurs propres⁷ réelles et il existe une matrice orthonormée⁸ P qui diagonalise A , c'est-à-dire telle que $P^T A P = \Lambda_A$ (en notant Λ_A la matrice diagonale dont les éléments diagonaux sont les valeurs propres de A).

Définition 1.15. (Matrices semi-définies positives) On dit que $A \in S_m$ est semi-définie positive (et l'on note $A \in S_m^+$ ou $A \succeq 0$) si $\forall x \in \mathbb{R}^m, x^T A x \geq 0$. On dit que $A \in S_m$ est définie positive ($A \in S_m^{++}$, $A \succ 0$) si $\forall x \in \mathbb{R}^m, x^T A x > 0$.

Théorème 1.2 (Caractérisation des matrices définies positives). Soit A une matrice symétrique ($A \in S_m$). Les propositions suivantes sont équivalentes :

- A est définie positive
- La valeur propre $\lambda_i(A)$ est strictement positive : $\lambda_i(A) > 0$ pour $i = 1, \dots, m$
- $\exists C \in M_m$ de rang m telle que $A = C^T C$.
- Pour n'importe quelle suite A_i de sous matrices principales «nested»⁹ de sous-matrices de A , $A_i \in S_i, i = 1, \dots, n$, on a $\forall i, \det(A_i) > 0$.

Théorème 1.3 (Caractérisation des matrices semi-définies positives). Soit A une matrice symétrique ($A \in S_m$). Les propositions suivantes sont équivalentes :

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

Théorème 1.4 (Matrices symétriques de taille 2). Soit a, b et c des nombres réels positifs. Alors la matrice

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

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

Théorème 1.5 (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 :

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

⁶La transposée de $A = (A_{i,j})$ est $A^T = (A_{j,i})$.

⁷Les valeurs propres sont les solutions du polynôme caractéristique de A : c'est $\det(A - \lambda I)$.

⁸c'est-à-dire dont les vecteurs colonnes sont 2 à 2 orthogonaux et de norme unitaire.

⁹Une suite «nested» est une suite de sous-ensembles d'indice $J_i, 1 \leq i \leq n$ telle que $J_1 \subset J_2 \subset \dots \subset J_n = \{1, 2, \dots, n\}$ de cardinal $|J_i| = i$. Les matrices A_i sont donc obtenues en effaçant des lignes et colonnes de la matrice originale A .

Théorème 1.6. Une matrice diagonale par bloc ¹⁰ est semi-définie positive si et seulement si tous ses blocs sont semi-définis positifs.

1.8.2 De la programmation linéaire à la SDP

On présente la SDP comme étant une extension de la programmation linéaire. Mais qu'est-ce que la programmation linéaire ? 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 [99].

1.8.2.1 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

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

Un programme linéaire est appelé :

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

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**.

¹⁰Une matrice par bloc est de la forme

$$A = \begin{pmatrix} \boxed{A_1} & & 0 \\ & \boxed{A_2} & \\ 0 & & \boxed{A_n} \end{pmatrix} \quad (1.9)$$

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 1.7 (Dualité en Programmation Linéaire). *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 1.8. (Condition nécessaire et suffisante d'optimalité en programmation linéaire) *Une paire (x, y) de solutions primales et duales sont les solutions optimales de leurs problèmes respectifs si et seulement si :*

$$y_i[Ax - b]_i = 0, \quad i = 1, \dots, m \quad (1.10)$$

ou, de façon équivalente,

$$c^T - b^T y = 0. \quad (1.11)$$

1.8.2.2 La programmation conique

On peut étendre la programmation linéaire à la programmation conique. Pour comprendre en quoi cette dernière consiste, définissons tout d'abord les cônes pointés convexes. Soit \mathbf{E} un espace Euclidien de dimension finie.

Définition 1.16 (Cônes pointés convexes). *Un cône pointé convexe \mathcal{K} est un sous-ensemble de \mathbf{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 \mathbf{E} . Cette relation est partielle. On la note $\geq_{\mathcal{K}}$:

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

Similairement, on définit la relation $>_{\mathcal{K}}$ par : $a >_{\mathcal{K}} b \Leftrightarrow a - b >_{\mathcal{K}} 0 \Leftrightarrow a - b \in \text{int}\{\mathcal{K}\}$ avec $\text{int}\{\mathcal{K}\}$ désignant l'intérieur de \mathcal{K} .

Définition 1.17 (Programmation conique). Soit \mathcal{K} un cône de \mathbf{E} (convexe, pointé et d'intérieur non vide). Soit $c \in \mathbb{R}^n$ un vecteur (vecteur objectif), A un mapping linéaire :

$$A : \begin{cases} x \mapsto Ax \\ \mathbb{R}^n \rightarrow \mathbf{E}. \end{cases} \quad \text{et } b \in \mathbf{E}. \text{ Le problème d'optimisation}$$

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

est un problème conique.

Définition 1.18 (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 1.9 (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, λ) .
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 $\lambda >_{\mathcal{K}} 0$ tel que $A^* \lambda = c$) alors le primal est solvable et $c^* = b^*$.
4. Supposons qu'au moins l'un des problèmes est borné et strictement faisable. Alors une paire faisable (x, λ) est une paire de solutions optimales pour son problème respectif

- (a) Si et seulement si $\langle b, \lambda \rangle = c^T x$ (pas de trou de dualité)
 (b) ou, si et seulement si $\langle \lambda, Ax - b \rangle = 0$.

Corollaire 1.1. *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.*

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.

1.8.2.3 La programmation conique-quadratique

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

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

$$\text{Alors } \mathcal{K} = L^{m_1} \times L^{m_2} \times \dots \times L^{m_k} = \left\{ y = \begin{pmatrix} y[1] \\ y[2] \\ \dots \\ y[k] \end{pmatrix} \mid y[i] \in L^{m_i}, i = 1, \dots, k \right\}.$$

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

Remarque. *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 cruciale est alors «le système de contraintes est-il représentable par des contraintes coniques-quadratiques?».

Un atout majeur de cette famille de cônes est qu'il est possible de construire une sorte de dictionnaire des ensembles de contraintes pouvant être représentées par des cônes quadratiques. On trouve dans [99] et dans [98] 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, en autres.

1.8.2.4 La programmation Semi-Définie Positive

Soit $\mathbf{E} = \mathbf{S}_m$ l'ensemble des matrices de taille $m \times m$ symétriques et $\mathcal{K} = \mathcal{S}_m^+$ le cône des matrices semi-définies positives (telles que définies en section 1.8.1). On remarque que la notation \succeq (respectivement \succ) de la section 1.8.1 est équivalente à la notation $\geq_{\mathcal{S}_m^+}$ (resp. $>_k$). Afin d'alléger les notations, nous l'utiliserons également dans la suite du manuscrit. On note que $\text{int}(\mathcal{S}_m^+) = \mathcal{S}_m^{++}$ est l'ensemble des matrices définies positives.

Un problème SDP est alors $\min_x \{c^T x \mid \mathcal{A}x - B \succeq 0\}$. Dans cette formulation x est un vecteur, \mathcal{A} est un mapping linéaire de \mathbb{R}^n dans \mathbf{S}_m et B une matrice symétrique ($B \in \mathbf{S}_m$). En fait, une façon naturelle de caractériser un mapping linéaire $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbf{S}_m$ est par la donnée d'une suite A_i de n matrices de \mathbf{S}_m telles que :

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

D'où la nouvelle expression SDP :

$$\min_x \{c^T x \mid \sum_{j=1}^n x_j A_j - B \succeq 0\} \quad (1.12)$$

Le problème dual associé est alors ¹¹ :

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

Ceci est la formulation utilisée par Nemirovski [99] et par Laurent et Rendl [74].

Formulation équivalente.

Une formulation équivalente se trouve dans la littérature (voir, par exemple dans ([50])). Notons tout d'abord que (1.12) est équivalent au problème suivant :

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

¹¹La notation $tr(A)$ désigne ici la trace de la matrice A , c'est-à-dire la somme de ses éléments diagonaux.

avec $X \in \mathcal{S}_m^+$, $c \in \mathbb{R}^{n^2}$, $b \in \mathbb{R}^m$ et $A \in M_{m,n^2}$. ⁽¹²⁾

On remarque que, en notant \mathbf{A} le mapping de \mathbf{S}_n dans \mathbb{R}^m associé et défini par ¹³ :

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

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

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

Alors la formulation SDP devient :

$$\begin{cases} \min \langle C, X \rangle \\ \text{tel que } \begin{cases} \mathbf{A}X = b \\ X \succeq 0 \end{cases} \end{cases} \quad \text{et le dual s'écrit alors : } \begin{cases} \max \langle b, y \rangle \\ \text{tel que } \begin{cases} \mathbf{A}^T y + Z = C \\ y \in \mathbb{R}^m, Z \succeq 0 \end{cases} \end{cases} \quad (1.13)$$

Nous utilisons dans la suite cette dernière formulation.

Propriété des problèmes SDP

Comme la SDP est un cas particulier de programmation conique, à la fois le problème primal et le problème dual sont des problèmes SDP.

Définition 1.19. *Un point X est strictement faisable pour le primal si il est faisable et satisfait $X \succ 0$. Similairement, une paire (y, Z) est strictement faisable pour le dual si elle est faisable et satisfait $Z \succ 0$.*

Théorème 1.10 (dualité forte). *Supposons qu'il existe une solution strictement faisable (\hat{y}, \hat{Z}) pour le dual. Soit :*

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

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 1.2. *Soit p^* et d^* définis comme dans le théorème précédent, alors :*

¹²Si X est la matrice $X = (x_{i,j}), 1 \leq i, j \leq n$ alors $\text{vec}(X)$ est le vecteur $(x_{1,1}, \dots, x_{1,n}, x_{2,n}, \dots, x_{n,n})$.

¹³Le produit scalaire considéré est le suivant : $\langle A, B \rangle = \text{tr}(B^T A)$. On peut alors remarquer qu'en fait $\mathbf{A} = \mathcal{A}^T$

- 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.

Des exemples montrant que toutes les hypothèses du théorème sont nécessaires pour assurer que le système est solvable sont disponibles dans [50].

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

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

Le problème primal a 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 $x_{22} \rightarrow +\infty$.

Ramana [110] présente une autre expression de la dualité pour SDP qui permet 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 rédhibitoire 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 [99, 98].

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. Nemirovski [99] propose quelques exemples tirés de la mécanique, l'électricité et la théorie des graphes. Laurent et Rendl [74] propose de nombreux exemples dans lesquels la relaxation SDP apporte de très bons résultats.

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 une paire (V, E) dans laquelle V est un ensemble de nœuds et E un ensemble de liens. Souvent, on associe les liens à des valeurs, appelées *poids*.

Une *coupe* consiste à partitionner l'ensemble des nœuds en 2 ensembles V_1 et V_2 . Le poids de la coupe est la somme des poids des liens reliant un nœud de l'ensemble V_1 à un nœud 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 nœud i au nœud j . Une coupe étant un partitionnement des nœuds, 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 :

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

Goemans et Williamson ont montrés en 1995 [42] que si OPT est la valeur optimale du problème de MAXCUT et SDPOPT la valeur optimale donnée par cette relaxation SDP, alors :

$$OPT \leq SDP \leq \alpha OPT, \text{ avec } \alpha = 1.138\dots$$

D'autres relaxations SDP peuvent être utilisées pour donner des approximations du problème de Max-Cut [50], [74]. 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 $\text{diag}(X) = e$, $X \succeq 0$ et montre la puissance de SDP.

Autres travaux sur la SDP

On peut citer les travaux faits sur la SDP de Anderson, Engebretsen et Hastad [7] 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 [50] et [74]. Des articles

proposent la SDP pour offrir des relaxations intéressantes de problèmes classique NP-complets [88, 70, 7].

Enfin, les méthodes algorithmiques de résolution de la SDP sont variées et ont donné 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 [50].

1.9 Conclusion

Nous avons présenté dans cette partie une famille de critères d'équité paramétrés par un réel α permettant à un opérateur de réseau de choisir un équilibre 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 critères d'équité qui prend en compte la nature des besoins des applications temps-réels utilisées dans les réseaux. 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 trois 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 de quantifier (notamment avec une mesure d'équité associée à chaque critère proposé) 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 à 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 [65].

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

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. On ignore alors quels seront leur impact sur les autres sources et des articles, tels [38] mettent en garde contre le fait que de tels protocoles, vendus pour être des «fast TCP» par exemple ne voient le jour, sans pour autant pouvoir quantifier leur impact. 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 média 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. La première est qu'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 linéaires.

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

Chapitre 2

Introduction aux réseaux terrestres.

Nous appliquons ici les critères d'équités introduits au précédent 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 au modèle.

2.1.1 Notations

Je rappelle que nous nous intéressons ici au cas d'un réseau composé d'un ensemble \mathcal{L} de liens. Chaque lien $l \in \mathcal{L}$ est caractérisé par sa bande passante associée c_l et $C = (c_l)_{l \in \mathcal{L}}$ est le vecteur de capacité. Sur ce réseau circulent un ensemble \mathcal{N} de connexions, c'est-à-dire de communications entre deux points. On note $\kappa = (\kappa_l)_{l \in \mathcal{L}}$ le vecteur de capacité résiduelle, c'est-à-dire de capacité des liens non-utilisé.

Nous considérons dans cette section deux scénarios possibles d'équité, appelés «connection-aware» et «network-aware». Dans le premier, les joueurs considérés sont les connexions. On essaye alors d'optimiser équitablement l'allocation donnée à chaque connexion. Dans la seconde, le but est alors de préserver au mieux le réseau en répartissant au mieux la bande passante résiduelle entre les différents liens du réseau, une fois les demandes des connexions satisfaites.

A chaque joueur j (une connexion ou un lien selon le type d'optimisation choisi) est associée une fonction d'utilité f_j ainsi que la donnée d'une demande de bande passante minimale (MR_j) et maximale (PR_j).

Ces notations sont résumées dans le tableau suivant :

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_j	demande minimale en bande passante du joueur j ,
PR_j	demande maximale en bande passante du joueur j ,
l	un lien,
\mathcal{L}	l'ensemble des liens du réseau,
c_l	la capacité du lien l ,
C	le vecteur de capacité des liens.
κ	le vecteur de capacité résiduelle (non utilisée) des liens.

Le problème est alors de répartir la bande passante disponible dans le réseau entre les différentes connexions tout en respectant un ensemble de contraintes. Nous les explicitons dans les sous-sections suivantes.

2.1.2 Demandes des utilisateurs

Les contraintes les plus évidentes sont celles liées aux demandes des utilisateurs. On suppose que chaque joueur j a une demande minimale MR_j et une demande maximale PR_j . Ces contraintes s'écrivent donc :

$$\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n \text{ pour l'optimisation «connection-aware»}$$

$$\forall l \in \mathcal{L}, MR_l \leq \kappa_l \leq PR_l \text{ pour l'optimisation «network-aware»}$$

et constituent alors un système de $|\mathit{Conn}|$ ou $|\mathcal{L}|$ inégalités linéaires indépendantes selon le type d'optimisation choisi.

Le deuxième type de contraintes dépend du type de routage choisi.

2.1.3 Types de routage

Une des particularités des réseaux filaires 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 paquet 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 fixe, 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 $|\mathcal{L}| \times |\mathcal{N}|$. 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 désire ne 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 suppose que l'on ne s'autorise à utiliser que la quantité $(1 - \kappa_l)c_l$ sur chaque lien l . Soit κ le vecteur de taille $|\mathcal{L}|$ des quantités résiduelles. La contrainte de capacité dans le cas du routage fixe 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. Le réseau est alors considéré comme un ensemble de nœuds reliés entre eux par des liens. Le but est alors de savoir, pour chaque paire de nœuds 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 nœuds u et v , on note $\phi_n^{u,v}$ le flux de la connexion n de u vers v et $\phi_n^{v,u}$ le flux opposé (de v vers u).

On note $N(u)$ l'ensemble des nœuds connectés au nœud u .

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

V est le nombre de nœuds du réseau.

Nous avons alors 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 nœuds, 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 nœud source :

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

5. Aucun flux ne sort du nœud $d(n)$ 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 $d(n)$. Les contraintes liées au routage fractionnaire s'écrivent :

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

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

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

D'où finalement :

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

On remarque que dans le cas du routage fixe comme du routage fractionnaire, le système de contraintes obtenu consiste en un ensemble de contraintes linéaires.

2.2 Types d'optimisations

Nous considérons ici 2 types d'optimisations appelées *connection-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 «connection-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 κ 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 utile si, en pratique, une fois un routage pour une connexion déterminé, on ne peut dynamiquement le modifier).

Pour ce schéma d'optimisation, le vecteur des allocations aux connexion x est une donnée du problème et le vecteur d'inconnues est κ . Le problème d'optimisation s'écrit alors :

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

sous les contraintes exhibées dans la section précédente.

On remarque que, le vecteur d'allocation x 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 à la fois le routage et les bandes passantes associées aux connexions sont fixées, alors la quantité disponible dans chaque lien est complètement déterminée.

2.2.3 Résumé

Alors qu'on avait a priori 4 schémas d'optimisations possibles, nous avons 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.

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

FIG. 2.1 – 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 «connection-aware». Dans le cas d'optimisation «network-aware», ce débit est fixé à la valeur minimale requise par les applications et on s'intéresse alors au partage équitable des bandes passantes résiduelles dans les liens l , qui sont notés κ_l .

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

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

2.3 Conclusion

Nous avons explicité dans ce chapitre trois 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 des optimisation de type «connection-aware», ou les valeurs minimales de bandes passante résiduelles pour 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. Cette approche n'est cependant 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 [128]. 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 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) où 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 l'étude du réseau linéaire et de 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 L nœuds, numérotés de 1 à L et de $N = L + 1$ connexions.

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

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

Le réseau se compose de L 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 «connection-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.

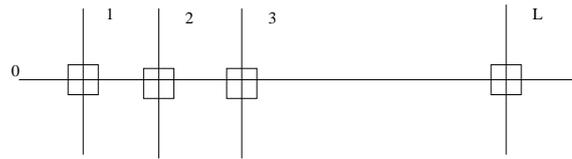


FIG. 3.1 – Un réseau linéaire

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

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

Le système de contraintes s'écrit :

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

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 :

$$\forall n \in [1; L], x_0 + x_n = c_n.$$

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

Nous faisons alors deux hypothèses simplificatrices :

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 $\forall n \in \mathcal{N}, n \neq 0, a_n = a_1, b_n = b_1, c_n = c_1$.

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

$$\begin{cases} g(x) = \frac{1}{1-\alpha} [(f_0(x))^{1-\alpha} + L(f_1(c-x))^{1-\alpha}] & \text{si } \alpha \neq 1 \\ h(x) = f_0(x)(f_1(c-x))^L & \text{sinon} \end{cases} \quad (3.1)$$

sous les contraintes

$$MR_0 \leq x \leq PR_0 \text{ et } MR_1 \leq c - x \leq PR_1.$$

Remarque. Le problème n'a de sens que si :

$$\begin{cases} MR_0 + MR_1 \leq c & \text{sinon le système n'est pas faisable} \\ PR_0 + PR_1 > c & \text{sinon on peut affecter les demandes maximales} \end{cases} \quad (3.2)$$

3.1.1.2 Solution du problème linéaire

En différenciant (3.1), on obtient :

$$\begin{cases} a_0(x - b_0)f_1(c - x)^\alpha = La_1(c - x - b_1)(f_0(x))^\alpha & \text{si } \alpha \neq 1 \\ a_0(x - b_0)(f_1(c - x))^L = La_1(c - x - b_1)(f_0(x))(f_1(c - x))^{L-1} & \text{sinon} \end{cases}$$

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 α :

$$a_0(x - b_0)(f_1(c - x))^\alpha = La_1(c - x - b_1)(f_0(x))^\alpha \quad (3.3)$$

3.1.1.3 Limites possibles

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

Lemme 3.1. Quand L 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 :

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

Comme f_0 est une fonction quadratique, 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 (1.7)), 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 L tend vers l'infini, la limite de l'allocation de la connexion 0 vaut :*

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

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

$$\text{Supposons que } \alpha \neq 1, \text{ on a : } \begin{cases} g(MR_0) = \frac{L}{1 - \alpha} f_1(c - MR_0)^{1-\alpha} \\ g(c - b_1) = \frac{1}{1 - \alpha} [f_0(c - b_1)^{1-\alpha} + L(f_1(b_1))^{1-\alpha}]. \end{cases}$$

Par définition b_1 est l'abscisse de l'apogée de f_1 et donc $f_1(b_1) \geq f_1(c - MR_0)$. 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 $(c - b_1 \geq MR_0) \Rightarrow (g(c - b_1) \geq g(MR_0))$.

$$\text{De même, si } \alpha = 1, \begin{cases} h(MR_0) = 0 \text{ et} \\ h(c - b_1) = f_0(c - b_1)(f_1(b_1))^L > 0. \end{cases}$$

Ainsi, quelle que soit la valeur de α :

$$\mathbf{Si } c - b_1 \geq MR_0 \mathbf{ alors } g(c - b_1) \geq g(MR_0) \mathbf{ et } h(c - b_1) \geq h(MR_0).$$

En fait on peut montrer que cette valeur est effectivement la limite de l'allocation.

Supposons que $\alpha = 1$. Comme f_0 est à valeurs positives sur l'ensemble de définition de x , d'après la définition de h , il nous suffit de regarder le comportement de $f_1(c - Id)$.

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) donc $\lim_{l \rightarrow \infty} \text{argmax} f_0(x)(f_1(c - x))^L = c - b_1$.

Considérons maintenant le cas où $c - b_1 \leq MR_0$. $f_1(c - Id)$ est une parabole dont l'apogée se trouve en $c - b_1$. Ainsi la courbe est décroissante sur $[c - b_1, \infty]$, et en particulier sur $[MR_0, PR_0]$ (car $c - b_1 \leq MR_0$). Donc son maximum se trouve en $c - MR_0$ et finalement $\lim_{L \rightarrow \infty} x = MR_0$.

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

Nous montrons dans la figure 3.2 comment le système converge vers la solution x_{lim} lorsque L tend vers l'infini dans le cas du NBS ($\alpha = 1$). Dans cet exemple, on a $c - b_1 \leq MR_0$.

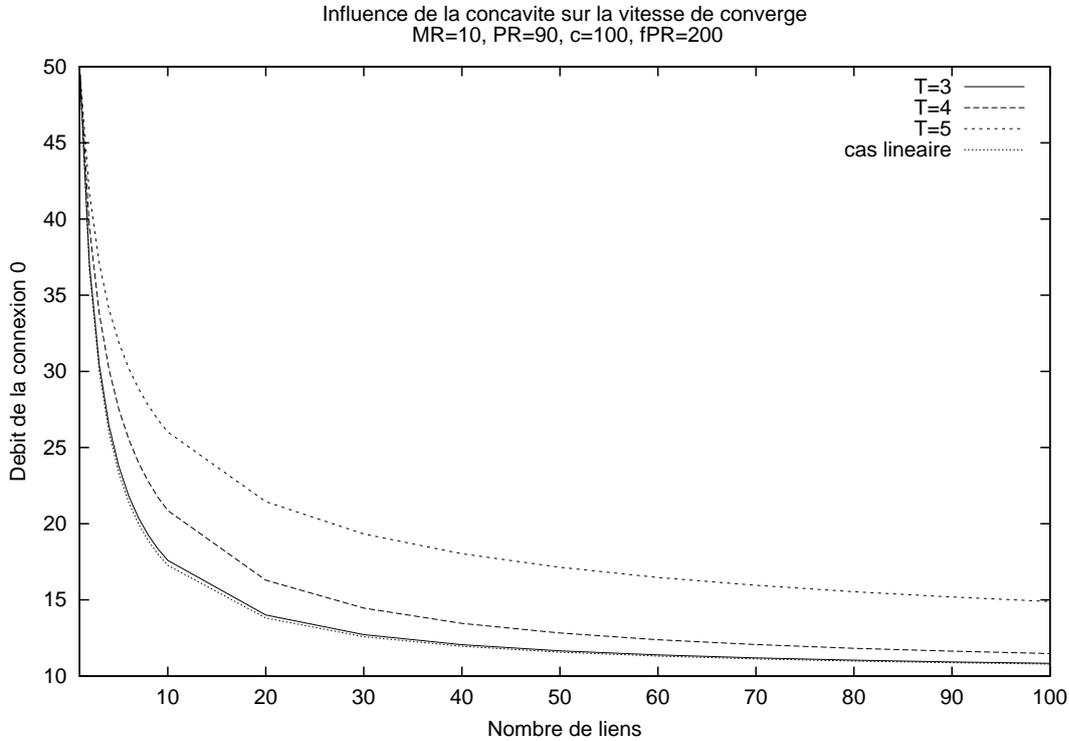


FIG. 3.2 – Influence de la concavité de la connexion 1 sur la vitesse de convergence de l'allocation

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 $c - b_1 < MR_0$ alors x vérifie :*

$$x - MR_0 \sim Z \text{ avec } Z^\alpha = \frac{1}{L} \frac{1}{f_0'(MR_0)^{\alpha-1}} \frac{f_1(c - MR_0)^\alpha}{f_1'(c - MR_0)}. \quad (3.4)$$

Si $c - b_1 \geq MR_0$ alors x vérifie :

$$x = c - b_1 + Z + o(1/L) \text{ avec } Z = \frac{1}{L} \frac{c_1^\alpha}{2a_1} \frac{f_0'(c - b_1)}{f_0(c - b_1)^\alpha}. \quad (3.5)$$

Démonstration. Pour un L fixé et une valeur de α donnée, notons $x(L)$ l'allocation de la connexion 0. Reprenons alors l'équation (3.3).

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 $MR_0 < b_0$ et $(f_1(c - MR_0))^\alpha \neq 0$)

Nous examinons maintenant le membre de droite de l'équation (3.3). Il peut être écrit :

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

ce qui conduit à l'équation (3.4) 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 : $-La_1z(c_0 - a_0(c - b_1 - b_0 + z)^2)$, d'où le résultat de l'équation (3.5). ■

Remarque. On peut reprendre l'équation (3.5) 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 [128]).
- Quand $\beta_1 \rightarrow 1/2$ on obtient :

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

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 3.3) 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 α , expliciter les allocations limites et donner un équivalent de ces allocations lorsque le nombre de liens devient grand.

3.1.3 Réseau grille

Ce réseau est la généralisation naturelle du réseau linéaire. Le réseau consiste en $J \times K$ liens parcouru par J connexions horizontales et K verticales comme montrées dans la Figure 3.3. Nous nous intéressons ici encore au cas du routage fixe, routage rendu évident de par le choix des connexions.

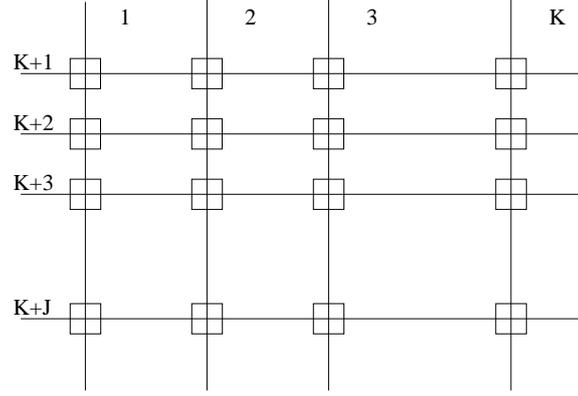


FIG. 3.3 – Un réseau grille.

Comme pour l'exemple précédent, nous supposons que pour chaque $k \in \{1, \dots, K\}$, $j \in \{1, \dots, J\}$, $MR_k + MR_{K+j} < c_{k,K+j}$ et $PR_k + PR_{K+j} > c_{k,K+j}$. Nous supposons également que les capacités des liens sont tous identiques $\forall i, k, c_{i,k} = c$.

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 $x_v = c - x$).

Nous voulons alors maximiser :

$$\begin{cases} \sum_{n \in \mathcal{N}} \frac{1}{1-\alpha} f_n(x_n)^{1-\alpha} = \frac{1}{1-\alpha} (K f_v(x_v)^{1-\alpha} + J f_h(x)^{1-\alpha}) & \text{si } \alpha \neq 1 \\ \prod_{n \in \mathcal{N}} f_n(x_n) = (f_h(x))^J * (f_v(c-x))^K & \text{sinon.} \end{cases}$$

Proposition 3.3. Dans le réseau grille, si $c - b_v < MR_h$, x vérifie : $x - MR_h \sim Z$ avec :

$$Z^\alpha = \frac{J}{K} \frac{1}{f_h'(MR_h)^{1-\alpha}} \frac{f_v(c - MR_h)}{f_v'(c - MR_h)}.$$

Sinon, $x = c - b_1 + Z + o(K/L)$ avec $Z = \frac{J}{K} \frac{c_v^\alpha f_h'(c - b_v)}{2a_v f_h'(c - b_v)}$.

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

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

3.2 Méthode Lagrangienne

La méthode de Lagrange a été proposée par [128] pour obtenir le NBS dans le cas particulier de fonctions d'utilités linéaires, et de l'optimisation «connection-aware» avec routage fixe. 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 rappelle que β est le paramètre de concavité des fonctions d'utilité). Cette méthode ne peut cependant pas être généralisée telle quelle pour les autres valeurs de α .

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 $\forall l \in \mathcal{L}, \sum_{n \in \mathcal{N}} a_{ln} MR_n < C_l$ (ce qui revient à dire qu'aucun lien n'est saturé¹), le NBS est caractérisé par :*
 $\forall l \in \mathcal{L}, \exists \mu_l \geq 0$ tels que $\forall n \in \mathcal{N}$, nous avons

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

Démonstration. Sous l'hypothèse $\sum a_{ln} MR_n < C_l$, l'ensemble 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 .² 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 :

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

Pour $l \in \mathcal{L}$, $\lambda_l \geq 0$ 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. $\mu_l \geq 0, l \in \mathcal{L}$ sont les multiplicateurs de Lagrange associés aux contraintes de capacité.

¹Cette hypothèse n'est pas restreignante. 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.

²C'est-à-dire continue, dérivable et sa dérivée est également continue.

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

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

On note que $\sum a_{ln}MR_n < C_l$ implique que $\forall n, \lambda_n = 0$ comme dans [128]. Enfin, l'on a soit $x_n = PR_n$ soit $\delta_n = 0$, ce qui mène à la conclusion. ■

On peut noter que :

Remarque. Quand $\beta \rightarrow 1$ nous obtenons la solution de [128] correspondant à l'utilité linéaire : $x_i = \min \left(PR_i, MR_i + \left[\sum_{l=1}^L \mu_l a_{l,i} \right]^{-1} \right)$.

Dans cette caractérisation sont donc introduits les paramètres μ_l représentant les coûts associés aux liens du réseau. Ainsi, μ_l 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, μ est le vecteur de multiplicateurs de Lagrange et \mathcal{X} l'ensemble des vecteurs d'allocations possibles.

Notre problème est de résoudre :

$$\min_{x \in \mathcal{X}} f(x) = \sum_{n \in \mathcal{N}} \ln f_n(x_n)$$

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

$$\max_{\mu \in \mathbb{R}_+^L} d(\mu) \text{ avec } d(\mu) = \min_{x \in X} \mathcal{L}(x, \mu) = \mathcal{L}(\bar{x}_i, \mu) \quad (3.6)$$

en notant \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 μ donné, on pose pour chaque connexion n :

$$x_n(\mu) = g_n \left(\sum_{l \in \mathcal{L}} \mu_l \cdot a_{l,n} \right). \quad (3.7)$$

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 $\mu \in \mathbb{R}^L$:

$$d(\mu) = \sum_{n=1}^N -\ln(f_n(g_n(\sum_{l \in [1..L]} \mu_l a_{l,n}))) + \sum_{l \in \mathcal{L}} \mu_l \sum_{n \in \mathcal{N}} a_{l,n} g_n\left(\sum_{l \in \mathcal{L}} \mu_l a_{l,n}\right) - \sum_{l \in \mathcal{L}} c_l \mu_l. \quad (3.8)$$

Comme dans [128], il n'y a pas de trou de dualité. L'idée est alors de considérer un pas judicieusement choisi et de poser à chaque itération :

$$\mu_l^{(k+1)} = \max(0, \mu_l^{(k)} + \gamma \cdot \frac{\partial d}{\partial \mu_l}(x(\mu^{(k)}))).$$

On peut alors montrer de façon similaire à [128] 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 [128] :

$$\frac{\partial d}{\partial \mu_l} = \sum_{n \in \mathcal{N}} \ln(f_n(g_n(\sum \mu_l a_{l,n}))) - c_l$$

L'algorithme se compose alors naturellement de deux éléments. D'une part, un *algorithme de liens* par lequel chaque lien l met à jour la valeur de son coût μ_l en fonction des bandes passantes des connexions qui le traversent et de sa capacité. De l'autre côté un *algorithme de nœuds* par 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 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 au réseau, 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 examiné 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 de l'équité proportionnelle et de fonctions d'utilité linéaires et de réseaux de topologie quelconques. Enfin, dans le cas de fonctions d'utilités quadratiques, on peut encore utiliser cette méthode basée sur l'utilisation des multiplicateurs de Lagrange, mais les paramètres des fonctions d'utilité de chaque connexion doivent alors être transmis au réseau, ce qui limite l'intérêt de cette méthode dans ce contexte.

L'utilisation du problème dual du Lagrangien a pour but 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-Définie Positive est un algorithme de résolution générale 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 forme générale des matrices SDP correspondant à notre problème. Dans une deuxième section, nous explicitons les particularités des matrices associées aux différentes valeurs possibles de α . Nous détaillons alors dans la section 4.3 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 d'optimisations «connection-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 joueurs et éventuellement celles liées au routage fractionnaire),
- Remplacer la fonction objectif (de la formule 2.1) par une seule variable.

On rappelle qu'une matrice par bloc 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 blocs de taille unitaire (c'est-à-dire des scalaires) et que notre fonction objectif peut être exprimée par une suite de blocs de taille 2×2 .

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. Nous verrons cependant que la taille de la matrice et le nombre de variables demeurent raisonnables dans ce type de problème.

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. Nous avons de plus remarqué dans la section 1.8 que de telles contraintes pouvaient être exprimées par des matrices de taille 1 (matrices scalaires).

Nous explicitons dans cette sous-section les contraintes relative au système de routage fixe (le cas du routage fractionnaire étant 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 $\forall l \in \mathcal{L}, (Ax)_l \leq c_l$.

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

Contraintes des joueurs.

Dans le cas des optimisations «connection-aware», elles sont données par : $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$. On introduit ainsi $2.N$ matrices scalaires $util$ telles que

$$util_{1,1}^i = \begin{cases} PR_i - x_i & \text{si } i \leq N \\ x_i - MR_i & \text{si } N < i \leq 2N \end{cases} \quad (4.1)$$

Similairement, pour l'optimisation «network-aware», il nous suffit d'introduire $2.L$ matrices scalaires telles que :

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

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 soit les capacités résiduelles des liens du réseau $\kappa_l, l \in \mathcal{L}$, soit les bandes passantes allouées aux connexions $x_n, n \in \mathcal{N}$. Les problèmes étant identiques, nous utilisons dans la suite de cette section la notation $x_n, n \in \mathcal{N}$ pour désigner la variable considérée.

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. Soit F_n la matrice :*

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

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

Démonstration. Il s'agit là 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 avec cet 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 toute valeur de α on a la relation $\frac{1}{1-\alpha} \sum w_n^{1-\alpha} \leq \frac{1}{1-\alpha} \sum f_n(x_n)^{1-\alpha}$ (il suffit de distinguer 3 cas selon la valeur de α : $\alpha < 1$, $\alpha = 1$ et $\alpha > 1$).

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

Lemme 4.1. *Soit z une variable réelle et $y_n, n \in \mathcal{N}$ 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 à la résolution du problème car :

$$z \leq \sum_{n \in \mathcal{N}} y_n \leq \sum_{n \in \mathcal{N}} \frac{w_n^{1-\alpha}}{1-\alpha} \leq \sum_{n \in \mathcal{N}} \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$$

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") :

$$y \leq w^{1-\alpha} \quad \text{si } 1-\alpha > 0, \quad y \geq w^{1-\alpha} \quad \text{sinon.} \quad (4.3)$$

4.2 Les différentes valeurs de α

Nous présentons dans cette section les matrices dont la positivité permet de lier les variables y aux variables w conformément à l'équation 4.3. 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 $y \leq 1/w$:*

$$H = \begin{pmatrix} w & 1 \\ 1 & y \end{pmatrix}. \quad (4.4)$$

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 $0 < \alpha < 1$.

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

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

Grâce à la proposition suivante, on peut donc résoudre le problème avec une précision aussi bonne que souhaitée.

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 :

$$\begin{pmatrix} y_{i-1} & y_i \\ y_i & w \end{pmatrix} \text{ si } c_i = 1 \quad \begin{pmatrix} y_{i-1} & y_i \\ y_i & 1 \end{pmatrix} \text{ si } c_i = 0$$

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 $y_p = y$. ■

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 de positivité 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 pour les autres valeurs de α , car l'on ne fait plus intervenir les variables intermédiaires w . Nous exhibons directement une suite de matrice dont la contrainte SDP assure que :

$$z \leq \prod_n w_n.$$

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 :*

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

Démonstration. Soit p le plus petit entier tel que $2^p \geq N$. On construit une fa-mille 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 :

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

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 $z_{1,2}^2 \leq w_1 w_2$, $z_{3,4}^2 \leq w_3 w_4$, $z_{1,4}^2 \leq z_{1,2} z_{3,4}$ et finalement $z_{1,2^p}^{2^p} \leq \prod_n w_n$. ■

Ainsi, la résolution du problème du NBS par la SDP, nous amène 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 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_n associée, la solution donnera $x_n = PR_n$.

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

Remarque. *Dans le cas de N connexions avec L liens, nous affirmons que nous aurons au plus $6N - 7$ variables, $4N - 5$ blocs de taille 2, et $4N + L - 4$ blocs 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 : $q + \sum_{i=0}^p q2^{i-1} = 3 \times 2^p - 1$ variables, $q + \sum_{i=1}^p q2^{i-1} = 2^{p+1} - 1$ blocs de taille 2 et $L + 2 \cdot 2^p = 2^{p+1} + L$ blocs 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 l'optimisation «connection-aware» avec routage fixé sur un cas simple de réseau.

On considère un réseau de $L = 4$ liens et $N = 3$ connexions. Les chemins suivis par les connexions sont définis 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 soit de 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

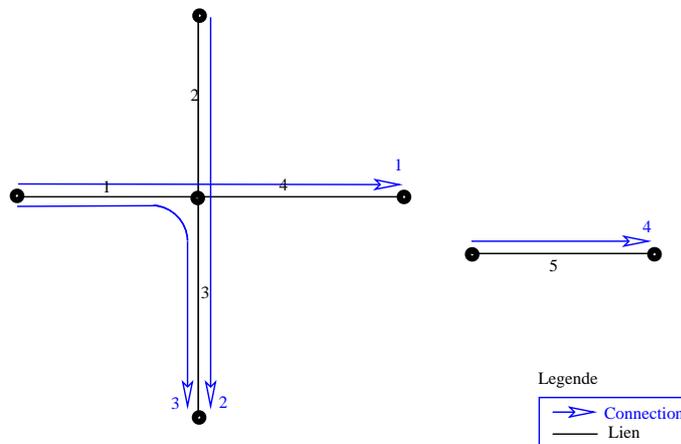


FIG. 4.1 – Exemple simple de calcul du NBS.

Étape 1

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

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

Étape 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 $\prod_n (c_n - a_n(x_n - b_n)^2)$.

Étape 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 :

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

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

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

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

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 construites 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-après. Le temps de calcul (y compris la partie affichage) dans les deux cas fut inférieur à la minute.

Dans le cas des optimisation «connection-aware», nous avons considérés des paramètres identiques pour tous les liens et les joueurs (alors que le programme permet de manipuler des capacités différentes sans augmenter la complexité). C'est ainsi que tous les liens ont la même capacité $C = 100$ et que les différents joueurs ont la même utilité quadratique 1.7 avec les paramètres $MR = 10$, $PR = 80$, $a = 1/490$, $b = 745$ et $c = 1102.5$.

En revanche, 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.

Pour chaque réseau, nous présentons une figure avec l'ensemble des liens et la quantité de bande passante attribuée à chaque connexion. Les paramètres de bande passante et les affectations sont données en pourcentage de la capacité totale du lien.

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. [107] 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 réseau représenté en figure 4.2. Il comprend $L = 10$ liens et $N = 11$ connexions. Nous nous intéressons ici à la résolution de l'optimisation «connection-aware» avec routage fixé dans le cas du NBS ($\alpha = 1$). Le routage est défini dans la matrice A ci dessous :

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

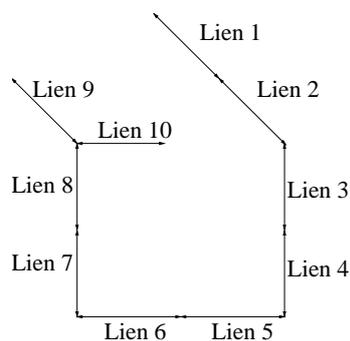


FIG. 4.2 – Petit réseau à routage fixé.

La solution fournie par la SDP est donnée en Figure 4.3. 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 blocs de taille 2, et 42 de taille 1).

Dans ce réseau 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.

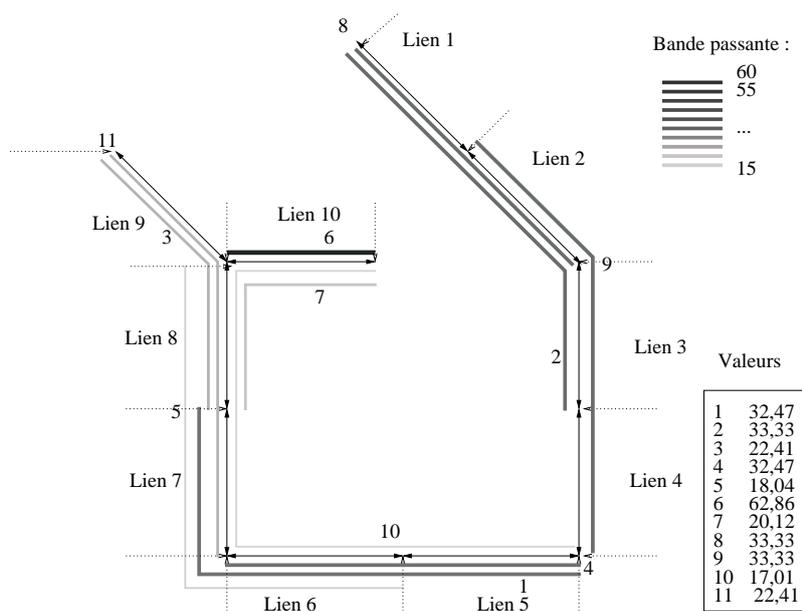


FIG. 4.3 – Premier réseau

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 déterminer, sans l'aide de l'ordinateur, les différentes allocations des connexions. Cependant, les résultats numériques sont conformes à l'intuition puisque la connexion 6 (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 (traversant le maximum de liens) obtient la plus faible (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 du nombre de liens saturés et non le nombre total de liens traversés 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 [1] et représenté dans la figure 4.4. Il se compose de 11 nœuds, représentant 11 grandes villes Européennes.

Dans les deux paragraphes suivants, deux schémas d'optimisations sont présentés. Le premier correspond à l'optimisation «connection-aware» avec routage fixé, le second

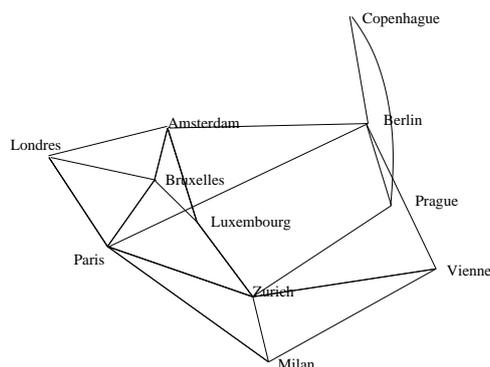


FIG. 4.4 – Réseau COST : liens.

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 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 ont 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 prévisions¹. 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.

Connexion	Bande passante	Connexion	Bande passante	Connexion	Bande passante
Londres-Paris	33.93	Zurich-Milan	71.58	Milan-Vienne-Berlin	37.00
Londres-Bruxelles	80.00	Copenhague-Berlin	80.00	Milan-Paris-Bruxelles	27.93
Londres-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-Vienne	63.00	Paris-Zurich-Vienne	25.48
Bruxelles-Paris	43.66	Milan-Vienne	63.00	Londres-Paris-Milan	24.74
Paris-Berlin	80.00	Berlin-Amsterdam-Luxembourg	27.11	Londres-Paris-Zurich	21.87
Paris-Zurich	33.19	Zurich-Prague-Berlin	50.00	Londres-Amsterdam-Berlin	23.73
Paris-Milan	47.34	Zurich-Luxembourg-Amsterdam	35.79	Vienne-Zurich-Paris-Londres	19.46
Zurich-Vienne	55.06	Zurich-Luxembourg-Bruxelles	35.79	Milan-Zurich-Luxembourg-Amsterdam	28.42

FIG. 4.5 – Allocation de bande passante dans les réseaux COST.

¹C'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

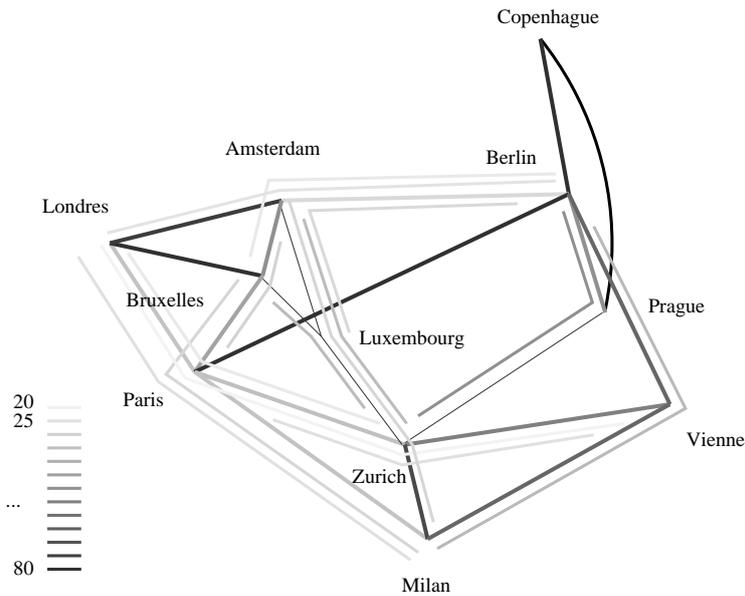


FIG. 4.6 – Le réseau COST : solution.

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.

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 α . Nous présentons 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 possibles². Les demandes ne sont plus identiques, comme illustré dans le tableau 4.1.

²Correspondant à l'ensemble de toutes les paires sources-destinations.

Paris - Milan	5	Paris - Zurich	6	Paris - Luxembourg	1	Amsterdam - Luxembourg	1
Paris - Londres	10	Paris - Vienne	2	Paris - Amsterdam	5	Paris - Copenhague	1
Paris - Prague	1	Paris - Berlin	11	Milan - Amsterdam	2	Milan - Luxembourg	1
Paris - Brussels	6	Milan - Berlin	9	Milan - Copenhague	1	Zurich - Luxembourg	1
Milan - Vienne	3	Milan - Brussels	2	Amsterdam - Copenhague	1	Luxembourg - Brussels	1
Milan - Zurich	6	Milan - London	3	Zurich - Amsterdam	3	Zurich - Copenhague	1
Zurich - Vienne	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 - Vienne	1	Vienne - Amsterdam	1	Vienne - Luxembourg	1
Prague - Milan	1	Vienne - London	2	Vienne - Copenhague	1	London - Amsterdam	5
Vienne - Brussels	1	Vienne - 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

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 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 - Milan	57	Zurich - Prague	87	Vienne - Berlin	96	Amsterdam - Londres	61
Paris - Zurich	68	Prague - Vienne	55	Berlin - Amsterdam	92	Amsterdam - Copenhague	29
Paris - Berlin	97	Prague - Berlin	93	Berlin - Copenhague	81	Amsterdam - Bruxelles	90
Paris - Bruxelles	68	Milan - Bruxelles	73	Prague - Luxembourg	32	Amsterdam - Luxembourg	62
Paris - Londres	91	Zurich - Vienne	41	Prague - Copenhague	61	Londres - Copenhague	87
Milan - Zurich	61	Milan - Vienne	81	Zurich - Luxembourg	73	Luxembourg - Bruxelles	61
Bruxelles - 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 - Bruxelles	73
Paris - Bruxelles	76	Zurich - Prague	76	Prague - Luxembourg	47	Amsterdam - Luxembourg	74
Paris - Londres	78	Prague - Copenhague	74	Amsterdam - Bruxelles	77	Londres - Copenhague	73
Milan - Bruxelles	76	Vienne - Berlin	92	Bruxelles - 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 α croît, on tend vers une coloration plus homogène des liens. Ceci est un résultat attendu puisqu' α est le paramètre d'équité. Quand α croit, 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 α croit de 0.5 à 5, les liens deviennent plus remplis en moyenne (le taux moyen d'utilisation des liens croit de 70.38 à 73.83), ce qui pourrait paraître surprenant, puisque la bande passante allouée reste constante de même que les capacités des liens. On peut en fait expliquer ce résultat de la façon suivante : lorsque α 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 et donc également l'utilisation du réseau.

Il est également intéressant de noter que le taux d'utilisation de certains liens (Berlin-Copenhague par exemple), a augmenté quasiment jusqu'à leur maximum d'utilisation

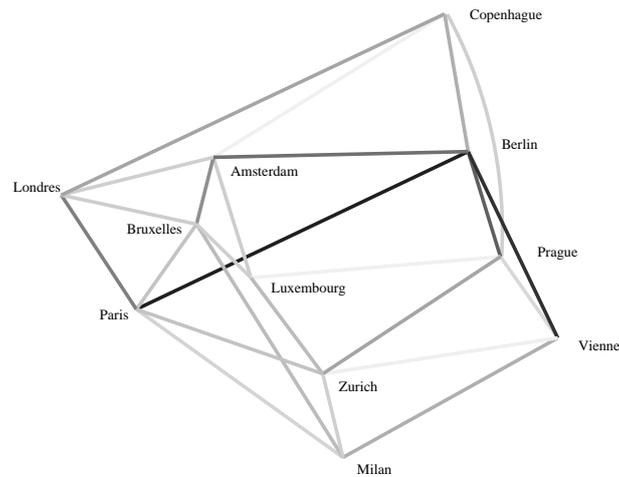


FIG. 4.7 – Allocation de bande passante pour le réseau COST, $\alpha = 0.5$.

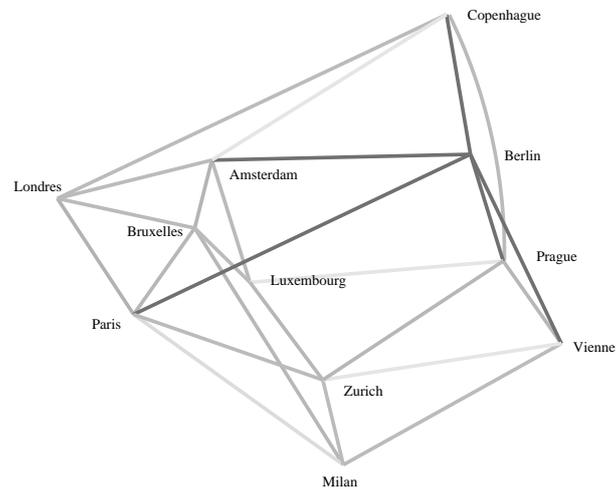


FIG. 4.8 – Allocation de bande passante pour les réseaux COST, $\alpha = 5$.

dans la solution pour $\alpha = 5$, qui est supposé être davantage équitable ! Ceci montre que la recherche de l'équité à une échelle globale pousse parfois quelques utilisateurs dans des situations critiques alors qu'ils sont à l'intérieur de leurs bornes dans des cas moins équitables.

4.5 Conclusion

Nous avons présenté dans ce chapitre une méthode de résolution centralisée basée sur la programmation SDP. Une telle approche est adaptée aux problèmes de dimensionnement de réseaux. Nous avons présenté une méthode de construction possible de la matrice semi-définie positive et avons montré quelques résultats numériques dans des contraintes différentes (roulage 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 appliqués à des fonctions d'utilité quadratiques pour des valeurs du paramètre d'équité α 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 les matrices utilisées sont très creuses puisque diagonales par bloc et que les blocs sont au plus de taille 2. 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 $\alpha \neq 1$, soit p l'entier servant dans l'approximation de α 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 $\mathcal{O}(N+L)$. Le nombre des blocs reste donc raisonnable ce qui permet en pratique d'utiliser la SDP pour résoudre des problèmes de taille réelle.

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 se justifie très bien dans ce domaine, 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 important 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.

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 où le nombre des autres connexions mais uniquement le coût associé aux liens qu'il utilise. Dans le cas de fonctions d'utilité linéaires, le réseau, de son côté, n'ayant pas à connaître les fonctions d'utilité des utilisateurs présents. Dans le cas de fonctions d'utilités quadratiques cependant cette connaissance est nécessaire.

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 solveur général de SDP, certains étant même disponibles dans le domaine public. 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 encore très ouvert. Il serait intéressant d'introduire à notre modèle d'autres notions fondamentales en télécommunications. Par exemple, les contraintes inhérentes à l'utilisation du *multi-cast* auraient tout à fait leur place dans cette étude.

De plus, dans le cas du routage fractionnaire, nous avons supposé que tous les chemins

possibles entre deux points étaient 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 α (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 [80].

Il a été montré [80] que l'équilibre obtenu par TCP Vegas satisfait une certaine forme d'équité proportionnelle :

$$\max_x \prod x_n^{d_n \cdot \delta_n}$$

Avec d_n le RTT associé à la connexion et δ_n un paramètre, représentant la différence maximale entre le débit attendu et le débit réel minimum auquel la source veut voir sa connexion maintenue³. (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 $d\delta \neq 1$, 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

³En 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, ce qui permet d'éviter les comportements oscillatoires

afin 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 [51] 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.

Atouts de la SDP.

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 une véritable révolution ces dernières années, tant au niveau des technologies utilisées que du nombre de leurs abonnés. Dans les réseaux de seconde génération (comme le GSM en Europe) chaque mobile se voit affecter un débit constant prédéfini. En d'autres termes, il n'existe qu'un seul type de connexion. Ceci est du en partie au fait qu'au moment 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 les 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 les concepts d'équité introduits dans le premier chapitre de cette thèse 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 système 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 résumons brièvement les principaux travaux récents d'allocations effectués dans le domaine des réseaux mobiles. Nous présentons dans un second notre approche et en particulier le modèle théorique que nous avons développé et qui permet de prendre en compte les besoins spécifiques des différents types de connexions du réseau. Enfin, dans un troisième chapitre nous appliquons ce modèle théorique et concluons cette partie 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 applicables 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, le temps est divisé 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 types de systèmes, plus le nombre d'utilisateur est grand, plus le débit sera réduit. En effet, dans les systèmes TDMA, le temps entre deux slots consécutifs alloués à une même connexion augmentera, tandis que en mode CDMA, une augmentation de la taille des codes sera nécessaire afin que le décodage puisse se faire sans erreur. Nous y reviendrons plus en détail 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 avec

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.

Nous mentionnons 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 fil.

Quelques articles se sont déjà penchés sur l'allocation équitable de ressources dans les réseaux sans-fil. L'équité a en effet été introduite dans des problèmes d'ordonnancement dans des systèmes TDMA et CDMA. Ainsi, l'article [82] s'intéresse à un ordonnancement 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. Une étude très similaire est présentée [14].

Dans la même famille d'articles 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 sont caractérisées par des besoins différents (en terme de nombre de slots par trame, de temps maximal d'émission et de débit d'information). Le but de l'ordonnanceur est alors d'allouer à chaque connexion des paires (slots, code) qui respectent les besoins des connexions.

On peut noter également l'article [11] qui s'intéresse au problème d'ordonnancement 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 orthogonaux¹ 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 ré-émission). Les auteurs proposent alors un ordonnanceur, 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.

¹Pour une définition précise de ce terme, on se référera à la section 8.1.6. 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.

Ainsi, dans cette famille de travaux, l'équité apparaît sous la forme d'informations d'agencement des paquets dans les buffers (voir également [90]). Nous souhaitons, au contraire, proposer des méthodes de calculs d'allocation équitables de débits aux sources.

Nous citons enfin [33] qui considère le lien descendant d'un système avec trois types de canaux : des canaux de circuit dédiés (pour les applications temps réel), un canal d'overhead 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

Nous présentons dans cette section quelques travaux effectués sur le contrôle des débits dans les réseaux sans fil.

Trafic non temps-réel.

Plusieurs articles étudient les problèmes d'optimisation de débit pour le trafic non temps réel. Ainsi, on cherche dans [117] 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 sans de perte, ou si l'on transmet avec à $2R$ 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 [108], 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.

La constatation simple que le débit alloué n'a pas la même conséquence sur le volume de donné échangé en fonction du type d'application est faite dans [5]. 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 [67] par exemple n'est pas adéquat pour décrire l'effet de l'allocation de débit pour les applications non temps réel. Pour contourner cette difficulté, Altman [5] décrit alors l'utilisation du réseau par le calcul de la capacité d'Erlang² en fonction du débit alloué aux applications.

Allocation de taux instantanés.

Une autre direction de recherche liée est l'allocation de taux de transmission instantanés des paquets au niveau des buffers dans les systèmes CDMA. Dans les articles qui s'y intéressent (telles que [31, 84, 126]), le véritable taux de transmission de la source n'est pas contrôlé. Par exemple [132] considère trois profils de trafic. 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 instantanés des connexions des différents profils sont calculés en fonction de bornes sur les pertes de paquets dans les buffers des mobiles.

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 [12] 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. 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

²Erlang : 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 sur une ligne pendant une heure et de la durée moyenne des sessions. Sa valeur est donc comprise entre 0 et 1 (à 1, la ligne est constamment occupée). A partir d'elle, il est ensuite possible de déterminer la valeur optimale du nombre de lignes nécessaires pour que la qualité de service (i.e taux de refus) soit acceptable.

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 [100] 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 décrite dans [49]. 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. Cependant, 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 de [49] considèrent le cas de la macro-diversité dans le lien montant, c'est-à-dire le cas où le signal du mobile peut être reçu par plusieurs stations de base et recomposé ensuite. Un article proche est [22].

Étude 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 [84] 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 [75] 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 profils de trafic.

Contrôle d'admission.

Le problème du contrôle d'admission dans un réseau DS-CDMA est étudié dans [10]. 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 [129] a présenté une étude statique du lien montant. Il propose un algorithme itératif pour trouver le vecteur de puissance minimal. Le débit des connexions étant fixé, la contrainte 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 énonce alors des conditions assurant la convergence de l'algorithme.

Enfin [4] 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, [67] 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 polynomiaux 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 [102], dans lequel les auteurs étudient le contrôle optimal à la fois de la puissance et du *spreading gain* (augmenter ce dernier est équivalent à diminuer le débit). Les auteurs restreignent leur étude au cas d'une cellule isolé et au cas montant. Leur but est de maximiser le débit total des sources NRT. Le modèle inclue le codage de canal (FEC) et une fonction générale du BER (Bit Error Rate), qui est une fonction du SIR. Des modèles similaires sont disponibles dans la littérature [103, 101]. 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 meilleur 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 avons vu dans ce chapitre quelques axes principaux de recherche en matière d'allocations de ressources dans les réseaux mobiles. Nous pouvons citer en outre l'article [31] qui propose une méthode général d'allocation de ressource en adoptant un modèle statistique. Chaque mobile génère un certain trafic (de voix) qu'il envoie à la station. Le but est alors pour cette dernière de 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 prises en compte ici.

Parmi toutes les directions de recherche mentionnées dans ce chapitre, notre axe est plus proche de [67] et [102] car il s'intéresse à l'allocation véritable des débits de transmission aux sources des applications temps-réel (plutôt qu'à l'intérieur du réseau). 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 [102] qui ne considère que le cas du lien montant sur un seul lien, ou de [33] 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 généralement 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. Cependant, dans les réseaux de téléphonie mobiles de type CDMA, chaque utilisateur est une source de bruit pour les autres. Ainsi, si le nombre d'utilisateurs devient trop important, le bruit total augmente et entraîne une chute de la qualité des communications. Un contrôle de débit de transmission des mobiles peut permettre de contrôler et limiter ce bruit et d'améliorer les performances globales 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 désormais utiliser des débits 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 section 6.3.3. Or cette qualité est une fonction des puissances d'émission des différentes sources. Pour en comprendre le principe, considérons une salle fermée dans laquelle plusieurs personnes parlent simultanément. On entend d'autant mieux son interlocuteur que celui-ci 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 à 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, mais 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 des débits ainsi obtenue n'est pas Pareto-optimale. Par un judicieux contrôle joint des ressources, 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 [67, 102]. 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, on 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, comme dans la partie précédente de cette thèse, du fait de la diversité des applications utilisées nous proposons de les appliquer non aux débits directement, mais aux utilités qui leur sont associées.

6.1.2 Un réseau, différents liens.

Nous donnons dans cette section 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 surfaces circulaires ou hexagonales. 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° , comme illustré à la Figure 6.1. On parle alors de trois *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).

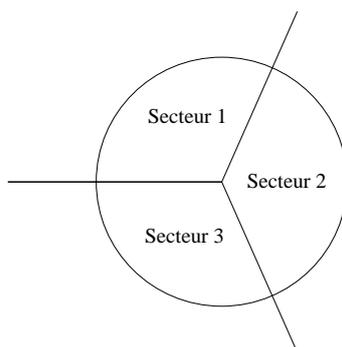


FIG. 6.1 – Un exemple de cellule avec 3 secteurs

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 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 cas du troisième terminal. En fait, afin d'éviter les «trous de couverture», les zones se recouvrent généralement légèrement. 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», on lui offre la possibilité de communiquer avec deux stations, afin d'augmenter les chances de succès de la communication. On parle alors de *handover*, ou *macro-diversité*. Nous y reviendrons dans la section 7.3 du prochain chapitre. Nous allons développer un modèle

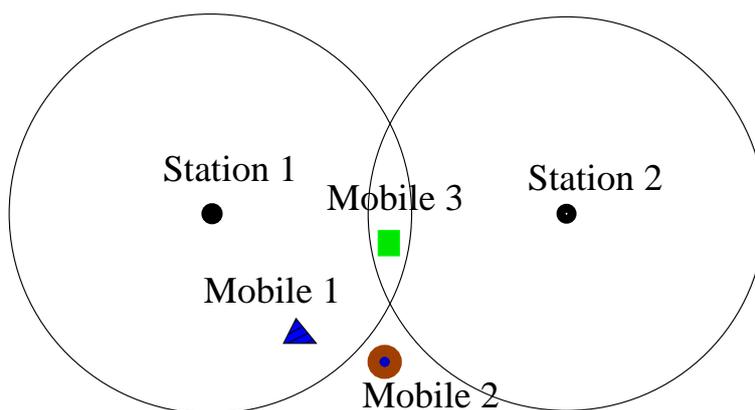


FIG. 6.2 – Un exemple de petit réseau sans fil avec 2 cellules

mathématique pouvant 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 (en particulier sur le plan industriel) et montrons comment elle se situe par rapport aux travaux déjà réalisés dans le domaine. Nous présentons ensuite notre modèle (section 6.3) 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 Intérêt technologique

Nous nous intéressons ici à la question de comment allouer des débits 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 α 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 α , 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 (Équation 1.6) 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 soit le débit total, soit les taux d'admissions, soit un compromis entre les deux.

Notre but 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 de 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 dépendent des constructeurs. Par exemple certains adaptent 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). Notons que ces stratégies sont sous-optimales au sens de Pareto car l'on ne modifie pas les 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 en dehors des solutions extrêmes : taux de rejet important (pour une QoS garantie), ou un taux de rejet faible (avec une QoS non garantie).

Les stratégies optimales et équitables introduites dans le contexte des réseaux fixes 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 et correspondant à une 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, alors la procédure d'admission d'un mobile est donné par l'algorithme de la Figure 6.3.

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 [67] que nous avons généralisé.

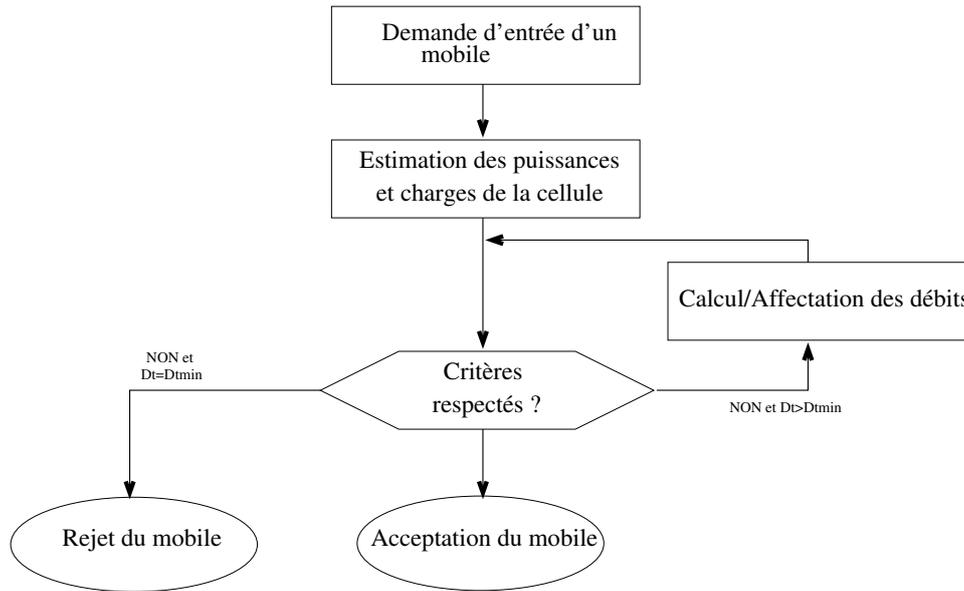


FIG. 6.3 – Algorithme de contrôle d'admission.

Le réseau se compose d'un ensemble \mathcal{N} de connexions. Soit $\overleftarrow{\mathcal{E}}$ l'ensemble des émetteurs et $\overrightarrow{\mathcal{E}}$ celui des récepteurs. Une connexion $n \in \mathcal{N}$ est une paire source - destination $\{s, t\}$ avec $s \in \overleftarrow{\mathcal{E}}$ et $v \in \overrightarrow{\mathcal{E}}$.

Si l'on étudie le lien montant, alors s désigne un mobile et t une station de base. Les rôles sont inversés dans le lien descendant.

Nous montrons dans la suite que l'optimisation équitable et optimale est un problème joint entre deux ressources : le débit (\overrightarrow{r}) et la puissance (\overrightarrow{p}) alloués. Les grandeurs \overrightarrow{r} et \overrightarrow{p} sont des vecteurs de taille $|\mathcal{N}|$ (le nombre de liens $\{s, t\}$) tel tout élément r_n (respectivement p_n) représente le débit instantané (respectivement la puissance instantannée) de la connexion $n, n \in \mathcal{N}$.

Dans la suite de cette section, nous présentons la fonction d'optimisation correspondant à l'allocation équitable des débits dans le contexte des réseaux mobiles (section 6.3.1) puis les contraintes existantes sur les vecteurs de puissances (section 6.3.2). Nous verrons dans la suite (sections 6.3.3 et 6.3.4) comment les débits et les puissances alloués sont liés. Enfin, nous terminerons cette section par l'énoncé du problème d'optimisation d'allocation équitable et optimale (section 6.3.5).

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

(Equation 1.6) 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 connexion $n = \{s, t\}$ 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 [11, 37] et les références incluses).

6.3.2 Modèle de puissance

Chaque source 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 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 $p_{\{s,t\}}$ la puissance du signal émis en vu du receveur 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 $\vec{P}_{tot} = (p_{tot,s})_{s \in \mathcal{E}}$.

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. Dans la suite du manuscrit, $SIR(\{s, t\})$ désigne le *SIR* exprimé au niveau du récepteur t du signal émis par s .

On peut alors montrer qu'un récepteur t peut correctement décoder le signal issu de l'émetteur s si la relation suivante est vérifiée (avec n la connexion $\{s, t\}$) :

$$\delta_n \leq \frac{SIR(n)}{r(n)}, \quad (6.2)$$

impliquant que $\delta_n r(n) \leq SIR(n)$.¹ $\delta_{\{s,t\}}$ 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.

¹Pour obtenir cette équation, nous avons implicitement supposé que le rapport entre l'énergie par bit

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 indépendants², 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 connexions. Ainsi, on notera indifféremment une connexion 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 macro-diversité, on peut toujours écrire le rapport signal sur bruit 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 :

- $N(\{m, b\}) > 0$ 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 macro-diversité).

et la densité spectrale de puissance d'interférence, notée $(E_b/I_o)_s$ ne dépend pas du taux de transmission $r(s)$. C'est une hypothèse courante dans la littérature de la modélisation, voir par exemple [67]. En pratique, cependant il peut dépendre de $r(s)$, voir par exemple [54, p. 151, 222, 239]. Mais, comme montré dans [54, Fig. 10.4, p. 222], il est proche d'une constante sur une grande plage de valeurs de débits. 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 $(E_b/I_o)_s$ sa moyenne sur l'intervalle $[MR_m, PR_s]$. Cependant, si une dépendance exacte est disponible analytiquement, elle peut être incluse dans notre modèle.

²Car les fréquences utilisées dans les deux sens sont différentes.

6.3.5 Résumé : formulation du problème d'équité

Nous achevons cette section en résumant 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 nous avons 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 (montant et descendant) 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 problème d'optimisation de la section 6.3.5. 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, nous transformons le système de contraintes Π^c par un ensemble de contraintes sur un seul vecteur.

6.4.1 Problème équivalent et concavité de la fonction objectif

On remarque que le système de contraintes (6.4) n'est pas convexe, ce qui rend la résolution ardue. Nous proposons ici (section 6.4.1.1) une formulation équivalente du problème basée sur un simple changement de variables, qui permet de rendre le système de contraintes convexe. En revanche, la fonction objectif obtenue n'est plus alors nécessairement concave. Nous montrons cependant dans la 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 $(\vec{p}, \vec{\rho}) \in \widehat{\Pi}^c = (P, \rho)$ qui maximise $Z(\vec{\rho}) := \sum_{n \in \mathcal{N}} \frac{1}{1 - \alpha} \left(\frac{\rho(n)}{1 - \delta_n \rho(n)} \right)^{1 - \alpha}$ où $\widehat{\Pi}^c = (P, \rho)$ 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 Z 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 $\rho(n)$, pour toute connexion 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. La condition $\alpha \geq 2$ est 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 [10]).

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)). Le vecteur \vec{p} étant proportionnel, composante par composante, au vecteur \vec{p}' , on peut raisonner indifféremment sur les vecteurs \vec{p} ou \vec{p}' . L'ensemble des vecteurs \vec{p}' acceptables correspondant est alors noté $P'(\vec{r})$ (respectivement $P'(\vec{\rho})$).

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 afin de découpler les équations en \vec{p}' et en $\vec{\rho}$.

Dans la première sous-section nous montrons que pour chaque 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 n une connexion, $n = \{m, b\}$. 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 :

$$C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\}) < \frac{C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}{C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}} = 1. \quad \blacksquare$$

Nous pouvons en déduire le lemme suivant :

Lemme 6.4. *Soit $\vec{\rho}$ un vecteur acceptable. Considérons le système linéaire suivant :*

$$\delta_{\{m, b\}} \rho(\{m, b\}) = \frac{p'_{\{m, b\}}}{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} p'_{\{m', b\}}}. \quad (6.6)$$

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 $(\vec{\mathbb{P}}, \vec{\rho}) \in \widehat{\Pi}^c$.

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 connexions n de 1 à $|\mathcal{N}|$. Alors on peut écrire les équations sous forme matricielle $AX = Y$ avec

- $Y = (\delta_n \rho_n N(n))_n, n \in \llbracket 1, |\mathcal{N}| \rrbracket$,
- $X = (p'_n)_n, n \in \llbracket 1, |\mathcal{N}| \rrbracket$
- $A = Id_M - B$
- Id_M la matrice d'identité de taille M
- B la matrice :

$$B = \begin{pmatrix} b_1 & \dots & b_1 \\ b_2 & \dots & b_2 \\ & \dots & \\ b_M & \dots & b_M \end{pmatrix}. \quad (6.8)$$

- et $\forall i, b_i = C \delta_i \rho_i$.

Alors, si U est un vecteur propre de A associé à la valeur propre λ , $AU = \lambda U = U - BU$. Ainsi U est un vecteur propre de B avec la valeur propre $1 - \lambda$. Mais $\text{rank}(B) = 1$ et $\text{trace}(B) = \sum_i b_i$. Alors A a uniquement deux valeurs propres qui sont : 1 et $1 - C \sum_i \delta_i \rho_i$.

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. \blacksquare

Nous pouvons alors énoncer le résultat majeur de cette section, nous permettant de définir le vecteur de puissance minimal associé à un taux de transmission acceptable donné.

Proposition 6.1. *A chaque vecteur de taux de transmission fixé \vec{r} (respectivement $\vec{\rho}$) correspond un unique vecteur de puissance minimale \vec{P}'^{min} composante par composante dans l'ensemble $P'(\vec{r})$. De plus $\vec{P}'^{min} = \vec{\mathbb{P}}'$, 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 [10] 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{P}'^{min} satisfaisant $\forall i \in \mathbb{N}, \vec{P}'^{min} \leq (P')^i$ composante par composante et

$$\forall n \in \mathcal{N}, n = \{m, b\}, P'_{\{m,b\}}^{min} = \delta_{\{m,b\}r}(\{m, b\}) \left(N(\{m, b\}) + C \sum_{\substack{m' \in \mathcal{M}_m \\ m' \neq m}} P'_{\{m',b\}}^{min} \right). \quad \blacksquare$$

Nous avons donc montré dans cette sous-section que si \vec{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 nouvelle expression de l'ensemble de contraintes Π^c ne dépendant que de $\vec{\rho}$. 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, $\vec{\rho}$.

6.4.2.2 Conditions nécessaires et suffisantes sur les vecteurs de débits et puissances minimales associées

Lemme 6.5. Soit $\vec{\rho}$ un vecteur de taux acceptable fixé. L'ensemble des allocations de puissances faisables $P'(\vec{r})$ satisfait $\forall \vec{p}' \in P'(\vec{r}), \forall n \in \mathcal{N}, n = \{m, b\}$:

$$\delta_n \rho(n) \left[C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\}) [N(\{m', b\}) - N(\{m, b\})] + N(n) \right] \leq \left(1 - C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\}) \right) p'_{\{m, b\}}. \quad (6.9)$$

Démonstration. Considérons une connexion $n = \{m, b\}$ et récrivons 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 \vec{r} acceptable (respectivement $\vec{\rho}$), $\overrightarrow{P'^{\min}}$ est donné par : $\forall n \in \mathcal{N}, n = \{m, b\}$,

$$P'_{\{m, b\}}^{\min} = \delta_{\{m, b\}} \rho(\{m, b\}) \frac{N(\{m, b\}) + C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\}) [N(\{m', b\}) - N(\{m, b\})]}{1 - C \sum_{m' \in \mathcal{M}_m} \delta_{\{m', b\}} \rho(\{m', b\})}. \quad (6.11)$$

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) $0 \leq p_{tot,s}^{min} \leq \overline{p_{tot,s}}$ avec $s = \begin{cases} m & \text{dans le lien montant,} \\ s & \text{dans le lien descendant} \end{cases}$ et $\overrightarrow{p_{tot}^{min}}$ associé à $\overrightarrow{P'^{min}}$, défini dans l'équation (6.11).

Démonstration. Si $\overrightarrow{\rho}$ est acceptable, alors (C1) est vérifié. La proposition 6.1 affirme que si 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). ■

Synthèse

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 6.4.1.2).

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 \overrightarrow{P} lui correspond), alors tous les vecteurs de puissance correspondants sont plus grands (composante par composante) que le vecteur de puissance correspondant à $\overrightarrow{P'^{min}}$ 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.

6.5 Conclusion

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 celles 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, ce pourquoi ils ne sont pas optimaux.

Nous avons introduit dans ce chapitre un modèle mathématique représentant les contraintes présentes dans ces réseaux. 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 respectivement.

Nous avons vu dans le chapitre précédent que la fonction d'utilité Z n'était concave que pour des valeurs de α 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 α 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 base¹. 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 α 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 ν_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 :

$$SIR_{m,b} = \frac{g_{\{m,b\}} p_{\{m,b\}}}{\nu_b + \sum_{\substack{m' \text{ quelconque} \\ m' \neq m}} g_{\{m',b\}} p_{\{m',b\}}}. \quad (7.1)$$

Nous faisons alors l'approximation suivante, qui est très souvent utilisée pour le cas du lien montant (voir par exemple [73]) :

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 :*

$$\sum_{m',b(m') \neq b} g_{\{m',b\}} p_{\{m',b\}} = \lambda \cdot \sum_{m',b(m')=b} g_{\{m',b\}} p_{\{m',b\}}. \quad (7.2)$$

¹Une introduction aux problèmes de bilan de liaison et atténuation de signal se trouve dans le chapitre d'introduction aux réseaux satellitaires, section 8.1.5.

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 [73]). Nous avons utilisé une hypothèse classique sur les signaux interférents nous permettant de ramener le système à l'étude d'une cellule isolée.

Nous avons montré que le système de contraintes pouvait alors s'écrire conformément au modèle du chapitre précédent (système (6.4)). On peut remarquer que l'analyse du cas montant est 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 Application des résultats du modèle

Nous appliquons désormais les résultats obtenus dans l'analyse théorique du chapitre précédent. En particulier, nous obtenons (voir également [73]) :

$$\forall m, n = \{m, b(m)\}, P_n^{\min}(\rho) = \frac{1}{g_n} \left(\frac{\nu_b \delta_n \rho(n)}{1 - (1 + \lambda) \sum_{m', b(m')=b(m)} \delta_{\{m', b(m)\}} \rho(\{m', b(m)\})} \right).$$

Pour toute cellule c , soit b_c la station de base associée et N_c le nombre de mobiles qui s'y trouvent, numérotés de 1 à N_c . Le problème peut alors s'écrire :

Trouver $\vec{\rho}$ sur chaque cellule maximisant

$$Z(\vec{\rho}) := \sum_{m=1}^{N_c} \frac{1}{1 - \alpha} \left(\frac{\rho(\{m, b\})}{1 - \delta_{\{m, b\}} \rho(\{m, b\})} \right)^{1-\alpha} \quad \text{tel que}$$

$$\left\{ \begin{array}{l} \text{(C1)} \quad \forall m, 1 \leq m \leq N_c, \frac{MR_{\{m, b_c\}}}{1 + \delta_{\{m, b_c\}} MR_{\{m, b_c\}}} \leq \rho(\{m, b_c\}) \leq \frac{PR_{\{m, b_c\}}}{1 + \delta_{\{m, b_c\}} PR_{\{m, b_c\}}}, \\ \text{(C2)} \quad (1 + \lambda) \sum_{m=1}^{N_c} \delta_{\{m, b_c\}} \rho(\{m, b_c\}) < 1, \\ \text{(C3)} \quad \forall m, 1 \leq m \leq N_c, \\ \quad \quad \quad 0 \leq \nu_{b_c} \delta_{\{m, b_c\}} \rho(\{m, b_c\}) \leq g_{\{b_c, m\}} \overline{p_{\{m, b_c\}}} \left(1 - (1 + \lambda) \sum_{m'=1}^{N_c} \delta_{\{m', b_c\}} \rho(\{m', b_c\}) \right). \end{array} \right.$$

On remarque que dans ce cas :

Remarque. *L'ensemble des contraintes est un polytope (convexe).*

De plus, si les quantités PR et δ vérifient une condition simple, alors la première condition n'est plus une contrainte du système. En effet :

Remarque. Une condition suffisante pour tous les vecteurs de taux pour satisfaire (C1) est :

$$(1 + \lambda) \sum_{m=1}^{N_c} \frac{\delta_{m,b_c} PR_m}{1 + \delta_{m,b_c} PR_m} < 1.$$

Nous concluons que pour $\alpha \geq 2w$ (et en particulier pour $\alpha \geq 2$, voir le lemme 6.2 et la remarque 6.4.1.2), 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, suivant une approche similaire à celle que nous avons suivi dans le cadre des réseaux terrestres (section 4). Nous ne détaillons donc pas les détails de l'implémentation spécifiques à l'étude des réseaux mobiles dans cette thèse.

On peut finalement noter que pour le cas d'une seule cellule, la solution précédente est une solution exacte (car l'hypothèse 7.1 n'est alors pas utilisée).

7.2 Cas du lien descendant

Nous nous intéressons désormais au lien descendant, c'est-à-dire aux signaux issus de la station de base à destination des différents mobiles. On peut noter que l'étude du lien descendant est moins aisée que celle du lien montant car chaque station émet dans plusieurs canaux simultanément.

L'expression que nous proposons est plus précise que dans [53] car notre modèle d'interférence est plus complet. De plus, les auteurs utilisent plusieurs hypothèses que nous n'imposons pas ici.

Comme dans la section précédente, nous supposons que chaque mobile ne reçoit des signaux que d'une seule station de base. On note comme précédemment $b(m)$ la station de base associée au terminal m .

La première sous-section est consacrée à l'étude du système et aux hypothèses faites et montre que, là 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 cas du lien descendant.

7.2.1 Notations et hypothèses

Nous présentons dans cette sous-section notre modélisation des contraintes de débits et de puissance sur le lien descendant. 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 mentionné, ce calcul est plus complexe que dans le cas du lien montant à cause de la multiplicité des canaux qu'utilise chaque station de base et de leurs différents statuts. Enfin, nous émettrons une hypothèse simple sur la puissance des interférences inter et intracellulaire, comparable à celle faite dans le cas du lien montant, nous permettant de nous ramener à l'expression utilisée dans le modèle du chapitre précédent.

7.2.1.1 Puissance d'émission

Chaque station de base utilise un certain nombre de canaux. On note alors :

- $P_{\{b,m\}}$ la puissance de transmission à la station de base b pour le mobile m (en général appelé Dedicated Physical Channel (DPCH)),
- $P_{SCH,b}$ la puissance du canal de synchronisation de la station b ,
- $P_{CCH,b}$ la puissance du canal commun de la station b

La puissance totale émise par une station de base b s'écrit alors tout naturellement :

$$P_{tot,b} = \sum_{m',b(m')=m} P_{\{b,m'\}} + P_{CCH,b} + P_{SCH,b}.$$

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 $g_{m,b}$. La raison en est que les atténuation du signal sont dues principalement aux obstacles se situant entre la station de base et le terminal mobile. Or ce type d'atténuation dépend de la fréquence utilisée, qui est différente dans le cas du lien montant ou descendant.

7.2.1.3 Première expression du rapport signal sur bruit

On note comme précédemment ν_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 de la cellule (c'est-à-dire $b(m)$) à la fois aux autres mobiles de la station et sur les canaux communs ;
- 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 β le facteur de synchronisation. On a bien évidemment $0 \leq \beta \leq 1$. Plus β 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. Ainsi dans la pratique :

- L'interférence dues aux signaux envoyé par $b(m)$ 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 [53], 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 [53] d'autres hypothèses sont effectuées que nous ne considérons pas dans notre modèle.

7.2.2 Application des résultats du modèle

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 précédente, on peut résoudre le problème d'optimisation au niveau de chaque cellule. Soit b la station de base correspondante. En appliquant les résultats de la section 6.4, nous obtenons le problème d'optimisation suivant :

Trouver $\vec{\rho}$ sur chaque cellule maximisant

$$Z(\vec{\rho}) := \sum_{m=1}^{N_c} \frac{1}{1 - \alpha} \left(\frac{\rho(\{m, b\})}{1 - \delta_{\{m,b\}} \rho(\{m, b\})} \right)^{1-\alpha} \quad \text{tel que}$$

$$\left\{ \begin{array}{l} \forall m, b(m) = b, \frac{MR_{\{m,b\}}}{1 + \delta_{\{m,b\}} MR_{\{m,b\}}} \leq \rho(\{m, b\}) \leq \frac{PR_{\{m,b\}}}{1 + \delta_{\{m,b\}} PR_{\{m,b\}}}, \\ C_D \sum_{m, b(m)=b} \delta_{\{m,b\}} \rho_{\{m,b\}} < 1, \\ 0 \leq p_{SCH,b} + p_{CCH,b} + \sum_{m, b(m)=b} \delta_{\{m,b\}} \rho(\{m, b\}) \times \\ \frac{N_D(\{m, b\}) + C_D \sum_{m', b(m')=b} \delta_{\{m',b\}} \rho(\{m', b\}) [N_D(\{m', b\}) - N_D(\{m, b\})]}{1 - C_D \sum_{m', b(m')=b} \delta_{\{m',b\}} \rho(\{m', b\})} \leq \overline{p}_{tot,b}. \end{array} \right. \quad (7.4)$$

Le système de contraintes ne parait pas linéaire. Cependant, en y regardant de plus près, on s'aperçoit que pour n'importe quelle valeur de δ , ρ , N et 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 c couverte par une station de base 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 :

$$\text{Trouver } \vec{\rho} \text{ qui maximise } Z(\vec{\rho}) := \sum_{m=1}^{N_b} \frac{1}{1-\alpha} \left(\frac{\rho(\{m, b\})}{1 - \delta_{\{m, b\}} \rho(\{m, b\})} \right)^{1-\alpha} \text{ tel que}$$

$$\left\{ \begin{array}{l} \frac{MR_{\{m, b\}}}{1 + \delta_{\{m, b\}} MR_{\{m\}}} \leq \rho(\{m, b\}) \leq \frac{PR_{\{m, b\}}}{1 + \delta_{\{m, b\}} PR_{\{m, b\}}}, \\ (1 + F)\beta \sum_{m=1}^{N_b} \delta_{\{m, b\}} \rho(\{m, b\}) < 1, \\ \sum_{m=1}^{N_b} \delta_{\{m, b\}} \rho(\{m, b\}) N_D(\{m, b\}) \leq (\overline{p}_{tot, b} - p_{SCH, b} - p_{CCH, b}) \left(1 - C_D \sum_{m'=1}^{N_b} \delta_{\{m', b\}} \rho(\{m', b\}) \right). \end{array} \right.$$

Nous avons alors de nouveau un problème de maximisation avec des contraintes linéaires, et pour $\alpha \geq 2w$ une fonction objective est concave. Ce système peut donc être aisément résolu. Comme pour le lien montant, 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. Malgré cela, le système de contrainte reste linéaire, permettant une résolution aisée, et donc une implémentation possible dans les terminaux.

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 utilisent le *hard-handover*. Cela signifie que lorsque la puissance de signal entre un mobile et la station de base devient inférieure à un certain seuil, la communication bascule alors vers une seconde station de base, dont le signal est plus clair (c'est-à-dire de plus grand *SIR*) si une telle station est disponible.

Ce type de technique permet ainsi 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 car il n'existe alors aucune station de base dont le signal reçu aura un *SIR* satisfaisant et la qualité de la communication risque d'être dégradée.

Dans les systèmes tels que l'UMTS, on utilise désormais le *soft-handover*. Il permet à un terminal d'avoir un lien actif avec deux stations de bases simultanément, ou deux

secteurs d'une même station. Chaque mobile en limite de couverture peut alors écouter les signaux des deux stations de base et les reconstituer.

On ne considère actuellement l'utilisation du soft-handover que dans le cas descendant. Ainsi, 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 macro-diversité dans le lien montant car cela requièrerait une collaboration entre les stations, ce qui est en pratique difficile.

7.3.0.1 Notations et hypothèses

Nous étendons le modèle développé dans le cadre du lien descendant afin de prendre en compte la macro-diversité. Au sein de chaque cellule, les valeurs des puissances des signaux interférents sont les mêmes que ceux précédemment vus. En revanche, les contraintes sur la qualité de signaux reçus au niveau des mobiles sont différentes. Il nous faut donc ré-écrire l'expression des SIR pour les mobiles en limite de couvertures.

SIR d'un terminal en situation de soft-handover

Considérons un mobile k ayant un lien actif avec deux stations de base ou secteurs l et s . Les rapports signal sur bruit s'ajoutent [53] et l'on note $SIR_{\{k\}}$ leur somme. Le signal peut alors être correctement reçu si

$$\delta_{\{k\}}r(\{k\}) \leq SIR_{\{k\}} = SIR_{\{k,l\}} + SIR_{\{k,s\}}.$$

Supposons que le lien avec la station s soit de meilleure qualité, c'est-à-dire que $SIR_{\{k,l\}} \leq SIR_{\{k,s\}}$. Posons 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 suivante :

Hypothèse 7.3. *La valeur de Δ_k ne dépend pas de k .*

Pour ce faire, en pratique nous proposons de prendre comme valeur de Δ la valeur moyenne des Δ_k sur l'ensemble des mobiles qui sont en soft handover dans le système.

SIR des mobiles dépendants d'une station de base donnée

Nous séparons alors l'ensemble des mobiles qui sont en communication avec la station de base b en trois groupes. Soient :

- I l'ensemble des mobiles de la cellule qui ne font pas de handover. Pour de tels mobiles nous avons :

$$SIR_{\{i\}} = SIR_{\{i,b\}} \text{ avec } SIR_{\{i,b\}} \text{ donné par la formule (7.3).}$$

- J les mobiles en situation de soft handover dont le meilleur lien est celui avec la station de base b . Alors,

$$SIR_{\{j\}} = (1 + \Delta_j)SIR_{\text{best link}} = (1 + \Delta_j)SIR_{\{j,b\}}. \text{ De nouveau, la valeur de } SIR_{\{j,b\}} \text{ est donné par la formule (7.3).}$$

- K l'ensemble des mobiles connaissant également le soft handover, mais dont le plus mauvais lien est celui avec la station b . Pour ceux la :

$$SIR_{\{k\}} = \frac{\Delta_k}{1+\Delta_k}SIR_{\text{worst link}} = \frac{\Delta_k}{1+\Delta_k}SIR_{\{k,b\}}.$$

Notons enfin que l'équation (6.1) s'écrit dans ce cas :

$$P_{tot,b} = \sum_{i \in I} P_{\{i,b\}} + \sum_{j \in J} P_{\{j,b\}} + \sum_{k \in K} P_{\{k,b\}} + P_{CCH,b} + P_{SCH,b}.$$

7.3.1 Application des résultats du modèle

Nous pouvons ainsi appliquer le modèle présenté au chapitre précédent en posant :

$$\begin{cases} N_I(b, m) = N_D(b, m), & C_I = C_D \\ N_J(b, m) = \frac{1}{\Delta+1} N_D(b, m), & C_J = \frac{1}{\Delta+1} C_D \\ N_K(b, m) = \frac{\Delta}{\Delta+1} N_D(b, m), & C_K = \frac{\Delta}{\Delta+1} C_D \end{cases}$$

avec N_D tel que défini dans la section 7.2.

Problème d'optimisation correspondant.

Soit b une station de base associée à une cellule c . Pour tout mobile m , soit $i(m)$ l'indice de l'ensemble auquel il appartient ($i(m) \in \{I, J, K\}$).

Le problème d'optimisation associé au lien descendant avec macro-diversité s'écrit alors

Trouver $\vec{\rho}$ sur chaque cellule qui maximise

$$Z(\vec{\rho}) := \sum_{m=1}^{N_b} \frac{1}{1 - \alpha} \left(\frac{\rho(\{m\})}{1 - \delta_{\{m\}}\rho(\{m\})} \right)^{1-\alpha} \text{ tel que}$$

$$\begin{cases} \frac{MR_{\{m,b\}}}{1 + \delta_{\{m\}}MR_{\{m,b\}}} \leq \rho(\{m\}) \leq \frac{PR_{\{m,b\}}}{1 + \delta_{\{m\}}PR_{\{m,b\}}}, \\ \sum_{m=1}^{N_b} C_{i(m)}\delta_{\{m\}}\rho(\{m\}) < 1, \\ \sum_{m=1}^{N_b} \delta_{\{m\}}\rho(\{m\})N_{i(m)}(\{m, b\}) \leq (\overline{p}_{tot,b} - P_{SCH,b} - P_{CCH,b}) \left(1 - C_{i(m)} \sum_{m'=1}^{N_c} \delta_{\{m'\}}\rho(\{m'\}) \right). \end{cases}$$

Discussion

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.

Le système de contraintes bien que simpliste est en fait plus général que ce que nous avons pu trouver dans la littérature. Par exemple, si nous regardons la méthode utilisée dans [53], 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 Δ_k était indépendant de k . En fait, on peut, si l'on désire avoir une image plus fine de la situation du réseau, discrétiser les valeurs possibles de Δ_k . Cela peut se faire simplement en sous-divisant les ensembles J et K en J_p et $K_p, 1 \leq p \leq n$ 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 considérer 3 ou 4 valeurs types de Δ_k pour représenter les divers scénarios. On se rend compte qu'augmenter le nombre possibles de valeurs de Δ n'augmente pas la complexité du problème d'optimisation mais celui de l'implémentation pratique. En effet, cela nécessite une mesure précise de tous les rapports signal sur bruit des terminaux en situation de soft-handover, ce qui est peu 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 α 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 $(\overrightarrow{p^{app}}, \overrightarrow{\rho^{app}})$ obtenue par cette méthode d'approximation est donc une paire valide pour le problème initial.

Enfin, la valeur obtenue par cette approximation $Z(\overrightarrow{\rho^{app}})$ est sous optimale, c'est-à-dire que $\exists(\overrightarrow{p}, \overrightarrow{\rho}) \in \widehat{\Pi}^c, Z(\overrightarrow{\rho}) \geq Z(\overrightarrow{\rho^{app}})$. 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 1².

7.4.2 Second schéma : 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 \overrightarrow{r} et non $\overrightarrow{\rho}$, 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 [75] qui considère le case d'une cellule unique). Notre approximation consiste alors à remplacer la dernière contrainte dans (6.4) par :

$$\delta_{\{m,b\}}r(\{m,b\}) \leq \frac{g_{\{b,m\}}P_{\{m,b\}}}{N_{\{m,b\}} + \sum_{m' \in \mathcal{M}_m} g_{\{m',b\}}P_{\{m',b\}}}. \quad (7.5)$$

Cela vient à remplacer le terme $\sum_{m' \in \mathcal{M}_m, m' \neq m} g_{\{m',b\}}P_{\{m',b\}}$ 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 $(\overrightarrow{p^{app2}}, \overrightarrow{r^{app2}})$ satisfait les contraintes de ce schéma d'approximation, alors il satisfait également les contraintes du système initial.

Nous avons étudié ce schéma d'approximation dans [124] dans le cas du lien montant.

L'approximation est raisonnable quand le système contient un grand nombre de mobiles. En effet, elle consiste à négliger la puissance reçue d'un signal devant celle de tous les autres. Dans un système où les puissances reçues sont relativement homogènes³ 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 donné 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.

²Ceci est en pratique vérifié lorsque le système contient un nombre raisonnablement grand de mobiles. En effet, rappelons que $\sum_n \delta_n \rho_n \leq 1$.

³C'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 $\frac{g'p'}{g''p''} \ll 1$.

7.4.3 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 il 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.

7.5 Quelques résultats numériques et conclusion

Résultats numériques.

Nous avons implémenté les solutions correspondants au lien montant pour le problème exact et approché (avec le second schéma d'approximation, nous permettant ainsi d'obtenir des solutions pour le cas $\alpha \rightarrow 0$). Nous présentons sur 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. Les mobiles sont modélisés par des cercles centrés à la position géographique du terminal et de rayon et proportionnel à la puissance (ou la bande passante) allouée.

Pour chacun des N mobiles, nous avons fixé les valeurs des paramètres à $\delta = 1$, $\nu = 1$, $\bar{p} = 1/N$, $MR_i = 1/4N$, et $PR_i = 1$. La position des mobiles est choisie aléatoirement dans le carré $[-1; +1] \times [-1; +1]$ et le gain g du signal est égal à $1/d^2$, où d est la distance au centre du carré (c'est-à-dire à la station de base). 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 légères 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 quelques 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

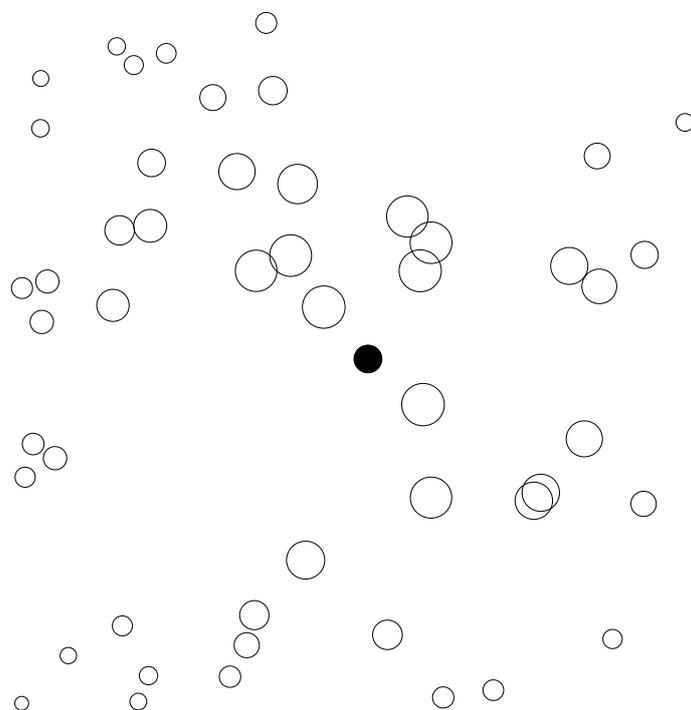


FIG. 7.1 – Allocation de la bande passante pour $\alpha = 0$.

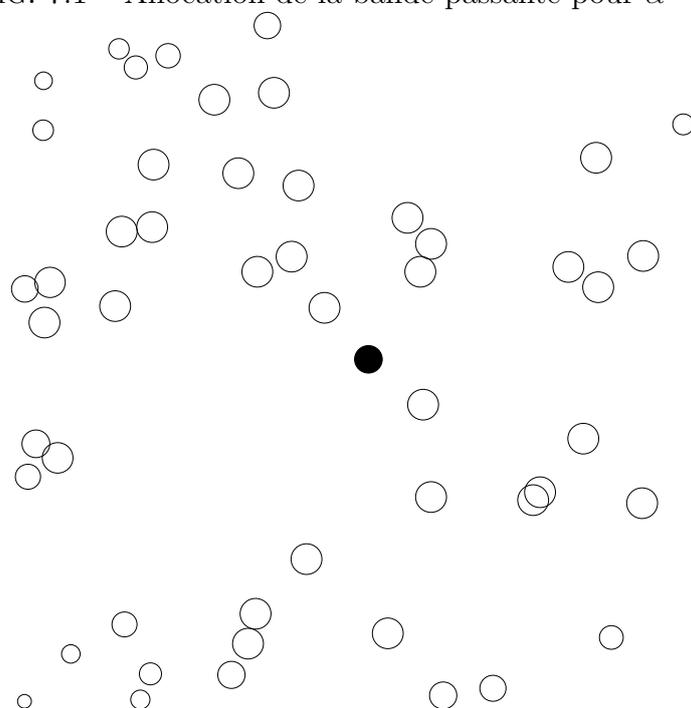


FIG. 7.2 – Allocation de la bande passante pour $\alpha = 1$.

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'équations linéaires, et qu'il est ainsi possible de les résoudre aisément, par exemple par l'utilisation de la SDP.

Enfin, nous avons vu précédemment que pour les faibles valeurs de α , la fonction d'optimisation n'était pas convexe. Nous avons donc proposé dans ce chapitre deux schémas d'approximations possibles pour traiter les valeurs de α arbitrairement proches de zéro.

Enfin, nous avons présenté 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 α sont proches des allocations optimales.

Conclusion

Nous nous sommes intéressés dans cette partie à l'allocation jointe des débits et des puissances d'émissions dans les réseaux sans fil de façon à la fois équitable et optimale. Nous avons présenté un modèle valide aussi bien pour le lien montant que descendant. Nous avons mis en valeur les similitudes entre les deux types de liens 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 macro-diversité dans le lien descendant. (Rappelons que l'idée de la macro-diversité 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.)

Notre système peut être résolu de façon efficace pour les valeurs de α 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, grâce au contrôle des puissances d'émissions des mobiles notre méthode augmente considérablement l'efficacité totale des systèmes par rapport aux algorithmes déjà implémentés dans le cadre des réseaux UMTS. La relation entre puissance et débit n'est pas aisée à modéliser car leur interaction est double :

- D'une part le débit que peut utiliser chaque émetteur est proportionnel à la puissance qu'il utilise
- D'autre part, une augmentation de chaque puissance d'émission accroît la quantité d'interférence des autres connexions et donc limite leur débit de chacun.

Nous avons 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.

Nos travaux font suite à beaucoup de références [65, 85, 94, 128] qui considèrent l'allocation équitable du débit lui-même dans d'autres contextes de télécommunications. Une question sur laquelle nous ne nous penchons pas est celle de savoir comment 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 a été très bien étudiée, voir par exemple [14, 90, 82].

Nous ne décrivons pas non plus la méthode de construction des matrices SDP utilisées dans ce contexte, celle-ci étant sensiblement identique à celle utilisée dans les réseaux terrestres.

On peut noter que nous avons dans cette partie présenté les concepts d'équité appliqués directement aux débits des connexions. En fait, il n'existe pas de problèmes (sur le plan de l'implémentation) à considérer les utilités associées, comme nous l'avons fait dans le cadre des réseaux terrestres.

Les voies d'extensions de ses travaux sont multiples. Tout d'abord, il serait intéressant d'étudier plus avant les performances de nos méthodes d'approximations pour les faibles valeurs du paramètre d'équité.

De plus, des travaux d'implémentations sont en cours, dans le cadre de travaux menés en collaboration avec France Telecom R&D. L'idée est d'utiliser notre critère pour la gestion en temps réel des 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 déjà présents. 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 sous-optimale. Nous tentons de l'améliorer en utilisant cette technique d'allocation puis en cherchant s'il est possible d'augmenter le débit de l'une des connexions sans violer les contraintes du système. On voit alors que la difficulté du problème réside dans la détermination de la connexion en question.

Troisième partie

Les réseaux satellites

Introduction

Cette dernière partie s'intéresse aux communications par satellites. 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 de ces derniers, 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 natures 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.

Les principales différences avec le système étudié dans la partie précédente sont de trois types. Tout d'abord, 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.

De plus, 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 uniquement entre les différentes zones géographiques dans lesquelles ils évoluent.

Enfin, 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, chacun faisant l'objet d'un chapitre consacré. 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*. 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.

Ce chapitre introductif se compose de trois sections. La première est une présentation générale sur les réseaux satellitaires, fortement inspirée du livre [6]. Le but est de comprendre quels sont les spécificités de tels systèmes, par rapport à ceux déjà présentés dans cette thèse.

La seconde section 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, dont nous donnerons la définition en sous-section 8.1.6).

Enfin, nous présentons dans une troisième section 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 afin 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 et 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'inter-connecter 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 multimédia 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 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. (*Petit rappel sur les ellipses*) On rappelle qu'une ellipse (figure 8.1) 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.

Ceci décrit des orbites parfaites. En fait, deux types de perturbations viennent modifier ces trajectoires : les **perturbations gravitationnelles** (dues au volume de la terre, à la présence d'autres astres à proximité du satellite, aux marées, etc) et les **perturbations non gravitationnelles** (dues 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.

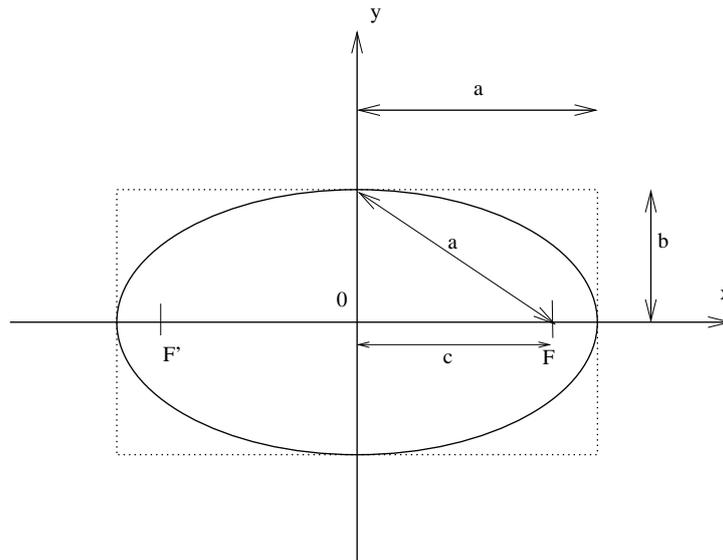


FIG. 8.1 – Allure générale d'une ellipse.

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 couverture d'un satellite est donc un disque sur la surface terrestre qui 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 le satellite paraît près de l'horizon, et plus la réception est mauvaise.

8.1.2 Altitudes possibles

Il existe de fortes contraintes sur les altitudes possibles auxquelles gravitent les satellites, dues à des phénomènes physiques.

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 ainsi 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 zone HEO, pour les satellites à orbite circulaire à 35786 km d’altitude. Les satellites 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 du développement des satellites à orbite basse est donc de réduire les délais de communications.

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 dans les applications quotidiennes sont :

- AM autour de 1 MHz,
- HF autour de 10 MHz,
- VHF autour de 100 MHz,
- UHF de 300 MHz à 900 MHz.

Quand à elles, 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 bande X était traditionnellement utilisée par les applications militaires.

Ainsi, 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 Spécificités des systèmes GEO

Les systèmes géostationnaires sont caractérisés par de grands délais de transmissions. De plus, ils gardent une position fixe au dessus de l'équateur et 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 en pratique que toute position qui a une latitude de plus de 75 degrés. Cependant, même à des latitudes comprises entre 45 et 75 degrés, l'élévation du satellite est faible 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, les villes de Paris, Londres, Berlin et Moscou sont toutes comprises dans cette zone. Il est possible de remédier à ce problème en utilisant des orbites légèrement inclinées les unes par rapport 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. Ces systèmes ont cependant l'inconvénient majeur d'utiliser trois fois plus de satellites.

D'autres difficultés se posent dans les systèmes géostationnaires. 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. De plus, 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 Bilans de liaison

Chaque satellite contient un certain nombre d'*antennes* (2 par exemple 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 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 ré-amplifié et démodulé. Le rapport P_r/P_e entre la puissance reçue et émise vaut

$$\frac{P_r}{P_e} = \mathcal{G}_r \cdot \mathcal{G}_e \left(\frac{c}{4\pi r F} \right)^2$$

avec \mathcal{G}_e et \mathcal{G}_r les gains des antennes d'émission et de réception, c la vitesse de lumière, F la fréquence en Hz et 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 sensibilité 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 donne la puissance minimale que l'on doit fournir pour tirer parti 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 qu'il faut une augmentation exponentielle du rapport signal/bruit pour obtenir un accroissement linéaire du débit. En fait, il existe des améliorations possibles de la capacité de Shannon : la polarisation permet simplement d'augmenter la capacité d'un facteur 2, tandis d'autres approches sont possibles pour démultiplier la capacité.

8.1.6 Méthodes d'accès

Dans les réseaux satellitaires, de même que pour les réseaux mobiles, chaque 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 filaires.

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 indépendamment. Par conséquent, il peut en résulter des collisions de paquets et donc des pertes.

Dans la 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 à sa propre bande de fréquence où elle peut émettre indépendamment des autres. L'avantage de cette méthode est qu'elle ne nécessite pas de synchronisation temporelle entre les sources. En revanche, cette méthode souffre de problèmes d'inter-modulation 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 alors des «trous» entre les bandes pour améliorer l'atténuation d'autres fréquences.

8.1.6.2 Systèmes TDMA

On définit ici un *cycle*, aussi appelé *trame*, découpé en tranches de temps, appelés *time-slot*. Dans chaque cycle, chaque source transmet pendant exactement un intervalle de temps. Afin de pouvoir allouer une bande passante différente aux sources, on définit aussi une extension de cette méthode, dans laquelle une source peut transmettre pendant plusieurs tranches de temps. La demande de chaque station s'exprime alors en un nombre entier de time-slots. 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, pour une puissance moyenne donnée, si la transmission n'est possible que pendant une fraction a de temps, il est nécessaire d'effectuer la transmission pendant cette période avec une puissance supérieure d'un facteur $1/a$. On définit enfin les systèmes MF-TDMA (Multiple Frequency TDMA) qui sont des systèmes combinés TDMA et FDMA.

8.1.6.3 Systèmes CDMA

Imaginons que l'on veuille transmettre une séquence de bits. 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 $\forall s, t, s \neq t, \langle a^s | a^t \rangle = \sum_{i=1}^N a_i^s \cdot a_i^t = 0$. 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 $b^s \cdot a^s + b^t \cdot a^t$. Alors, $\langle b^s \cdot a^s + b^t \cdot a^t | a^s \rangle = b^s \langle a^s | a^s \rangle + b^t \langle a^t | a^s \rangle = b^s$. De même, $\langle b^s \cdot a^s + b^t \cdot a^t | a^t \rangle = b^t$. Ainsi, bien 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 sinusoïdal de fréquence élevée avant sa transmission dans le canal. Le CDMA a des avantages par rapport au TDMA et FDMA. Il est notamment plus robuste aux phénomènes d'atténuation par des multi-chemins et aux interférences.

8.2 Précédents travaux sur l'allocation de slots dans les systèmes TDMA et MF-TDMA

De nombreux articles traitant des problèmes d'allocation de ressources dans les systèmes TDMA et MF-TDMA sont disponibles 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 rafales. Nous mentionnons ensuite quelques problèmes annexes (section 8.2.3), et citons dans la dernière sous-section des articles présentant des systèmes satellitaires généraux (section 8.2.4).

8.2.1 Le problème TSA (Time Slot Allocation)

Dans un système SS/TDMA (Satellite Switch / Time Division Multiple Access), les stations terrestres communiquent entre elles 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. 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 (TDMA), chaque station demande une quantité différente de temps pour transmettre de l'information à une autre station terrestre, correspondant à un nombre entier de time-slots. 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 rafales (ou en anglais *burst*) (cf. sous-section 8.2.2).

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 est notée $d_{i,j}$, alors 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.

Dans la suite de cette partie, nous nous intéressons tout d'abord à l'énoncé puis les méthodes de résolutions du problèmes TSA ainsi que quelques unes de ces variantes. Nous verrons ensuite deux cas de switches plus complexes, pour lesquels des études spécifiques ont été menées.

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 pour chaque lien. 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 alors nécessairement de taille strictement supérieure à 1.

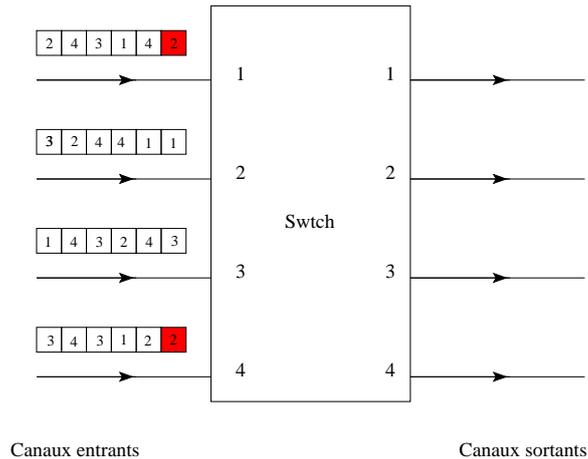


FIG. 8.2 – Un switch TDM classique

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, du spot i au spot 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 l'exemple de la figure 8.2, 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 liens 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-slots d'un même canal montant.

Le problème TSA est alors d'exprimer une matrice de trafic donnée comme somme de matrices de modes 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 (c'est-à-dire d'une configuration du switch à une 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 en matrice de modes 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^{ème} 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 de matrices de trafic en matrice de mode 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 résolution

L'un des plus anciens articles traitant du problème TSA est [55]. 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 [111].

Dans [17], on introduit de nouvelles contraintes dues 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) et 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 par 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. Bongiovanni et al. [17] ont alors présenté 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. A partir de nombreux résultats numériques générés à partir de matrices de trafic aléatoires, ils ont montré qu'en moyenne, le nombre de matrices de mode est sensiblement inférieur à la valeur des bornes.

8.2.1.3 Variantes du problème

Nous mentionnons dans cette section quelques variantes intéressantes du problèmes TSA.

Minimisation du délai moyen

Gopal, Coppersmith et Wong remarquent dans [44] que les résolutions de minimisations de modes 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.

Reprenons l'exemple de la sous-section précédente. 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, ce qui 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, toutes les matrices de modes de la décomposition ne sont pas des permutations. La proposition de Gopal et al. est donc d'utiliser des matrices 0 – 1 les plus «pleines» possibles (c'est-à-dire au nombre de 1 maximal) en début de décomposition.

Limitation du nombre de modes de switch

Dans [77], les auteurs proposent de ne considérer qu'un faible nombre de configuration de switch (en $\mathcal{O}(n)$ au lieu de $\mathcal{O}(n!)$ comme précédemment considéré). Ainsi, le nombre de configurations possibles du switch embarqué est considérablement réduit et donc les délais de changement de configuration le sont 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 alors une borne du temps ainsi perdu.

Minimisation des reconfigurations

Dans [45] les auteurs font l'hypothèse que le temps nécessaire au switch pour se reconfigurer est grand par rapport au temps de mode. 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 suppose le contraire. Il montre d'abord que ce problème est NP-complet et propose une heuristique basée sur la théorie des graphes minimisant le nombre de reconfiguration du switch. L'algorithme trouve tout d'abord l'ensemble des trames avec un minimum de reconfigurations, puis sélectionne celle qui minimise le temps de transmission. L'algorithme est ensuite validé par des expérimentations numériques.

Trafic hétérogène

Bonuccelli, Gopal et Wong considèrent dans [19] un système à deux types de trafic : un trafic de voix, supposé relativement statique et 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 considéré est de construire une nouvelle trame TDMA qui ne change l'allocation d'aucun des time-slots déjà alloué mais alloue autant de paquets que possible dans les slots vides. Les auteurs montrent que ce problème est NP-complet et proposent trois 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. Ces dernières sont toutes basées sur une relaxation Lagrangienne.

Extension au multicast

Dans [30], les auteurs étendent le problème du TSA au système multicast. Le problème multicast peut être résolu en utilisant les techniques vues précédemment si les paquets sont autorisés à être scindés en paquets 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, contenant l'ensemble des receveurs. Pour résoudre le TSA dans le cas multicast, deux heuristiques sont proposées, l'une basée sur une méthode 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 disjoint fait l'objet d'une étude spécifique. On peut noter que même ce cas particulier est NP-complet.

8.2.1.4 TDM hiérarchiques

Le principe des TDM hiérarchiques est de regrouper les utilisateurs finaux utilisant des connexions à faible bande passante sur des multiplexeurs. L'intérêt de tels systèmes 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 8.3.

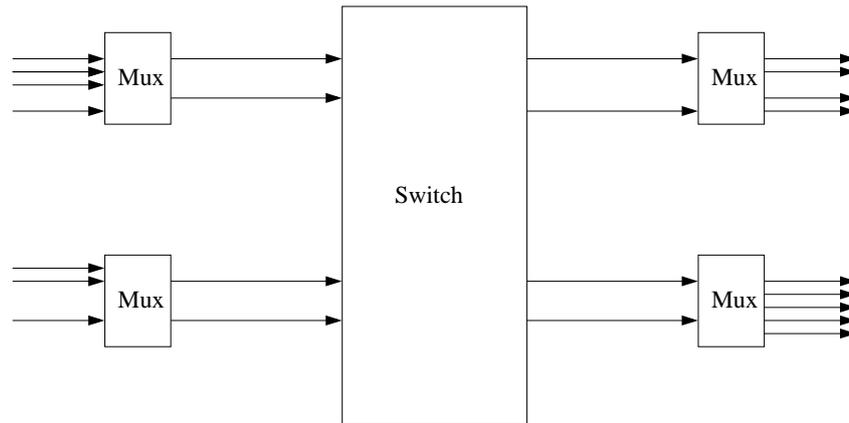


FIG. 8.3 – Un exemple de système TDMA hiérarchique

On peut énoncer des conditions nécessaires et suffisantes sur le trafic des utilisateurs pour affirmer l'existence d'un TSA valide pour acheminer le trafic entre tous les utilisateurs finaux [36].

Dans [18], Bonuccelli s'intéresse au cas symétrique (M utilisateurs 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 de circulation¹ sur ce réseau. Une version améliorée a été proposée par Tham dans [120] qui utilise les propriétés des graphes bipartites². La complexité est alors de $\mathcal{O}(M^4)$. Un article très proche est [119].

De nombreux autres algorithmes ont été développés dans le cadre des switches symétriques de taille N à M utilisateurs en entrée et en sortie. On peut citer les travaux de Fumabiki et Takefuji [40] dont l'heuristique est basée sur l'utilisation de réseaux de neurones. Leur algorithme tourne en parallèle sur $N^2 \times K$ processeurs, où K est le nombre de time-slots du problème.

¹La 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).

²Un 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 autre approche intéressante est celle de Yeung [130], dont l'algorithme itératif a une complexité en $\mathcal{O}(N^2)$ pour le problème non-hiérarchique et en $\mathcal{O}(M^2)$ pour le problème hiérarchique. L'idée est proche de celle de Inukai [55] dans le cadre des systèmes non hiérarchiques : à 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. Les simulations tendent à montrer que les résultats obtenus sont très proches de la solution optimale. Nous illustrons la méthode suivie dans l'exemple suivant (dans le cas simple des systèmes non-hiérarchiques).

Exemple 8.1. (*méthode des éléments critiques*) Soit la matrice de routage suivante :

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

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 maximale. Dans les matrices suivantes, les éléments correspondants apparaissent en italique. La méthode des éléments critiques consiste alors à prendre à chaque étape la matrice de mode contenant des 1 aux coordonnées des éléments marqués 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 bandes passantes différentes

Un autre type de système est appelé système SS/TDMA avec faisceaux de tailles variables (voir Fig. 8.4). L'idée est d'utiliser des faisceaux de bandes passantes α_i différentes. On ne peut malheureusement pas appliquer les méthodes utilisées dans les systèmes TDMA hiérarchiques pour résoudre de tels problèmes. Chalasani et Varma [27] modélisent alors ce problème de TSA comme un problème de flot de réseau, tout comme Bonuccelli dans [18] puis proposent alors deux heuristiques. 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, α 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 $\mathcal{O}((M + N)^3 \log(L))$.

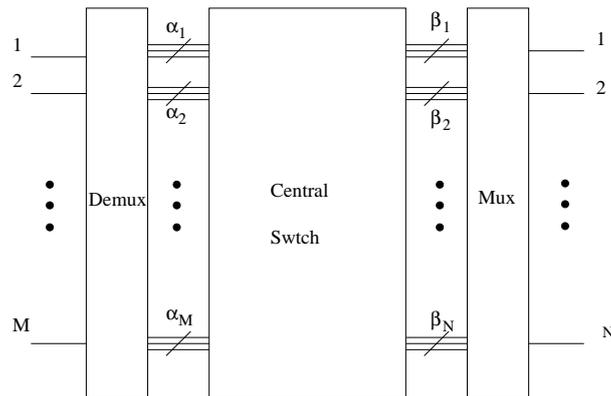


FIG. 8.4 – SS/TDMA avec faisceau de taille variable

8.2.2 Allocation des rafales

On considère ici le problème consistant à allouer aux utilisateurs un ou des time-slots dans le lien montant en rafales (ou par groupes), c'est-à-dire sur des intervalles de temps consécutifs de la trame.

Approche classique

Une solution est proposée dans [106] 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-slots 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. Le problème considéré est de satisfaire les demandes des terminaux tout en s'assurant que les time-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.

En fait, dans les systèmes MF-TDMA, 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'une rafale. Ce problème est considéré dans [57].

Résolution jointe du problème d'allocation de rafales et de TSA.

Une approche intéressante est de considérer les deux problèmes (d'allocation de rafales et de TSA) de façon jointe. C'est ce que propose [119]. Dans le système considéré la zone couverte par un satellite est divisée en *spots*. L'auteur considère 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 (canaux), qui sont différentes pour les liens montants et descendants. La question est de 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.

Plus précisément, l'idée est de regarder la matrice de trafic liant chaque source à chaque destination. Le maximum des sommes des valeurs de chaque ligne et de chaque colonne de la matrice correspondant à chaque spot 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. A chaque étape du processus, on détermine alors 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

Considérons un système SS/TDMA dans lequel un certain nombre d'appels sont en cours. Des rafales de trafic sont alloués aux stations terrestres une à une. Au début de chaque rafale se trouve une entête (overhead) et un intervalle de sécurité sépare deux rafales consécutives. 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 [127] 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 du switch (car elles sont coûteuses en temps) et les effectuer 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

Nous mentionnons ici deux études intéressantes liées aux problèmes d'allocations de slots.

Spots interférents.

En pratique, dans un système SS/TDMA, les spots interfèrent entre eux. Dans [43], les auteurs proposent une approche en deux étapes. Dans la première, ils allouent des polarisations orthogonales aux stations pour réduire les interférences. Ils supposent que les spots interfèrent 2 à 2. Le but de l'allocation est alors d'empêcher 2 zones interférentes d'émettre en même temps. Ils montrent que le problème est solvable en temps polynomial dans la plupart des cas et le modélise par un graphe dont les nœuds sont les spots : il

existe une arrête entre deux nœuds si et seulement si ils interfèrent entre eux. L'allocation proprement dite des time-slots aux stations constitue la seconde phase (le système étant SS/TDMA, il y a donc une porteuse par spot). L'objectif est encore de minimiser la durée de l'allocation sur la trame. L'article propose des solutions heuristiques.

Synchronisation des sources

Dans [23], les auteurs considèrent la difficulté pour les différentes sources de se synchroniser entre elles afin d'émettre leur 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 Descriptions de systèmes complets

Nous terminons notre présentation des précédents travaux sur les systèmes satellitaires en mentionnant quelques articles décrivant des systèmes complets et les défis soulevés. Ils sont de deux types : ceux traitant des différents types de satellites et ceux s'intéressant à un système satellitaire particulier.

8.2.4.1 Études générales des systèmes satellitaires.

On peut citer en particulier l'article [131] qui présente une description sommaire des satellites utilisant la bande Ka classés par pays (cette étude date de 1999). Une autre étude générale intéressante est [15], qui décrit en particulier les atténuations dues aux conditions atmosphériques dans de tels systèmes.

Les problèmes technologiques liés à l'embarquement des systèmes de traitements de signal (en matière de poids, de consommation d'énergie, etc) sont très bien décrit dans [62]. L'étude date malheureusement de 1987 mais présente les évolutions technologiques envisagées à cette époque.

Enfin, on peut noter que Sinha [115] présenta précisément dès 1976 les systèmes satellitaires TDMA. En particulier il y décrit les équipements terrestres et embarqués, et donna une première formulation mathématique du problème d'allocation des rafales et du TSA.

8.2.4.2 Description globale de systèmes

Enfin, quelques articles présentent de systèmes satellitaires complets. Nous en mentionnons quelques uns parmi les plus caractéristique.

Intelsat

Minoux et Brouder présentent dans [91] 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 de 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 par programmation dynamique d'un ensemble de sous-blocs dont la somme est la plus grande possible, tout en restant inférieure à celle du bloc critique,
- 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 [68]. On y trouve en particulier une description de la composition de la trame, des 2 types de rafales (de données et de parole), la méthode pour agréger les flots au niveau des stations émettrices et les équipements utilisés dans le système.

Services RNIS³ et ATM

Dans [63] les auteurs s'intéressent aux choix technologiques nécessaires dans le cas d'un système satellitaire intégrant des services RNIS. Le système offre aux utilisateurs un ensemble de débits entre 64 et 1536 kbps. Les auteurs proposent 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. L'inconvénient 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 était prévu pour la fin de l'année 1991.

³Réseau Numérique à Intégration de Services (ISDN en anglais)

Un autre article intéressant est [133]. Il s'agit ici d'un réseau de type VSAT (Very Small Aperture Terminal) pour des services ATM.

Systeme ACTS

ACTS (Advanced Communications Technology Satellite) [105] est un satellite expérimental, basé sur une combinaison de FDMA et de TDMA. Il est constitué de 2 spots chacun associé à une antenne du satellite. Chaque antenne visite périodiquement une série de zones terrestres. La durée du cycle des visites est égal à celle de la trame, c'est-à-dire de 1ms. Les slots sont regroupés en rafales, 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). L'idée est d'attribuer préalablement à chaque communication un certain nombre de time-slots, et de comparer les façons possibles de ré-allouer des time-slots affectée à une connexion qui ne les utilise pas à une autre connexion qui n'en n'a pas assez.

Autres systemes

Systeme à 2 types de données ou d'utilisateurs

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, se trouve dans [35].

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 vidéo 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 (Multiple Carrier TDMA)

Enfin, [92] présente l'étude d'un système complet MC-TDMA. L'allocation est décomposée en différentes étapes. La première s'intéresse à la question suivante : le trafic entre les différentes zones géographiques 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 TSA 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 alors à un problème de bin-packing.

Enfin, un réarrangement des rafales est nécessaire afin que deux rafales 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 rafales simultanément sur 2 porteuses différentes. Ils proposent pour ce dernier problème une méthode de branch-and-bound.

8.3 Conclusion et choix du système considéré

Nous avons vu que le problème d'allocation des slots a été largement étudié dans la littérature. Cependant, il semble que jusqu'à ce jour les notions d'équité n'aient pas reçues l'attention qu'elles méritent. Les modèles précédemment mentionnés diffèrent les uns des autres sur les hypothèses faites : multiplicité ou non des porteuses, et dans le second cas, faisceaux de tailles différentes ou non, allocation en rafales ou non et finalement la façon dont sont prises en compte des conditions d'interférences.

Nous présentons ici brièvement le modèle physique ainsi que la formulation générale du problème que nous nous proposons d'étudier et commentons brièvement son positionnement par rapport aux travaux précédents.

8.3.1 Système physique

Nous considérons un système géostationnaire MF/TDMA utilisant la bande Ka. Nous décrivons brièvement les hypothèses faites sur l'emprunte du satellite et la façon de prendre en compte les interférences.

8.3.1.1 Spots et réutilisation spatiale

La zone géographique couverte par le satellite est divisée en régions hexagonales appelées *spots*, eux-même divisés en *zones*. Un faisceau différent est utilisé pour chaque spot. La subdivision en zones 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*. Il y a une affectation statique des couleurs aux spots, que nous ne considérons pas ici. Dans la pratique, le nombre de couleurs est inférieur au nombre de spots. Comme deux spots ayant la même couleur interfèrent et que l'interférence est fonction de la distance, l'affectation a pour but d'espacer au maximum les spots de même couleur. Enfin, comme deux spots ayant des couleurs différentes sont associés à des problèmes différents, dans la suite nous calculerons les allocations indépendamment sur chaque groupe de couleurs.

Un exemple d'allocation de couleurs est illustré par la figure 8.5 avec 4 couleurs dans un système de 14 spots.

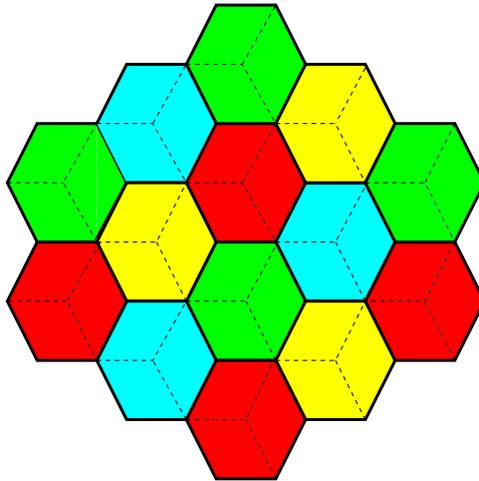


FIG. 8.5 – Exemple d'emprunte de satellite avec 14 spots, 4 couleurs et 3 zones par spots.

8.3.1.2 Zones

La subdivision en zone peut avoir deux rôles différents. Nous considérons donc deux partitionnement en zones, selon le problème considéré.

Atténuation du signal

Un inconvénient de la bande Ka est sa grande sensibilité à la pluie. Or, deux utilisateurs se trouvant à des endroits différents d'un même spot peuvent être soumis à des conditions atmosphériques différentes. D'une façon générale, la localisation géographique des utilisateurs dans un spot peut avoir des conséquences sur la qualité de leur liens. Pour prendre ceci en considération, nous partageons les spots en zones de petites tailles, donc soumises aux mêmes conditions de liens. On peut alors supposer qu'en tout points d'une zone donnée, les opérateurs utilisent le même taux de codage.

Interférences

L'utilisation de couleurs permet évidemment de réduire, mais non de supprimer les interférences. Or la position géographique d'un terminal à l'intérieur d'un spot a une influence sur l'identité des spots sur lesquels il va provoquer des interférences, ainsi que sur la puissance de ces interférences. Nous introduisons alors une seconde division de spots en zones pour permettre de prendre en compte ce phénomène.

On peut remarquer que dans ce modèle les interférences sont prises en compte de façon plus fine que dans [43] 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.

8.3.2 Formulation du problème et synthèse

Nous supposons que les utilisateurs peuvent utiliser n'importe laquelle des porteuses du système. Ces dernières appartiennent cependant à des types différents, c'est-à-dire qu'elles correspondent à des quantités de bandes passantes différentes. Parallèlement, les terminaux des utilisateurs utilisent des débits différents. On mesure donc la demande des utilisateurs en terme de nombre et de type de porteuses. Nous considérerons dans la pratique 5 types de porteuses.

Les données du système sont alors les demandes des opérateurs (en terme de nombre minimal et maximal pour chaque type de porteuses), les poids associés aux demandes et les conditions d'interférences. Les poids être choisis de multiples façons. Une possibilité est de les attribuer 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 ainsi qu'un positionnement possible de ces time-slots 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 les articles mentionnés précédemment dans la mesure où différentes difficultés sont regroupées : système multi-spots, multi-porteuses, avec plusieurs types de porteuses et un modèle d'interférence plus complexe. A cela s'ajoute des considérations d'équité.

Les deux chapitres suivants présentent respectivement les deux parties de la méthode de résolution proposée. Le premier décrit la résolution du problème dit de *planification radio*, pour laquelle 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 le second chapitre nous proposons une méthode de calcul du nombre de time-slots ainsi que de leur disposition dans le plan de fréquence de façon à satisfaire l'allocation trouvée dans la première partie en fonction des conditions d'interférences, c'est la phase d'*allocation des slots*.

Chapitre 9

Première partie de l'allocation : la planification radio.

Ce chapitre présente la résolution de la première phase de l'allocation. Elle consiste en l'allocation des porteuses à des fournisseurs de service (opérateurs) dont les clients 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 due 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 allocation qui équitable et possède des propriétés d'optimalité. La solution est l'entrée de la seconde phase, dite d'allocation de slots, que nous présentons dans le chapitre suivant.

Nous proposons une méthode combinatoire pour obtenir une allocation optimale équitale en temps polynomial en prenant en compte les problèmes pratiques comme des contraintes d'intégrité inhérente au système.

Le chapitre est organisé de la façon suivante. La section suivante introduit les notions et notations relatives à ce problème. Nous détaillons ensuite l'algorithme proposé (section 9.2) et présentons enfin quelques résultats numériques (section 9.3) avant de conclure.

9.1 Introduction et notations

Nous considérons un système satellite géostationnaire dans lequel un manager veut allouer de la bande passante entre différents opérateurs. On note qu'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, un opérateur local peut avoir besoin d'utiliser une plus grande partie de sa bande passante pour de l'information redondante (i.e. un plus fort taux de codage pour la correction d'erreur), et donc le débit effectif d'information diminue. Ainsi, si le manager décide maximiser le débit total du système, 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 dont les conditions de transmission sont mauvaises.

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 \mathcal{O} des opérateurs se voit affecter un nombre entier de porteuses dans chaque zone dans un ensemble de T types de porteuses : $\mathcal{T} = \{1, \dots, T\}$. La bande passante totale d'une porteuse de type t est alors notée B_t et nous supposons que $B_1 > \dots > B_T$.

Précisons que 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 réduit d'environ 10% par les overheads utilisés entre autres pour la signalisation et les marges de fréquences. 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 est prêt à payer pour une porteuse de type t dans une zone z . 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, ou plus généralement proportionnellement à la quantité réelle de débit allouée. Le problème devient ainsi d'allouer équitablement (et optimalement) le débit.

Nous supposons qu'il existe un nombre minimal et maximal de porteuses de type t par zone z demandées par chaque opérateur o . On note respectivement ces grandeurs $D_t^{\min}(z, o)$ et $D_t^{\max}(z, o)$. Comme dans les problèmes considérés 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, soit $\mathcal{O}(f)$ est l'ensemble des opérateurs auxquels sont assignés les porteuses qui contiennent une certaine fréquence f , et $\mathcal{F}(f)$ leurs porteuses associées. Si $\overline{\mathcal{F}}(f)$ est la porteuse de plus grande bande passante parmi $\mathcal{O}(f)$, alors les bandes de fréquence de toutes les porteuses de $\mathcal{F}(f)$ doivent être contenues dans la bande de fréquence de $\overline{\mathcal{F}}(f)$. De plus, 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. Nous constatons également que 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.

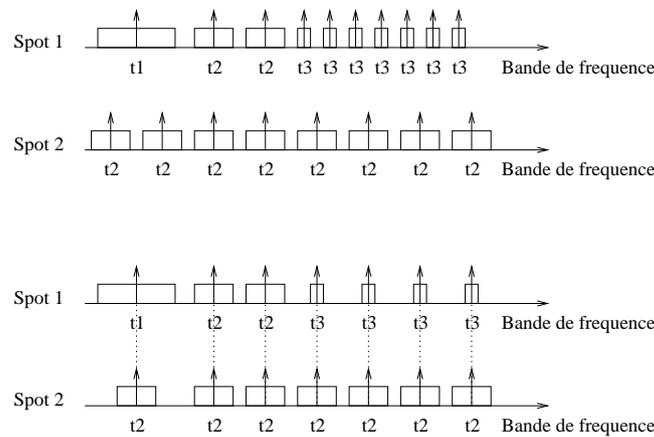


FIG. 9.1 – Deux allocations avec et sans ISCC.

9.1.3 Notations (résumé)

Le tableau suivant résume les notations utilisées dans ce chapitre :

Données du problème :

o / \mathcal{O}	un opérateur / l'ensemble des opérateurs
$s / \mathcal{S}(c)$	un spot / l'ensemble des spots de couleur c ,
$z / \mathcal{Z}(s)$	une zone / l'ensemble des zones du spot s
t / \mathcal{T}	un type de porteuse / l'ensemble des types ($ \mathcal{T} = T$),
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 les valeurs de J_t pour $t = 1, \dots, T$ et de $D_t(z, o)$ où J_t est le nombre de porteuses de type t et $D_t(z, o)$ est le nombre de porteuses de type t associées à la zone z et à l'opérateur o , comme résumé dans le tableau suivant.

Résultats fournis :

$D_t(z, o)$	le nombre de porteuses de type t attribuées à l'opérateur o dans la zone z ,
J_t	le nombre de porteuses de type t .

9.1.4 Critères d'équité

Dans le contexte considéré, on peut interpréter et implémenter les concepts d'équité, au niveau géographique de plusieurs façons. En effet, 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.

Considérons par exemple le partage équitable des débits entre les opérateurs au niveau global. La formulation générale du critère d'équité (équation (1.6)) s'écrit alors : Étant donnée une constante positive $\alpha \neq 1$, et une couleur c , considérons la maximisation de

$$\frac{1}{1-\alpha} \sum_{o \in \mathcal{O}} \left[\sum_{\substack{s \in \mathcal{S}(c), z \in \mathcal{Z}(s) \\ t \in \mathcal{T}}} (D_t(z, o) - D_t^{\min}(z, o)) \right]^{1-\alpha}$$

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 α -allocation de débit.

De la même façon, on peut définir l' α -allocation pondérée de débit comme étant l'unique solution de la maximisation

$$\frac{1}{1-\alpha} \sum_{o \in \mathcal{O}} \left[\sum_{\substack{s \in \mathcal{S}(c), z \in \mathcal{Z}(s) \\ t \in \mathcal{T}}} C_t(z, o) (D_t(z, o) - D_t^{\min}(z, o)) \right]^{1-\alpha}. \quad (9.1)$$

Nous proposons au contraire de résoudre le problème d'allocation équitable des porteuses entre les opérateurs au niveau local, comme expliqué dans la prochaine sous-section.

9.1.5 Formulation du problème

Nous nous intéressons au critère d'équité obtenu par les solutions de la maximisation de

$$\sum_{t \in \mathcal{T}, s \in \mathcal{S}(c), z \in \mathcal{Z}(s), o \in \mathcal{O}} \frac{\left[C_t(z, o) (D_t(z, o) - D_t^{\min}(z, o)) \right]^{1-\alpha}}{1-\alpha}. \quad (9.2)$$

Ceci est en adéquation avec la définition précédente dans le cas de l'équité appliquée à une échelle locale (c'est-à-dire de chaque zone). Elle peut être également interprétée comme le cas où l'on identifie un sous-opérateur comme étant responsable des demandes d'un type donné pour un opérateur donné dans une zone fixée.

Les contraintes de notre problème peuvent être résumées comme ceci :

$$\begin{aligned} \forall z \in \mathcal{Z}(c), \forall o \in \mathcal{O}, \forall t \in \mathcal{T}, \quad & D_t^{\min}(z, o) \leq D_t(z, o) \leq D_t^{\max}(z, o), \\ \forall s \in \mathcal{S}(c), \forall t \in \mathcal{T}, \quad & \sum_{i \in \mathcal{T} \text{ t.q } i \leq t} J_i \geq \sum_{z \in \mathcal{Z}(s), o \in \mathcal{O}} D_t(z, o),^1 \\ \forall t \in \mathcal{T}, \quad & J_t \geq 0, \\ & \sum_{i \in \mathcal{T}} J_i B_i \leq B. \end{aligned}$$

9.2 Algorithme d'allocation proposé

Dans cette partie nous proposons un algorithme de résolution du problème formulé dans la section 9.1.5. Il 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 fixé pour tous les spots et en déduisons la répartition des porteuses entre les opérateurs. Dans la seconde phase, nous montrons comment déterminer le nombre minimal 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

Nous définissons 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 alors satisfaire au moins $D_t^{(min)}(z, o)$ d'entre elles.

¹Ceci provient du fait que nous pouvons affecter à un opérateur une porteuse de plus grand type que celui qu'il demande.

Soit \mathcal{R} l'ensemble des requêtes. A l'intérieur d'un spot s , le nombre total de requêtes est alors $|\mathcal{R}| = R = \sum_{z \in \mathcal{Z}(s), o \in \mathcal{O}, t \in \mathcal{T}} D_t^{max}(z, o)$. Pour une requête r , on note t_r son type (où $t_r \in \{1, \dots, T\}$), et on lui associe une *valeur*, notée v_r . Nous expliquerons en section 9.2.3 comment choisir les valeurs v_r 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 , $i = 1, \dots, T$) et maximise la valeur $V = \sum_{r \in \text{Accept}} v_r$. 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.

```

k := min{r such that J_r > 0} - 1
for r := 1 to R
  begin
    if t_r > k then
      begin
        accept[r] := 1
        for l := k + 1 to T
          if  $\sum_{m \leq r \text{ and } t_m \leq l} \text{accept}[m] = \sum_{i \leq l} J_i$ 
            then k := l;
          end
        else accept[r] := 0;
      end;
    end;
  end;

```

FIG. 9.2 – Un algorithme d'acceptation des requêtes pour maximiser la somme de leur valeur.

La sortie de l'algorithme est un vecteur «accept» de R éléments booléens. Donc l'ensemble **Accept** se compose des 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. La preuve de l'optimalité de l'algorithme est donné dans l'annexe B.2.

9.2.2 Obtention du vecteur J minimal

Dans chaque spot s nous définissons

$$J_t^s = \sum_{o \in \mathcal{O}} \sum_{z \in \mathcal{Z}(s)} D_t^{\min}(z, o), \quad \text{et} \quad J^s = (J_1^s, \dots, J_N^s).$$

Le vecteur 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 $J^* = \max(\{J^s\}_{s \in \mathcal{S}(c)})$ le vecteur contenant les maxima des valeurs de J_s sur tous les spots.² 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 en attribuant la valeur 1 aux $D_t^{\min}(z, o)$ premières requêtes de chaque opérateur dans chaque zone et pour chaque type et la valeur 0 aux autres.

9.2.3 Étapes de résolution du problème d'allocation

Pour une couleur donnée, la procédure suivante permet d'obtenir le vecteur global $J = (J_1, \dots, J_T) \in \mathbb{N}^T$, et donc la solution de l'allocation optimale et équitable.

Pour chaque vecteur J admissible (c'est-à-dire vérifiant $J \succeq J^*$), nous déterminons 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 obtenue pour cette valeur de J en utilisant notre critère global d'équité défini par l'équation (9.2). 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 α . Les deux paragraphes suivants détaillent ces étapes.

²Nous pouvons considérer un ordre partiel sur 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

$$\begin{aligned} J'_1 &\geq J_1, \\ J'_2 + J'_1 &\geq J_2 + J_1, \\ &\vdots \\ &\vdots \\ \sum_{t=1}^T J'_t &\geq \sum_{t=1}^T J_t. \end{aligned}$$

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 $J^* \succeq J$ et $J^* \succeq J'$, et telle que pour toute allocation J'' telle que $J'' \succeq J$ et $J'' \succeq J'$, nous avons $J'' \succeq J^*$.

9.2.3.1 Maximisation des bénéficiaires du surplus (pour $\alpha \geq 1$).

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 devient infini. C'est pourquoi nous souhaitons minimiser le degré de nullité du polynôme $\prod_{z,t,o} (D_t(z, o) - D_t^{min}(z, o))$ en maximisant le nombre d'opérateurs à qui l'on peut allouer strictement davantage que leur demande minimale.

Pour cela, nous utilisons l'algorithme de la section 9.2.1, pour chaque opérateur, dans chaque zone et chaque type de porteuse et fixons les valeurs des valeurs des requêtes à :

- 2 pour $D_t^{min}(z, o)$ requêtes
- 1 pour une requête additionnelle
- 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 requièrent (il s'agit des opérateurs dont l'une des requêtes de valeur 1 a été acceptée). Nous restreignons alors le problème uniquement à ces opérateurs et appliquons la seconde étape de l'algorithme, telle que décrite dans le paragraphe suivant.

9.2.3.2 Redistribution équitable des valeurs (toute valeur de α).

Nous utilisons l'algorithme de la section 9.2.1 dans lequel l'on choisit judicieusement les variables v_r , $1 \leq r \leq R$ 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 des $D_t^{min}(z, o)$ requêtes à $+\infty$ (dans la pratique, n'importe quelle borne supérieure Λ 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 (i.e. correspondant à $D_t^{min}(z, o) + 1$) à la valeur Λ afin de s'assurer que l'allocation de tous les opérateurs 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 :

$$\begin{cases} \log \left(\frac{j - D_t^{min}(s, o)}{j - 1 - D_t^{min}(s, o)} \right) & \text{si } \alpha = 1^3 \\ C_t(s, o)^{1-\alpha} [(j - D_t^{min}(s, o))^{1-\alpha} - (j - 1 - D_t^{min}(s, o))^{1-\alpha}] & \text{sinon.} \end{cases}$$

³On peut noter que la valeur v des requêtes est alors indépendante de C . Ceci est du au fait que le NBS est invariant par multiplication par une constante.

9.3 Expérimentations

Nous présentons brièvement dans cette section quelques résultats numériques obtenus par implémentation de l'algorithme présenté à la section précédente.

9.3.1 Valeurs des paramètres

On considère un exemple de réseau satellitaire contenant 32 spots équi-répartis sur 4 couleurs. Le nombre de spots par couleur vaut donc 8. Comme les problèmes associés à deux couleurs sont différentes, nous nous concentrons sur une seule des quatre couleurs du système.

9.3.1.1 Types et utilités 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, c'est-à-dire la marge nécessaire entre deux porteuses est de 50%. Ceci signifie qu'en pratique la largeur de la bande occupée par une porteuse est de 150% la valeur de sa bande passante. Ainsi, si toute la bande passante était utilisée pour un seul type de porteuse, alors le système pourrait accepter un maximum de 4 porteuses du premier type, 8 du second, 16 du troisième, 32 du quatrième et 128 de la dernière. Chaque spot est divisé en 3 zones, et les valeurs des utilités $C_t(z, o)$ de chaque type de porteuse dans l'un des spots est donné par le tableau suivant (on suppose ici que la valeur de $C_t(z, o)$ est indépendante de l'opérateur o et du spot considéré) :

$C_t(z, o)$	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, ce qui explique les différences de valeur considérées entre les zones.

9.3.1.2 Génération des demandes des opérateurs

Les demandes des opérateurs ont été générés aléatoirement de sorte à respecter les contraintes suivantes : le système est utilisé par 40 opérateurs, chacun utilisant 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 selon une loi uniforme. 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 $1/2$. S'il ne sélectionne 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 (avec une loi uniforme) 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_{\min} porteuses de type t et espère avoir au plus $d_{\max} - d_{\min}$ porteuses de type $\min(t+1, T)$. En d'autres termes, $D_t^{\min} = D_t^{\max} = d_{\min}$, $D_{\min(t+1, T)}^{\min} = 0$, $D_{\min(t+1, T)}^{\max} = d_{\max} - d_{\min}$, et pour tout $s > \min(t+1, T)$, $D_s^{\min} = D_s^{\max} = 0$. 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 α . La première figure représente l'allocation équitable et optimale selon l'algorithme présenté précédemment. La seconde illustre le cas (plus simple) dans lequel on ne prend pas en compte les contraintes ISCC.

Pour chaque valeur du critère d'équité, nous traçons la bande passante allouée à chaque opérateur dans les différents spots par types de porteuse croissant sous forme d'un parallélépipède⁴. Les allocations de chaque opérateur sont représentées avec une couleur spécifique, en couleur pleine pour les allocations minimales et 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 du parallélépipède 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.

Les expériences montrent clairement que, lorsque α augmente, les porteuses de plus grands types sont favorisées au détriment de celles de plus grande capacité.

⁴En fait nous représentons la parallélépipède 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) porteuse allouée soit en fait de type inférieur, c'est-à-dire de capacité supérieure à celle demandée.

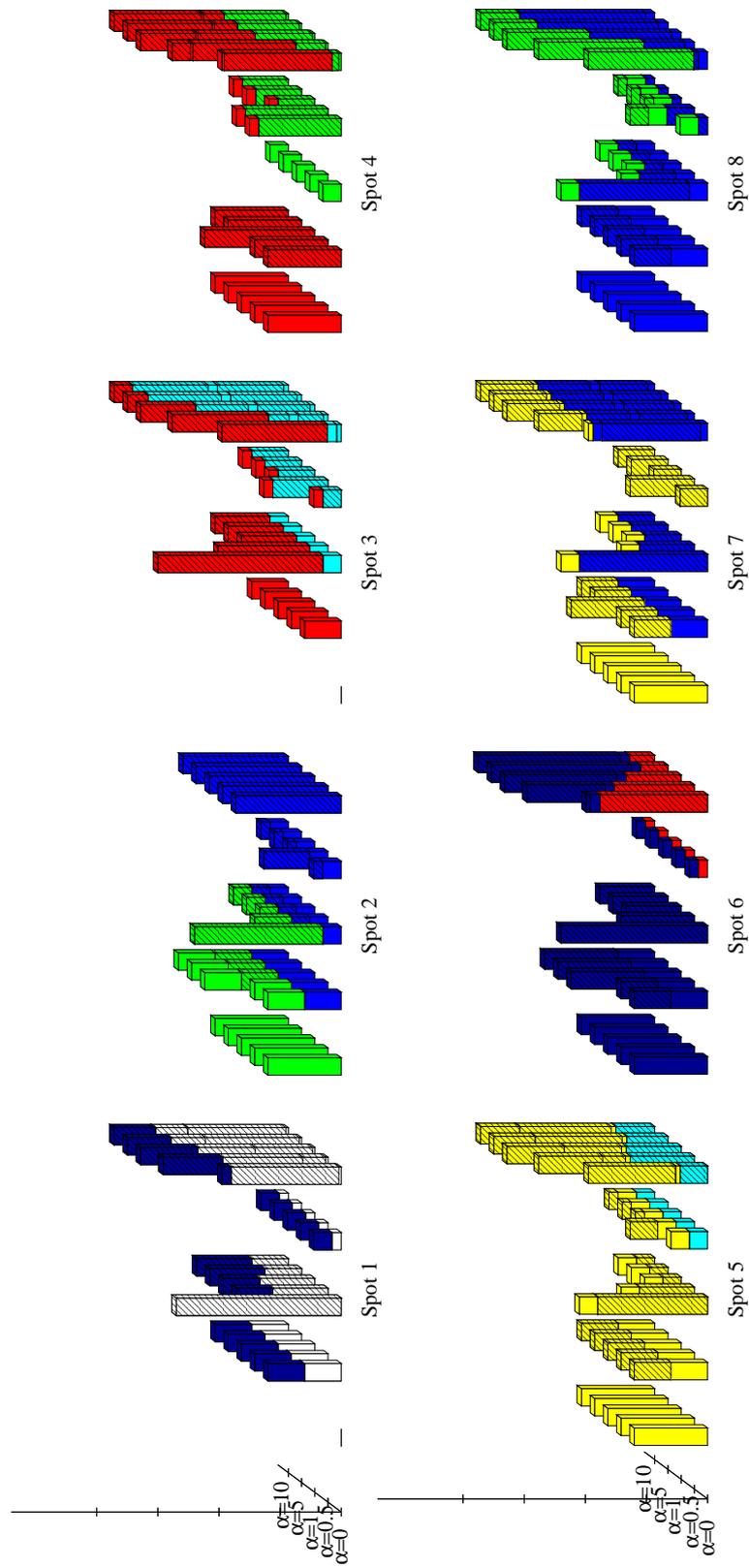


FIG. 9.3 – Allocation équitable avec ISCC

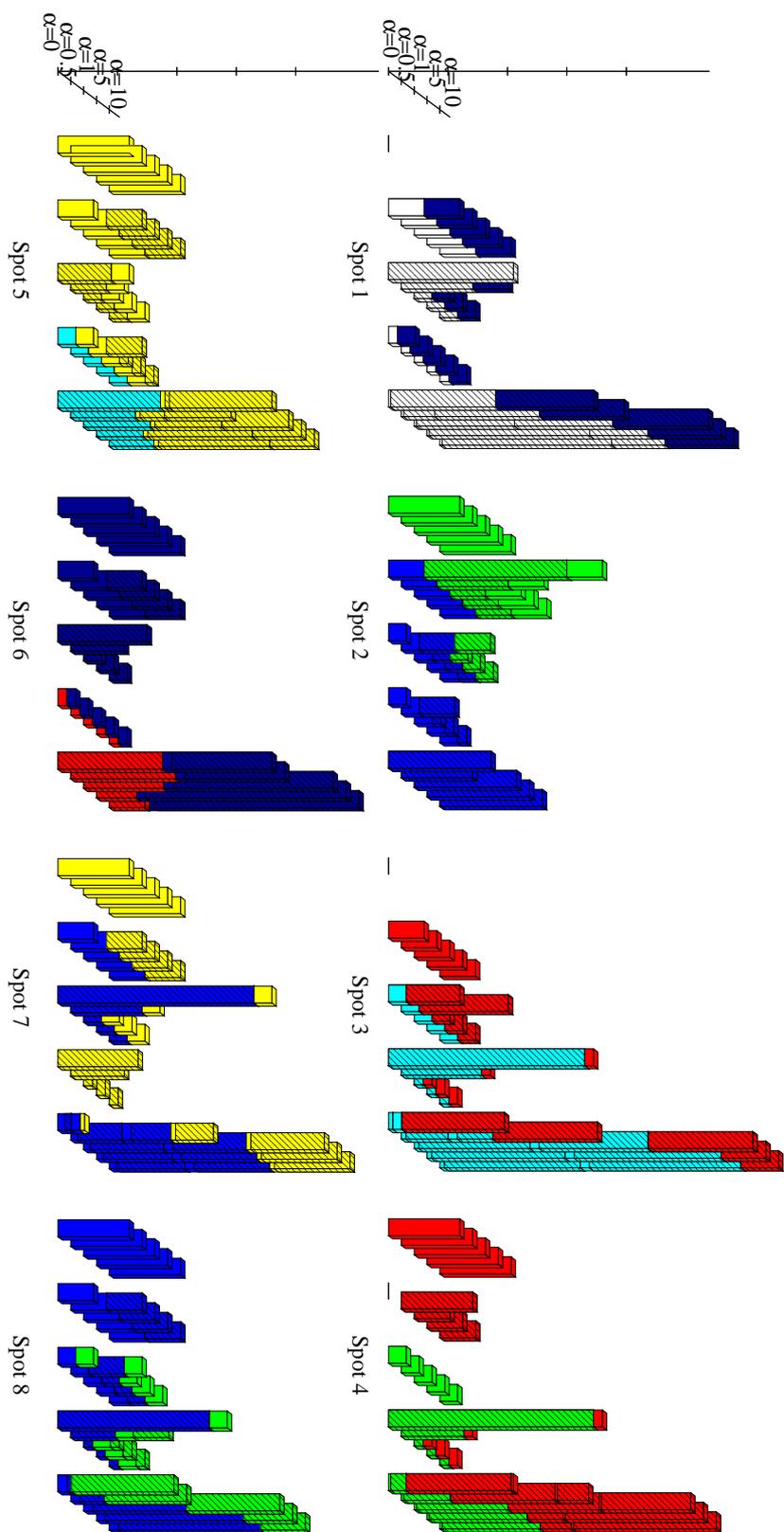


Fig. 9.4 – Allocation équitabile sans ISCC

Enfin, la figure 9.5 montre l'efficacité du système, mesuré en pourcentage moyen de bande passante utilisée sur un spot en fonction du paramètre d'équité α . Lorsque α augmente on remarque que l'efficacité est réduite ce qui est le prix logique d'une plus grande équité. Nous remarquons également que la capacité du système perdue avec les contraintes ISCC reste inférieure à 10%, ce qui nous conforte dans l'idée que le coût de l'ISCC reste raisonnable.

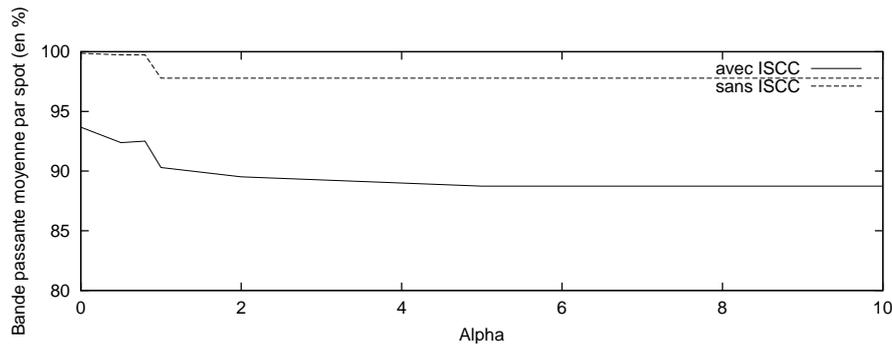


FIG. 9.5 – Efficacité de l'allocation de la bande passante avec et sans l'ISCC

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 présentée ne prenaient en compte qu'un ou deux types de porteuses. De plus, la méthode proposée peut être utilisée sur des systèmes relativement grands sans que le temps de calcul ne devienne rédhibitoire. Enfin, cette étude est, à notre connaissance, la première introduisant des notions d'équité.

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 ce 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 l'étude de l'équité.

Chapitre 10

Deuxième partie de l'allocation : l'allocation des slots.

Nous nous intéressons ici au problème d'allocation des slots dans un système MF-TDMA. Les contraintes sont que sur chaque spot, un slot ne peut être alloué simultanément à plus d'une zone. De plus, les spots interférant entre eux, un slot ne peut être alloué à un opérateur dans une zone donnée que si l'interférence qu'il subit est inférieure à un certain seuil. Enfin, 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). Il est donc important que l'algorithme proposé ne nécessite pas plus de quelques minutes pour s'exécuter. Le but de l'allocation est alors 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. Cependant, en castant le problème en une coloration de graphes on peut montrer qu'il est NP-complet¹.

Nous proposons deux méthodes de résolution. La première est basée sur le recuit simulé et la seconde sur la programmation linéaire. Nous les mentionnons ici afin de compléter la solution proposée du problème d'allocation de ressources dans le cas des réseaux satellitaires. Cependant, une version détaillée des deux solutions se trouve dans [123] pour l'approche par recuit simulé et [3] pour la programmation linéaire.

La structure du chapitre est la suivante : après une brève introduction sur le modèle d'interférence considéré (section 10.1), nous consacrons une courte section à la présentation générale de chacune de deux approches (section 10.2 et 10.3).

10.1 Le modèle d'interférence.

Une même bande de fréquence est potentiellement réutilisée d'un spot à l'autre. Cette réutilisation n'est cependant possible que si le niveau d'interférence générée par l'ensemble des spots utilisant simultanément la ressource radio est acceptable.

¹C'est-à-dire qu'il n'est pas possible de le résoudre en un temps polynomial

Les hypothèses sont les suivantes. Chaque zone génère une certaine interférence sur tous les autres spots $I(s, z)$ (si z est une zone de s , i.e. si $z \in \mathcal{Z}(s)$ alors $I(s, z) = 0$). Chaque zone possède également un gain $G(z)$. Si $spot(z)$ est le spot contenant la zone 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 G en décibels est prise aléatoirement entre deux bornes. Le seuil d'interférence σ est une donnée du problème. Si σ 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) . Réciproquement, plus le seuil est élevé, plus le nombre de spots pouvant être actifs simultanément et dans la même bande de fréquence est petit.

On note que les couleurs correspondant à des bandes de fréquences différentes, les problèmes d'interférence considérés ne se produisent qu'entre deux spots d'une même couleur.

10.2 Le recuit simulé

Le problème étant NP-complet, nous proposons une approche heuristique basée sur le recuit simulé [89, 69] et analysons ses performances. Comme notre nouvelle allocation doit être générée chaque heure, les solutions sont nécessaires en moins de quelques minutes, ce qui induit des choix d'implémentations que nous présentons ici. 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 était généré aléatoirement, et de l'ordre de la centaine sur chaque spot.

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'en améliorer les performances dans notre contexte (sous-section 10.2.3). Des expériences numériques sont enfin proposées pour tester nos améliorations (sous-section 10.2.4).

10.2.1 Méthodes heuristiques et recuit simulé

L'article [69] présente une belle introduction sur les méthodes heuristiques et en particulier du recuit simulé. Il montre également l'inconvénient majeur des algorithmes basés sur des heuristiques : ils sont plutôt spécifiques d'un problème. Il n'y a ainsi aucune garantie qu'une heuristique efficace pour un système le soit pour un autre.

Les problèmes comme celui du voyageur de commerce² 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 familles principales d'heuristiques : *divide and conquer* (dans lequel on divise successivement le problème en sous-problèmes) et les méthodes itératives. Le recuit simulé appartient à la seconde catégorie. Le risque d'une méthode dans laquelle on améliore progressivement la solution est que l'algorithme peut alors converger vers un optimum local et non global. Pour éviter cela, une possibilité est alors d'utiliser l'algorithme à partir de plusieurs configurations initiales et de ne retenir que le meilleur résultat. Comme nous le montrons dans la suite, le recuit simulé propose une autre approche.

Le recuit simulé est né dans les années 50 des travaux de Metropolis [89]. L'idée est née d'une analogie avec la physique, plus précisément de l'étude de la façon dont les atomes s'agrègent (en moyenne) aux 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 proches 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 comprend alors que le choix de cette fonction (c'est-à-dire 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 [48] énonça des conditions que doivent remplir les schémas de refroidissement pour assurer une convergence vers le minimum absolu.

10.2.2 Implémentation du recuit simulé

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 acceptons le candidat si les conditions précédentes 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 la nouvelle allocation.

²Le problème du voyageur de commerce peut être formulé comme ceci. Étant donné une liste de villes et des coûts associés au 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 son coût total.

Pour décider que faire à chaque étape, nous considérons la mesure de performance \mathcal{G} qui est la somme du nombre de slots alloués pondérée par le taux de codage (elle correspond donc au débit utile du système qui est notre fonction à maximiser). Nous avons implémenté et présenté dans [123] une méthode de recuit simulé hiérarchique avec comme second critère la minimisation de l'interférence créée.

L'algorithme s'arrête si l'une des conditions suivantes est vérifiée :

- La limite de temps est dépassée,
- La 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 l'utilisation de taux de refroidissement linéaires et logarithmiques. On peut noter que le refroidissement linéaire est le plus souvent utilisé, car il est beaucoup plus rapide que le logarithmique. Cependant, la convergence asymptotique à un optimum global n'est plus garantie [48].

10.2.3 Améliorations de l'algorithme

Nous avons amélioré 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 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 les plus basses fréquences de la bande passante. Dans l'autre moitié des spots, ces porteuses ont été placées dans les plus grandes fréquences. Ceci nous permet de réduire simplement les interférences entre les porteuses de grande bande passante. Nous l'appelons «intelligent placement of carriers» (iPoC).

Une seconde amélioration a été d'effectuer le recuit simulé en deux étapes plutôt qu'une. Chaque phase dispose de la moitié du temps limite initialement alloué à 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 avons ensuite effectué des simulations avec différentes limites d'interférences permises et avons tracé le débit efficace obtenu, normalisé par celui obtenu avec la borne supérieure. Les résultats sont présentés sur la figure 10.1 pour différents schémas

d'utilisation du recuit simulé : avec refroidissement linéaire ou logarithmique (ces derniers sont indiqués par «LT» sur la figure), avec iPoC ou non, et en une ou deux étapes. La meilleure performance est obtenue avec notre version optimisée, combinant l'utilisation hiérarchique du recuit simulé (en prenant comme second critère la minimisation de l'interférence générée), l'iPoC et la résolution en deux étapes.

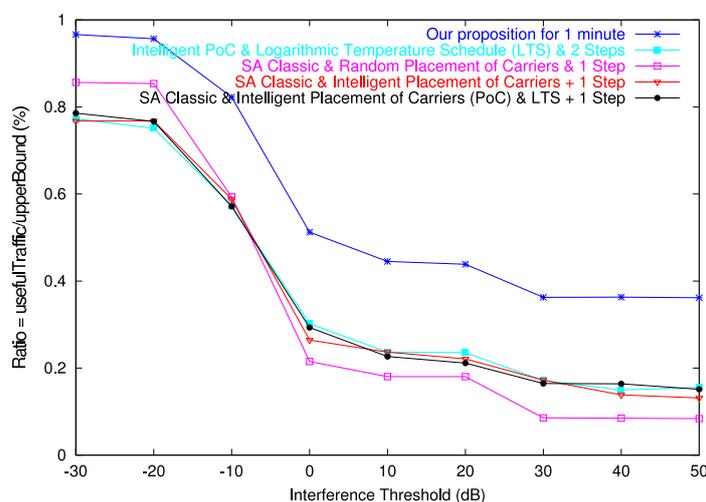


FIG. 10.1 – Comparaison pour une minute

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 jusqu'à 600 % pour les grands seuils ($+50$ dB) par rapport aux implémentations directes du recuit simulé à ce problème.

10.2.4.1 Conclusion

Nous avons proposé une approche de recuit simulé pour le problème d'allocation des slots dans le lien montant d'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 comparées numériquement à d'autres approches. Nous avons observé qu'un gain substantiel est obtenu.

En fait, le recuit simulé est une amélioration des techniques utilisée dans des travaux entrepris par l'entreprise Alcatel Space Industries. Certes, le gain est substantiel par rapport à ceux reposant notamment sur une affectation aléatoire, mais encore très sous-optimale. Nous avons développé une seconde technique, bien plus performante car mieux adaptée au problème que nous évoquons brièvement dans la sous-section suivante. Nous n'avons pas jugé utile de présenter de comparaison des deux méthodes, les différences de performances étant trop importantes, notamment en matière de taille maximale de 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 développé une seconde méthode de résolution du problème d'allocation des slots basée sur une programmation linéaire. Nous la mentionnons ici pour compléter l'étude de l'allocation de ressources dans les réseaux satellitaires et en expliquons le principe. Le lecteur désireux d'en connaître les détails pourra se référer à [3].

Nous présentons dans une première sous-section les notations utilisées dans la suite. 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 (section 10.3.2 et 10.3.3). Nous terminons enfin cette présentation par quelques résultats numériques et une brève conclusion (section 10.3.4).

10.3.1 Présentation du problème

10.3.1.1 Notations

On notera S le nombre de spots de chaque couleur et B la taille de la bande passante totale. En pratique, on s'intéresse à $S \leq 32$. Chaque zone de chaque spot est notée sous la forme : *spot.zone*.

Les terminaux ont des capacités variables d'émission, qui sont prises en compte 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 dans d'un time-slot est une constante du problème. Ainsi, pour tout type t , 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 Δ .

$$\forall t = 1, \dots, T, \quad \Delta = t_k^b t_k^t.$$

Chaque spot a une demande spécifique à satisfaire, exprimée en nombre de time-slots³ par types de terminaux et par zone.

$$\forall z_i, \forall t, \quad d_{z_i}^t := \text{demande dans la zone } z_i \text{ en type } t$$

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 interférant. N'ayant pas à notre portée de simulations du niveau du signal interférant dans toutes les zones géographiques, nous adoptons

³Les time-slots d'un type donné correspondent à l'allocation de porteuses de ce type obtenu dans la phase de planification radio. Remarquons que dans la phase d'allocation de ressources, nous ne considérons plus les notions équité et donc que le concept d'opérateur n'apparaît pas.

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 $S_V(z)$ et l'interférence globale, générée par l'ensemble des spots notée $S_G(z)$. Enfin, la valeur finale de l'interférence générée sur la zone z est prise comme étant un barycentre à poids fixe (noté γ) de ces deux valeurs :

$$S_V(z) = \sum_{z' \text{ active dans un spot voisin}} I(\text{spot}(z), z'), \quad S_G(z) = \sum_{z' \text{ active}} I(\text{spot}(z), z')$$

$$I(z) = \gamma S_V(z) + (1 - \gamma) S_G(z).$$

Le critère d'allocation (10.1) devient alors

$$\frac{G(z)}{I(z)} \geq \sigma. \quad (10.2)$$

10.3.2 Définition des familles et notions associées

Dans cette section, nous introduisons et étudions la notion de *famille* de zones (section 10.3.2.1). En pratique, comme le nombre de familles est trop important pour qu'elles puissent toutes être générées, nous introduisons les concepts de *super-zone* (section 10.3.2.2), de *famille générique* (section 10.3.2.3) et de *statut* d'un spot (section 10.3.2.4).

10.3.2.1 Qu'est-ce qu'une famille ?

Pour comprendre la modélisation, considérons le système simple représenté sur la figure 10.2. Il se compose d'un seul type de terminal (comme en GSM) et de trois spots, chacun contenant deux zones.

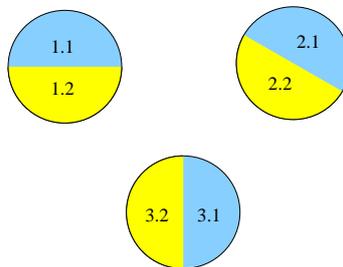


FIG. 10.2 – Petit exemple sur 3 spots.

Le canal radio est découpé en time-slots dans chacun desquels 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*⁴) dans un time-slot donné. Dans l'exemple illustratif considéré, on choisit la valeur du gain de chaque zone ainsi que des interférences qu'elle génère conformément aux valeurs du tableau 10.1.

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 cas cette zone peut-elle émettre ? Cela dépend des états des zones qui interfèrent sur le spot 1. 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 [3].

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.

Si le seuil d'admissibilité est fixé à 0.30, on peut voir que la zone 1.1 peut émettre pour toutes les configurations des spots 2 et 3, sauf une, correspondant à l'allumage simultané des zones 3.2 et 2.2. Nous montrons les configurations permettant l'émission de la zone 1.1 simultanément à l'émission sur les spots 2 et 3 sur la figure 10.3.

Nous appellerons dans la suite *famille* une configuration admissible.

Pour une meilleure résolution du problème, d'autres notions peuvent être utiles. Par exemple, notons que les familles 1 et 3 (de la figure 10.3) diffèrent seulement par le spot 2. En fait, il est correct de dire qu'il est possible d'allumer *tout* le spot 2 simultanément aux zones 1.1 et 3.1. Nous créons ainsi 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 utilisons également quelquefois des familles qui ont «trop» de zones allumées dans le sens où certaines resteront inutilisées pendant l'utilisation effective du système. Ceci est résumé dans la notion de *super-zone* et de *famille générique*.

⁴Similairement on dit qu'un spot est allumé si l'une des zones qu'il contient l'est (on rappelle que deux zones d'un même spot ne peuvent être simultanément allumées car elles sont gérées par un même faisceau du satellite).

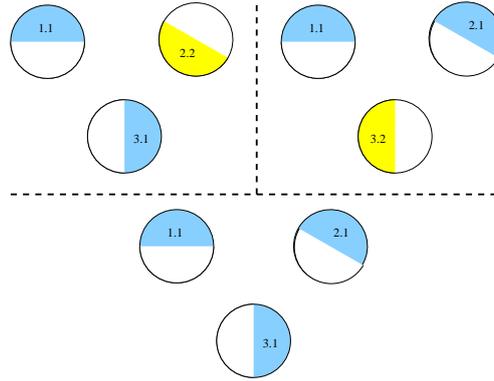


FIG. 10.3 – Configurations d'émissions possibles de la zone 1.1 simultanément à celle des spots 2 et 3 pour un seuil à 0.30.

10.3.2.2 Définition des super-zones.

Le nombre de familles possibles est en pratique trop grand pour que l'on puisse les décrire de manière exhaustive. Par conséquent nous introduisons 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, alors chacune des zones de cette dernière peut émettre en respectant les conditions d'interférences. Ainsi, dans l'exemple de la figure 10.2, $\{2.1, 2.2\}$ est une super-zone et une famille admissible est $1.1 - \{2.1, 2.2\} - 3.1$. Les paramètres d'une super-zone ζ 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) . Si l'on utilise la modélisation classique en cellule hexagonales, on peut noter que mis à part les spots en limite de 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 dans lesquels plusieurs spots voisins sont éteints.

Si σ 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.

Quand le seuil d'interférence σ 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

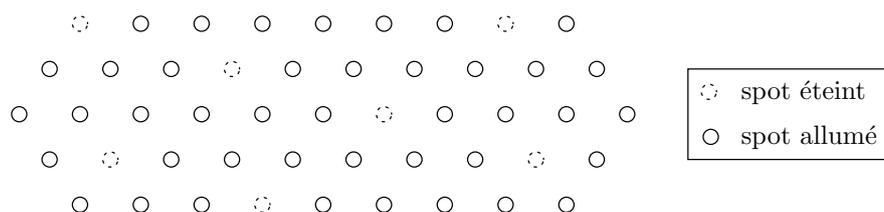


FIG. 10.4 – Illustration d'une famille 6/7

le schéma 5/7 dans lequel 5 spots sur 7 sont allumés. Le schéma 5/7 est représenté dans la 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.

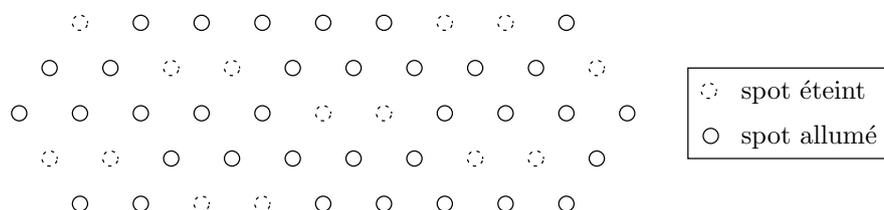


FIG. 10.5 – Illustration d'une famille 5/7

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 σ est très élevé.

10.3.2.4 Statut des spots

Dans la sous-section précédente, nous avons introduit la distinction entre spot allumé (ou actif) et spot éteint. Or, tout spot comporte un certain nombre de zones géographiques (typiquement 2 ou 3 zones). Il existe donc plusieurs configurations possible 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 donc la notion de *statut* d'un spot qui prend valeur dans l'intervalle $[0, 2^{Z(s)} - 1]$, où $Z(s)$ désigne le nombre de zones du spot s . Pour un spot contenant 3 zones, son statut peut donc 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 valant 7, ceci veut dire que nous pouvons allouer les ressources à n'importe laquelle des 3 zones.

Une autre conséquence de la notion de statut de spot est relative au calcul du critère d'allocation (10.2) : au lieu de devoir 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.

En pratique, 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 $2^{Z(s)} - 1$). 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 $2^{Z(s)} - 1$, les spots éteints ayant le statut 0. Nous testons successivement la validité de ces familles et les séparons en deux ensembles : les familles 6/7 valides d'une part et les familles 6/7 non valides de l'autre. Nous pouvons faire de même avec les familles 5/7, 4/7, etc., selon le seuil d'interférence σ considéré. Pour qu'une famille non valide devienne valide, faut éteindre certaines de ses zones allumées. Nous proposons une heuristique permettant de 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 des listes de zones pouvant 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

La planification de l'utilisation d'un terminal de type t pendant un time-slot consiste en fait au placement schématiquement d'un rectangle d'une certaine surface dans l'intervalle temps-fréquence. La figure 10.6 illustre l'utilisation dans un même spot de deux terminaux de types différents.

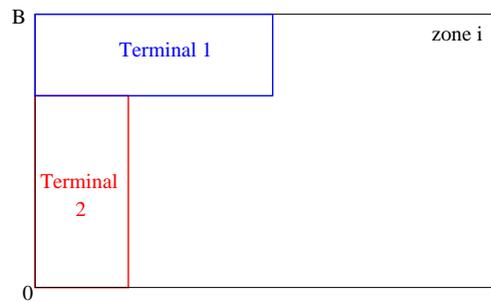


FIG. 10.6 – Utilisation de terminaux de types différents.

On propose 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 à allouer des familles typées simultanément dans l'ensemble des spots.

Pour chaque famille on peut choisir un type de transmission t qui sera diffusé sur tous les spots concernés. On place alors 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 *multiples*. Si t est le type de la famille de plus large bande et t' celui de largeur de bande la plus étroite, alors la famille utilisera l'aire (t_k^b, t_k^t) . Sur chaque spot, la famille se composera de times-slot de type différents, mais de même nombre (puisque les valeurs des aires $\Delta = t_k^b t_k^t$ sont supposés indépendantes du type choisi, cf section 10.3.1). Nous proposons de déterminer les familles par programmation linéaire. Une fois cela effectué, il reste à leur attribuer un type.

On remarque qu'étant donné une famille non-typée, il existe T^S manières différentes de la typer, soit bien plus qu'il n'est raisonnable d'entrer dans le programme linéaire ! Ce pourquoi nous avons utilisé la méthode classique de *génération de colonnes* que nous ne détaillons pas ici, mais dont l'implémentation est donnée dans [3].

10.3.4 Résultats numériques et conclusion

Nous avons testé différents types de configurations de systèmes contenant 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 (des six spots voisins) et pour 15%, comme une interférence globale.

La sortie «classique» de notre programme 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 types de 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 time-slot de type 1, 2 ou 3 selon les spots avec $n = t_3^t / t_1^t$.

Si l'on considère des configurations à 8 spots par couleur, notre programme peut trouver la solution flottante optimale pour des temps de calcul de l'ordre de la minute. 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 de solutions en temps réel.

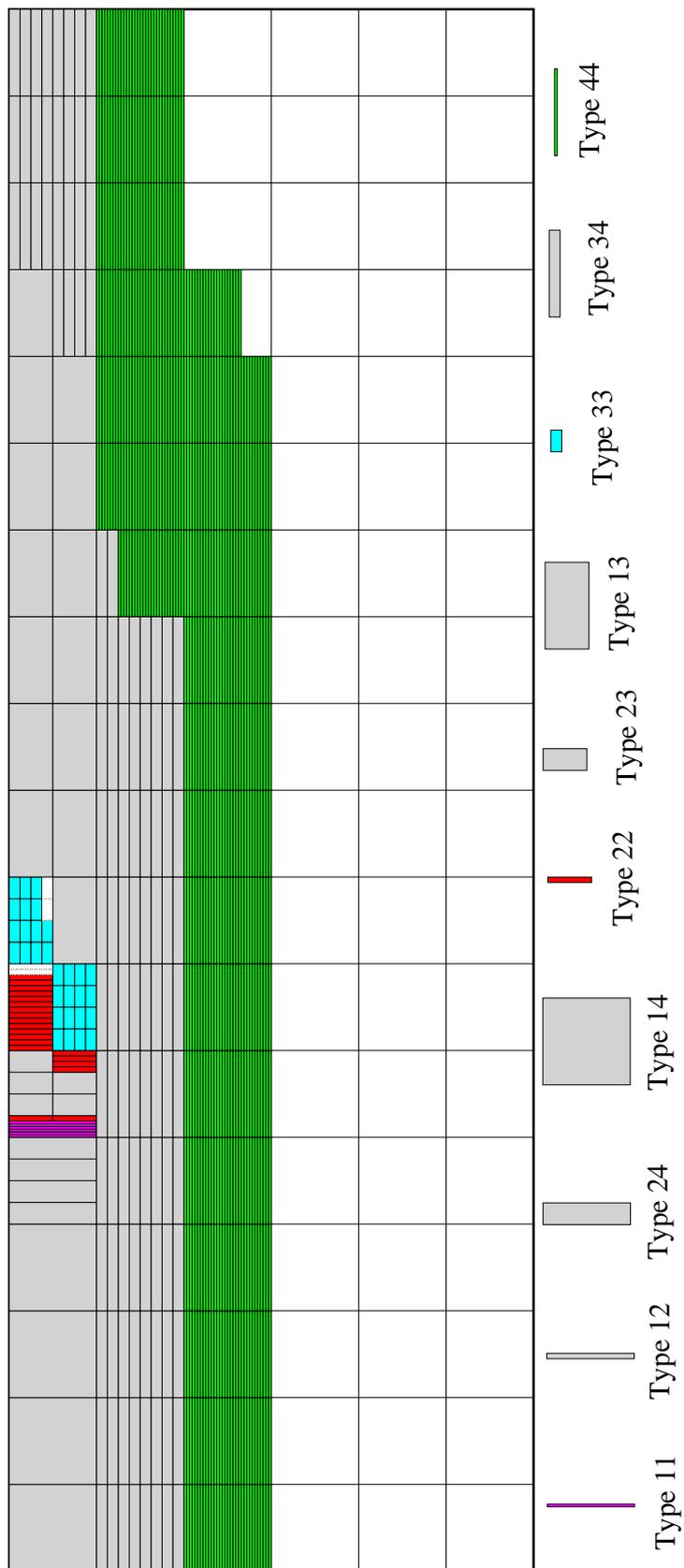


FIG. 10.7 – Un exemple de résultat d'allocation de ressources.

10.4 Conclusion

Nous avons considéré dans ce chapitre le problème d'allocation des time-slots, c'est-à-dire la question du placement effectif des porteuses (allouées dans la phase de planification radio) dans l'espace bi-dimensionné temps-fréquence. Le problème étant NP-complet, nous avons proposé deux méthodes algorithmiques d'approximations.

La première est basée sur une implémentation de recuit simulé hiérarchique, pour laquelle nous avons proposé quelques mécanismes simples permettant de considérablement améliorer les performances obtenues.

Notre seconde résolution est basée sur constatation simple que pour tout couple temps-fréquence (t, f) , il existe une famille 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 [46]. Dans le cas du coloriage de graphe, on s'appuie sur un ensemble de familles de sommets indépendants. Ici, nous utilisons des familles de zones pouvant communiquer simultanément. La résolution du problème peut alors être obtenue par la programmation linéaire avec génération de colonnes.

Comme nous l'avons vu dans la partie introductive, les approches classiques ne considèrent pas de problème d'interférences couplé à un choix multiple de types 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 aurait été 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 d'émettre dans un time-slot quelconque. Cependant, nous pouvons espérer que les systèmes futurs permettront de telles libertés, qui offriront alors, une bien meilleure utilisation du spectre.

Conclusion

Dans cette dernière partie, nous avons proposé une méthode originale de résolution du problème d'allocation équitable et optimale de porteuses à des opérateurs dans un réseau satellitaire géostationnaire MF-TDMA en deux étapes. Les résultats offrent une très bonne utilisation des ressources dans les systèmes TDMA.

La méthode fournie 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, l'approche est novatrice dans la mesure où elle intègre des notions d'équité.

Nous pouvons noter ici que nous n'avons pas utilisé notre concept général d'équité associé aux fonctions d'utilité mais uniquement sa version pondérée. La raison en est que les allocations sont adressées à des opérateurs et non des utilisateurs finaux. Ainsi, elles ne sont pas associées à une seule mais à un ensemble d'applications.

L'intérêt de ce type de réseaux était également de fournir un exemple typique de situation dans laquelle des contraintes d'intégrité apparaissent. De par la simplicité relative du problème (puisque 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 la SDP.

Conclusion

Conclusion

Les travaux présentés dans cette thèse portent sur l'adaptation de concepts généraux d'équité au cadre des réseaux de télécommunication. L'idée fondamentale fut d'appliquer une formulation globale d'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. Nash s'est en effet intéressé à des problèmes de « marchandages » : des situations dans lesquels deux joueurs possédant des biens de natures différentes collaborent afin de trouver un accord satisfaisant pour tous. Chacun associe une valeur relative à chaque objet de la transaction.

La formulation de l'équité que nous utilisons est dérivée de celle introduite par Mo et Walrand dans [94]. Elle est définie sous la forme d'un problème d'optimisation paramétré. 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 ne nous sommes pas intéressé ici à la façon de déterminer les fonctions d'utilité. On trouve dans la littérature de nombreuses études sur les perceptions de qualité issues de l'utilisation de différents types d'applications temps réels, telles que la vidéo ou la téléphonie notamment. A la lumière de ces travaux, nous avons étudié l'allure de ces fonctions et avons proposé de les approcher par des fonctions quadratiques paramétrables.

Nous avons ensuite étudié précisément trois types de réseaux. En effet, si l'allocation équitable se justifie a priori dans tout réseau dans lequel évoluent des applications temps-réel et dans lequel on peut allouer une quantité variable de ressource, nous avons observé que jusqu'à présent seules quelques maigres tentatives d'intégration de l'équité avaient vu le jour, principalement dans le cadre des réseaux terrestres. Nous avons montré que les problèmes d'optimisation correspondant aux allocations dans les différents réseaux sont de natures très diverses. En particulier, le système de contraintes correspondant aux réseaux utilisant des communications radio 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 les systèmes satellitaires géostationnaires MF-TDMA. Le premier exemple est le plus simple à résoudre car le système de contraintes consiste en une suite d'inégalités linéaires. Dans le cadre des réseaux mobiles la difficulté du problème vient de l'interdépendance entre la puissance utilisée et de débit alors utilisable. Enfin, dans les réseaux satellitaires la difficulté provient de l'existence de contraintes d'intégrité (en effet, on ne peut allouer n'importe quelle quantité de débit mais seulement un nombre entier de porteuses de capacité prédéfinie).

Nous avons proposé des méthodes algorithmiques de résolution pour chaque système étudié. 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. Dans ce manuscrit, nous avons présenté en détail son implémentation dans le cadre des réseaux terrestres. Lors de l'étude des systèmes cellulaires, nous avons réalisé une étude analytique rigoureuse du système de contraintes et donné une nouvelle formulation du problème dans lequel les contraintes s'expriment simplement, permettant alors sa résolution par la SDP. Nous avons alors montré que l'implémentation qui en découle est 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 spécifique et avons montré qu'il était de complexité polynomiale.

Les possibilités d'extensions de ces travaux sont multiples. Un axe majeur serait de considérer des méthodes décentralisées permettant d'obtenir des équilibres correspondant aux différents critères d'équité, en particulier dans les réseaux terrestres. En effet, si dans les réseaux mobiles et satellitaires l'allocation est gérée de façon centralisée par la station de base ou le satellite, dans les systèmes tels que l'Internet, il n'existe pas d'institution pouvant jouer ce rôle de coordonnateur : il est alors nécessaire de trouver des mécanismes permettant d'obtenir les allocations de façon décentralisée. 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 [24]. Il serait intéressant de pouvoir en étudier plus avant les propriétés et relations éventuelles avec les équilibres obtenus avec notre critère d'équité généralisé.

Quatrième partie

Annexes

Annexe A

Programme de création des matrices d'entrées.

Le programme donné ci-dessous (écrit en langage *C*) est un exemple de ce que l'on peut utiliser pour créer les matrices d'entrées au solveur *csdp*. Afin d'en limiter la taille, il ne correspond que au cas du réseau terrestre avec routage fixé dans le cas du NBS (i.e. pour $\alpha = 1$). Il s'agit du programme utilisé pour résoudre l'exemple présenté dans la section 4.4.1.

On note qu'ici la matrice de trafic ainsi que tous les arguments du problème font partie du programme. Dans l'implémentation réelle, ces paramètres, ainsi que la valeur de α sont bien évidemment des arguments d'entrée du programme.

```
main(int argc, char *argv[]) {

/* Paramètres du problème */
// Topologie du réseau
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}};

// Paramètres des connexions (toutes identiques)
double T = 3;
double PR = 80;
double MR = 10;
double fPR = 200;
double cap = 100;
```

```

// Calcul des paramètres des paraboles
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; /* le nombre de variables additionnelles */
int compteur=0;
int vraiN=1;
int temp, temp2, compt, compt2;

// calcul de vraiN, p and prof.
// prof et vraiN sont telles que pour tout t < prof et 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); /* nombre total de variables */

/* Remarque : nombre total de blocs =
    vraiN (pour les produits) + (vraiN-1) (relations entre les "fi")
    + liens (capacité des liens) + vraiN (contraintes des MR) *
    + vraiN (contraintes des PR) */
// Remarque : p=2*vraiN-1
nblocs = p+2*vraiN+liens;
printf("%d = nblocs\n",nblocs);

// Ligne donnant la taille des différents blocs
printf("{2}");
for (i=1; i<p;i++) printf(", 2");
for (i=1; i<=2*vraiN+liens; i++) printf(", 1");
printf("}\n");

// vecteur L (à optimiser)
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 à AvraiN */
for (i = 1; i<=vraiN; i++) {
    printf("%d %d 1 2 1\n", i, i); /*lien xi et 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 à 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;
}

/* dernière variable */
printf("%d %d 1 2 1\n", vraiN+p, p);

/* contraintes maximales */
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 sur 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

$$\begin{aligned} J'_1 &\geq J_1, \\ J'_2 + J'_1 &\geq J_2 + J_1, \\ &\vdots \\ \sum_{t=1}^T J'_t &\geq \sum_{t=1}^T J_t. \end{aligned}$$

La relation \succeq est un ordre partiel. Nous avons la propriété suivante :

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 $J'' \succeq J'$, on a $J'' \succeq J^*$.*

Démonstration. Posons

$$\begin{aligned} J_1^* &= \max\{J_1, J'_1\} \\ &\vdots \\ J_t^* &= \max\left(\sum_{k=1}^t J_k, \sum_{k=1}^t J'_k\right) - \sum_{k=1}^{t-1} J_k^* \\ &\vdots \\ J_T^* &= \max\left(\sum_{k=1}^T J_k, \sum_{k=1}^T J'_k\right) - \sum_{k=1}^{T-1} J_k^* \end{aligned}$$

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 équitales : 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, de la forme :

$$\sum_{t=1}^T B_t J_t \leq C$$

avec C la bande passante totale et B_t la bande passante d'une porteuse de type t . Il convient d'être prudent quand à la prise en compte de cette contrainte, comme indiqué dans la remarque suivante :

Remarque. Soit \mathcal{B} l'ensemble des allocations satisfaisant les contraintes de bande passante. Il est possible que pour un $J \in \mathcal{B}$ et $J' \in \mathcal{B}$ nous ayons $\max(J, J') \notin \mathcal{B}$: en d'autres termes, une allocation peut être localement mais pas globalement acceptable.

Dans l'algorithme d'allocation nous considérerons les vecteurs J qui satisfont les contraintes des utilisateurs (contraintes minimales) et qui sont maximaux dans l'ensemble des vecteurs acceptables (c'est-à-dire les vecteurs $J \in \mathcal{B}$ 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 l'algorithme de la section 9.2.1 trouve une allocation optimale des porteuses pour un vecteur fixé donné J .

Dans un contexte général, nous considérons un ensemble fini \mathcal{I} d'éléments aux valeurs positives. Soit v_i la valeur associée à l'élément i .

Définition B.1. Soient \mathcal{A} et \mathcal{B} deux sous-ensembles disjoints de \mathcal{I} . On note alors $Union(\mathcal{A}, \mathcal{B}, k)$ un sous ensemble de \mathcal{I} contenant au plus k éléments de $\mathcal{A} \cup \mathcal{B}$ tels que $\sum_{i \in \mathcal{I}} v_i$ est maximum.

Notons que si $\mathcal{I} = \{1, \dots, p\}$ et $v_1 \geq \dots \geq v_p$, alors $Union(\mathcal{A}, \mathcal{B}, k)$ se compose des k plus petits indices de $\mathcal{A} \cup \mathcal{B}$.

Définition B.2. Soit \mathcal{A} et \mathcal{B} deux sous ensembles de \mathcal{I} . On dit que \mathcal{A} domine \mathcal{B} ssi $|\mathcal{A}| \geq |\mathcal{B}|$ et il existe une injection φ de \mathcal{B} à \mathcal{A} telle que

$$\forall i \in \mathcal{B} \quad v_i \leq v_{\varphi(i)}.$$

Proposition B.2. Soit \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} des sous ensembles de \mathcal{I} tels que

- \mathcal{A} domine \mathcal{B} ,
- \mathcal{C} domine \mathcal{D} ,
- $\mathcal{A} \cup \mathcal{B}$ et $\mathcal{C} \cup \mathcal{D}$ sont disjoints.

Alors, pour tout $k \in \mathbb{N}$, $Union(\mathcal{A}, \mathcal{C}, k)$ domine $Union(\mathcal{B}, \mathcal{D}, k)$.

Démonstration. Soit φ_1 une injection de \mathcal{B} vers \mathcal{A} , et φ_2 de \mathcal{D} vers \mathcal{C} . On défini ψ de $\mathcal{B} \cup \mathcal{D}$ vers $\mathcal{A} \cup \mathcal{C}$ comme suit :

$$\begin{aligned} \psi(v) &\mapsto \varphi_1(v) && \text{si } v \in \mathcal{B}, \\ \psi(v) &\mapsto \varphi_2(v) && \text{si } v \in \mathcal{D}. \end{aligned}$$

De façon évidente, la restriction de ψ à $Union(\mathcal{B}, \mathcal{D}, k)$ génère un ensemble d'au plus k éléments dans $\mathcal{A} \cup \mathcal{C}$, qui est par définition dominé par $Union(\mathcal{A}, \mathcal{C}, k)$. ■

Nous considérons maintenant que à chaque élément i de \mathcal{I} est associé un type t et pour tout sous-ensemble \mathcal{A} de \mathcal{I} appelons

$$\mathcal{A}_t = \{i \in \mathcal{A} : \text{type}(i) = t\}.$$

Alors une allocation faisable est un ensemble \mathcal{A} tel que

$$\forall t \in \{1, \dots, T\} \quad |\mathcal{A}_1 \cup \dots \cup \mathcal{A}_t| \leq J_1 + \dots + J_t.$$

Théorème B.1. Soit $\mathcal{A}^{(k)}$ construite comme suit :

$$\begin{aligned} \mathcal{A}^{(0)} &= \emptyset \\ \mathcal{A}^{(k+1)} &= Union(\mathcal{A}^{(t)}, \mathcal{I}_{t+1}, J_1 + \dots + J_{t+1}) \end{aligned}$$

alors $\mathcal{A}^{(T)}$ est une solution de notre problème d'allocation qui domine toutes les autres solutions.

Démonstration. $\mathcal{A}^{(T)}$ est évidemment une solution de notre problème d'allocation. Soit \mathcal{B} une autre allocation faisable. Nous montrons par récurrence que $\mathcal{A}^{(t)}$ domine $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_t$.

En fait, on a $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_{t+1} = Union(\mathcal{B}_1 \cup \dots \cup \mathcal{B}_t, \mathcal{B}_{t+1}, J_1 + \dots + J_{t+1})$. Clairement \mathcal{I}_{t+1} domine \mathcal{B}_{t+1} , et par récurrence $\mathcal{A}^{(t)}$ domine $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_t$. D'où le résultat par la proposition B.2. ■

Annexe C

Bibliographie

- [1] Cost 239. « Ultra High Capacity Optical Transmission Networks ». ISBN 953-184-013-X (1999). Rapport final du projet COST 239.
- [2] Micah Adler, Jin-Yi Cai, Jonathan K. Shapiro, et Don Towsley. « Estimation of congestion price using probabilistic packet marking ». *Preprint* (2002).
- [3] Sara Alouf, Eitan Altman, Jérôme Galtier, Jean-François Lalande, et Corinne Touati. « Un algorithme d'allocation de bande passante satellitaire ». Rapport technique, INRIA, 2003. (à paraître).
- [4] Tansu Alpcan, Tamer Başar, Rayadurgam Srikant, et Eitan Altman. « CDMA uplink power control as a noncooperative game ». Dans *IEEE Conference on Decision and Control* (Orlando, Florida, décembre 2001), pp. 197–202.
- [5] Eitan Altman. « Capacity of multi-service CDMA cellular networks with best-effort applications ». Dans *Mobicom' 02* (2002).
- [6] Eitan Altman, Afonso Ferreira, et Jérôme Galtier. *Les réseaux satellitaires de télécommunication*, Dunod edition. SNEL, 1999.
- [7] Gunnar Andersson, Lars Engebretsen, et Johan Håstad. « A new way to use semidefinite programming with applications to linear equations mod p ». *Journal of Algorithms* **39** (février 2002), 162–204.
- [8] Ambalavanar Arulambalam et Xiaoqiang Chen. « Allocating fair rates for available bit rate service in ATM networks ». *IEEE Communications Magazine* (1996), 92–100.
- [9] Sanjeewa Athuraliya, Victor H. Li, Steven H. Low, et Qinghe Yin. « REM : active queue management ». *IEEE Network* **15**, numéro 3 (2001), 48–53. Version détaillée dans les actes de ITC'17, Salvador, Brésil, Sept. 2001.
- [10] Deepak Ayyagari et Anthony Ephremides. « Power control for link quality protection in cellular DS-CDMA networks with integrated (packet and circuit) services ». Dans *Mobicom' 99* (Seattle Washington, USA, 1999), pp. 96–102.
- [11] Andrea Baiocchi, Francesca Cuomo, et Cristina Martello. « Optimizing the radio resource utilization of multiaccess systems with a traffic-transmission quality adaptive packet scheduling ». *Computer Networks* **38** (2002), 225–246.

- [12] Randall Berry. *Power and delay trade-offs in fading channels*. Thèse de doctorat, Massachusetts Institute of Technology, juin 2000. Disponible sur <http://www.ece.nwu.edu/~rberry/pubs.html>.
- [13] Dimitri Bertsekas et Robert Gallager. *Data Networks*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
- [14] Vaduvur Bharghavan, Songwu Lu, et Thyagarajan Nandagopal. « Fair queuing in wireless networks, issues and applications ». *IEEE Personal Communications* **6**, numéro 1 (février 1999).
- [15] Vineet Birmani. « Master of Science ». Rapport de DEA, University of Maryland, 1999.
- [16] Thomas Bonald et Laurent Massouillé. « Impact of fairness on Internet performance ». Dans *Sigmetrics* (2001).
- [17] Giancarlo Bongiovanni, Don Coppersmith, et Chung Kei Wong. « An optimum time slot assignment algorithm for an SS/TDMA system with variable number of transponders ». *IEEE Trans. Communications* (mai 1981), 721–726.
- [18] Maurizio A. Bonuccelli. « A fast time slot assignment algorithm for TDM hierarchical switching systems ». *IEEE Trans. Communications* **37** (1989), 870–874.
- [19] Maurizio A. Bonuccelli, Inder Gopal, et C. K. Wong. « Incremental time-slot assignment in SS/TDMA satellite systems ». *IEEE Trans. Communications* **39** (juillet 1991), 1147–1156.
- [20] Jean-Yves Le Boudec. *Rate Adaptation, Congestion Control and Fairness : A Tutorial*. Ecole Polytechnique Federale de Lausanne, décembre 2000.
- [21] Stephen Boyd, Laurent El Ghaoui, Eric Feron, et Venkataramanan Balakrishnan. *Linear matrix inequalities in system and control theory*. Springer, 1994.
- [22] Joseph P. Burke et James R. Zeidler. « Data Throughput in a Multi-Rate CDMA Reverse Link : Comparing Optimal Spatial Combining vs. Maximal Ratio Combining ». Dans *IEEE Globecom* (San Antonio, novembre 2001), volume Vol. 1, pp. 700–705.
- [23] Joseph S. Campanella et Roger J. Colby. « Network Control for Multibeam TDMA and SS/TDMA ». *IEEE J. on Sel. Areas in Communications* **1** (janvier 1983), 174–187.
- [24] Xi Ren Cao. « Preference functions and bargaining solutions ». Dans *IEEE CDC* (Florida, décembre 1982), pp. 164–171.
- [25] Xi-Ren Cao, Hong-Xia Shen, Rodolfo Milito, et Patrica Wirth. « Internet Pricing With a Game Theoretical Approach : Concepts and Examples ». *IEEE/ACM Transactions on Networking* **10**, numéro 2 (avril 2002), 208–216.
- [26] Gregory W. Cermak. « Subjective Quality of Speech Over Packet Networks as a Function of Packet Loss, Delay and Delay Variation ». *International Journal of Speech Technology* **5**, numéro 1 (janvier 2002), 65–84. ISSN 1381-2416.

-
- [27] Suresh Chalasani et Anujan Varma. « Efficient time-slot assignment algorithms for SS/TDMA systems with variable-bandwidth beams ». *IEEE Transactions on Communications* **42**, numéro 2/3/4 (février 1994), 1359–1370.
- [28] Suresh Chalasani et Anujan Varma. « Efficient Time-Slot Assignment Algorithms for SS/TDMA Systems with Variable-Bandwidth Beams ». *IEEE Trans. Communications* **42** (février/mars/avril 1994), 1359–1370.
- [29] Anna Charny, David Clark, et Raj Jain. « Congestion Control With Explicit Rate Indication ». Dans *IEEE International Conference on Communications* (1995), pp. 1954–1963.
- [30] Wen-Tsuen Chen, Pi-Rong Sheu, et Jiunn-Hwa Yu. « Time Slot Assignment in TDM Multicast Switching Systems ». *IEEE Trans. Communications* **COM-42** (janvier 1994), 149–165.
- [31] Michael Cheung et Jon W. Mark. « Resource allocation in wireless networks based on joint packet/Call levels QoS constraints ». Dans *IEEE Globecom* (novembre 2000), pp. 271–275.
- [32] Dah-Ming Chiu et Raj Jain. « Analysis of the increase and decrease algorithms for congestion avoidance in computer networks ». *Computer Networks and ISDN systems* **17** (1989), 1–14.
- [33] Jaeweon Cho et Daehyoung Hong. « Downlink throughput and fairness analysis in multi-rate CDMA systems ». *IEEE VTC* (2001), 1001–1005.
- [34] Jon Crowcroft et Philippe Oechslin. « Differentiated end-to-end Internet services using a weighted proportional fair sharing TCP ». *ACM Sigcomm Computer Communication Review* **28**, numéro 3 (juillet 1998), 53–69.
- [35] Ahmed K. Elhakeem, Stephan Bohm, Mohamed Hachicha, Tho Le-Ngoc, et H. T. Mouftah. « Analysis of a New Multiaccess/Switching Technique for Multibeam Satellites in a Prioritized ISDN Environment ». *IEEE J. Selected Areas in Comm.* **10** (février 1992), 378–390.
- [36] Kai Y. Eng et Anthony S. Acampora. « Fundamental Conditions Governing TDM Switching Assignments in Terrestrial and Satellite Networks ». *IEEE Trans. on Communications* **35** (juillet 1987), 755–761.
- [37] Mirko Ferracoli, Velio Tralli, et Roberto Veerdone. « Channel adaptive scheduling for a wideband TDD/TCDMA wireless system under heterogeneous traffic conditions ». *Computer Networks* **38** (2002), 207–223.
- [38] Sally Floyd. « Congestion control principles ». Internet draft RFC 2914, octobre 1999.
- [39] Errin W. Fulp et Douglas S. Reeves. « The Fairness and Utility of Pricing Network Resources Using Competitive Markets ». Rapport technique, North Carolina State University, 2002.
- [40] Nobuo Funabiki et Yoshiyasu Takefuji. « A Parallel Algorithm for Time-Slot Assignment Problems in TDM Hierarchical Switching Systems ». *IEEE Trans. Communications* **COM-42** (octobre 1994), 2890–2898.

- [41] NATALIE GIROUX, éditeur. *Traffic Management Specification, Version 4.0*. AF-TM-0056.000. ATM Forum Traffic Management Working Group, avril 1996.
- [42] Michel Goemans et David Williamson. « Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming ». *Journal of ACM* **42** (1995), 1115–1145.
- [43] Inder S. Gopal, Maurizio A. Bonuccelli, et Chung K. Wong. « Scheduling in Multi-beam Satellites with Interfering Zones ». *IEEE Trans. Communications* **COM-31** (août 1983), 941–951.
- [44] Inder S. Gopal, Don Coppersmith, et Chung K. Wong. « Minimizing Packet Waiting Time in a Multibeam Satellite System ». *IEEE Trans. Communications* **COM-30** (février 1982), 305–316.
- [45] Inder S. Gopal et Chung K. Wong. « Minimizing the Number of Switchings in an SS/TDMA System ». *IEEE Trans. Communications* **COM-33** (juin 1985), 497–501.
- [46] Martin Grötschel, László Lovász, et Alexander Schrijver. « The ellipsoid method and its consequences in combinatorial optimization ». *Combinatorica* **1** (1981), 169–197.
- [47] Ellen L. Hahne. « Round-Robin Scheduling for Max-Min Fairness in Data Networks ». *IEEE Journal on Selected Area in Communications* **9**, numéro 7 (septembre 1991), 1024–1039.
- [48] Bruce Hajek. « Cooling schedules for optimal annealing ». *Mathematics of Operations Research* **13**, numéro 2 (mai 1988).
- [49] Bassam Hashem et Halim Yanikomeroglu. « Power control for Code-Division Multiple Access Cellular Systems ». Dans *International Congress on Dynamics and Control of Control Systems (DYCONS'99)* (Ottawa, Canada, août 1999).
- [50] Christoph Helmberg. *Semidefinite Programming for Combinatorial Optimization*. ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustraße 7, D-14195 Berlin, Allemagne, octobre 2000. Habilitationsschrift, disponible sur <http://www.zib.de/helmberg>.
- [51] Thomas H. Henderson, Emile Sahouria, Steven McCanne, et Randy H. Katz. « On Improving the Fairness of TCP Congestion Avoidance ». Dans *IEEE Globecom conference* (Sydney, 1998).
- [52] Enrique J. Hernandez-Valencia et Lbachir Benmohamed. « Rate Control Algorithms for the ATM ABR service ». *European Trans. Telecom* **8** (1997), 7–20.
- [53] Kimmo Hiltunen et Riccardo de Bernardi. « WCDMA downlink capacity estimation ». Dans *VTC' 2000* (Tokyo, Japan, 2000), volume 992–996.
- [54] Harri Holma et Antti Toskala. *WCDMA for UMTS*. Revised Edition, J. Wiley & Sons, 2001.
- [55] Thomas Inukai. « An Efficient SS/TDMA Time Slot Assignment Algorithm ». *IEEE Trans. Communications* **27**, numéro 10 (octobre 1979), 1449–1455.

-
- [56] Yasuhiko Ito, Yoshiyori Urano, Takuro Muratani, et M. Yamaguchi. « Analysis of a switch matrix for an SS/TDMA system ». Dans *IEEE* (mars 1977), volume 65, pp. 411–419.
- [57] Bijan Jabbari et David McDysan. « Performance of Demand Assignment TDMA and Multicarrier TDMA Satellite Networks ». *IEEE J. on Sel. Areas in Communications* **10** (février 1992), 478–486.
- [58] Van Jacobson et Michael J. Karels. « Congestion avoidance and Control ». *ACM Computer Communication Review* **18** (1988), 314–329. Egalement dans les actes de Sigcomm' 88, Stanford, CA, Août 1988.
- [59] Jan Janssen, Danny De Vieeschauwer, Maarten Büchli, et Guido H. Petit. « Assessing Voice Quality in Packet Based Telephony ». *IEEE Internet Computing* **6**, numéro 3 (2002), 48–56.
- [60] Koushik Kar, Saswati Sarkar, et Leandros Tassiulas. « Optimization based rate control for multirate multicast sessions ». Dans *IEEE Infocom' 01* (Anchorage, Alaska, avril 2001), pp. 123–132.
- [61] Koushik Kar, Saswati Sarkar, et Leandros Tassiulas. « A Simple Rate Control Algorithm for Maximizing Total User Utility ». Dans *IEEE Infocom' 01* (Anchorage, Alaska, avril 2001).
- [62] Shuzo Kato, Takemi Arita, et Kozo Morita. « Onboard Digital Signal Processing Technologies for Present and Future TDMA Arid SCPC Systems ». *IEEE J. on Sel. Areas in Communications* **5** (mai 1987), 685–700.
- [63] Shuzo Kato, Masahiro Morikura, et Shuji Kubota et al. « A TDMA Satellite Communication System for ISDN Services ». *IEEE J. Sel. Areas in Comm.* **10** (février 1992), 456–464.
- [64] Frank P. Kelly. « Charging and rate control for elastic traffic ». *European Trans. on Telecom.* **8** (1998), 33–37.
- [65] Frank P. Kelly, Aman Maulloo, et David Tan. « Rate control for communication networks : shadow prices, proportional fairness and stability ». *Journal of the Operational Research Society* **49** (1998).
- [66] D. Kennedy et al. « TDMA burst scheduling within the INTELSAT system ». Dans *Globecom' 82* (Nov. 1982), pp. 1263–1267.
- [67] Seong-Lyun Kim, Zvi Rosberg, et Jens Zander. « Combined power control and transmission selection in cellular networks ». Dans *IEEE Vehicular Technology Conference* (1999).
- [68] C. King et al. « INTELSAT TDMA/DSI burst time plan development ». *Int. J. Satellite Commun.* **3**, numéro 1-2 (1985), 35–43.
- [69] Scott Kirkpatrick, Daniel C. Gelatt Jr., et Mario P. Vecchi. « Optimization by Simulated Annealing ». *Science* **220**, numéro 4598 (1983), 671–679.
- [70] Jon Kleinberg et Michel Goemans. « The Lovasz theta function and a semi-definite programming relaxation of vertex cover ». *SIAM J. Discrete Math* **11** (1998).

- [71] Srisankar Kunniyur et Rayadurgam Srikant. « A time scale decomposition approach to adaptive ECN marking ». Dans *Infocom* (2001), pp. 1330–1339.
- [72] Srisankar Kunniyur et Rayadurgam Srikant. « A time scale decomposition approach to adaptive ECN marking ». Dans *IEEE Infocom 01* (Anchorage, Alaska, avril 2001).
- [73] Jaana Laiho et Achim Wacker. « Radio network planning process and methods for WCDMA ». *Ann. Telecommun.* **56**, numéro 5-6 (2001).
- [74] Monique Laurent et Francz Rendl. « Semidefinite Programming and Integer Programming ». Une version préliminaire est parue en avril 2002, rapport PNA-R0210, CWI, Amsterdam, décembre 2002.
- [75] Rangan Leelahakriengkrai et Rajeev Agrawal. « Scheduling in Multimedia CDMA Wireless Networks ». Rapport technique ECE-99-3, ECE Dept., University of Wisconsin - Madison, juillet 1999.
- [76] Rangan Leelahakriengkrai et Rajeev Agrawal. « Scheduling in Multimedia DS-SS Wireless Networks ». Rapport technique ECE-99-3, ECE Dept., University of Wisconsin - Madison, juillet 1999.
- [77] James L. Lewandowski, Jane W. S. Liu, et C. L. Liu. « SS/TDMA Time Slot Assignment with Restricted Switching Modes ». *IEEE Trans. Communications* **31** (janvier 1983), 149–154.
- [78] Abdel Lisser et Franz Rendl. « Telecommunication clustering using linear and semidefinite programming ». *Math. Programming* (2001). A paraître.
- [79] Lázló Lovász. « On the Shannon capacity of a graph ». *IEEE trans. on Information Theory* **25** (1979).
- [80] Steven H. Low, Larry L. Peterson, et Limin Wang. « Understanding TCP vegas : a duality model ». Dans *Sigmetrics/Performance* (2001), pp. 226–235.
- [81] Steven H. Low et Rayadurgam Srikant. « A Mathematical Framework for Designing a Low-Loss, Low-Delay Internet ». *Network and Spatial Economics* (2003). A paraître.
- [82] Songwu Lu, Vaduvur Bharghavan, et Rayadurgam Srikant. « Fair scheduling in wireless packet networks ». *IEEE/ACM Trans. on Networking* **7**, numéro 4 (1999), 473–489. Egalement ACM SIGCOMM Août 1997.
- [83] Duncan Luce et Howard Raiffa. *Games and Decisions*. New York : Wiley, 1957.
- [84] Jon W. Mark et Shihua Zhu. « Power control and rate allocation in multirate wideband CDMA systems ». *IEEE WCNC* (septembre 2000), 168–172.
- [85] Laurent Massoulié et James W. Roberts. « Bandwidth sharing and admission control for elastic traffic ». *Telecommunication Systems* (2000).
- [86] Laurent Massoulié et James Roberts. « Bandwidth sharing : objectives and algorithms ». Dans *IEEE Infocom* (1999), volume 3, pp. 1395–1403.
- [87] Ravi Mazumdar, Lorne G. Mason, et Christos Douligeris. « Fairness in network optimal flow control : optimality of product forms. ». *IEEE Trans. on Comm.* **39** (1991), 775–782.

-
- [88] Jeremy W. McLean et Hugo J. Woerdeman. « Spectral factorization and sums of squares representations via semidefinite programming ». *SIAM J. Matrix Anal. Appl.* **23**, numéro 3 (mai 2001), 646–655.
- [89] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, et Edward Teller. « Equation of State Calculations by Fast Computing Machines ». *The Journal of Chemical Physics* **21**, numéro 6 (juin 1953), 1087–1092.
- [90] Gyorgy Miklós et Sándor Molnár. « Fair allocation of elastic traffic for a wireless base station ». Dans *IEEE Globecom* (décembre 1999), pp. 1673–1678. Rio de Janeiro.
- [91] Michel Minoux et Christian Brouder. « Models and algorithms for optimal traffic assignment in SS/TDMA switching systems ». *Int. J. Satellite Commun.* **5** (1987), 33–47.
- [92] Takeshi Mizuike, Yasuhiko Ito, Lan N. Nguyen, et Eijiro Maeda. « Computer-Aided Planning of SS/TDMA Network Operation ». *IEEE J. on Sel. Areas in Communications* **9** (janvier 1991), 37–47.
- [93] Jeonghoon Mo et Jean Walrand. « Fair end-to-end window-based congestion control ». Dans *SPIE'98, International symposium on voice, video and data communications* (1998).
- [94] Jeonghoon Mo et Jean Walrand. « Fair End-to-End Window-based Congestion Control ». Dans *SPIE' 98, International Symposium on Voice, Video and Data Communications* (1998).
- [95] Abhinay Muthoo. *Bargaining theory with applications*. Cambridge University Press, 1999.
- [96] John Nagle. « Congestion Control in IP/TCP Internetworks ». Internet draft RFC 896, janvier 1984.
- [97] John F. Nash Jr. « The bargaining problem ». *Econometrics* **18** (1950), 155–162.
- [98] Arkadi Nemirovski. « What can be expressed via Conic Quadratic and Semidefinite Programming? ». *Lecture notes* (1998).
- [99] Arkadi Nemirovski. « Five Lectures On Modern Convex Optimization ». CORE Summer School on Modern Convex Optimization, août 2002.
- [100] Aleksandar Neskovic, Natasa Naskovic, et George Paunovic. « Modern Approaches in Modeling of Mobile Radio Systems Propagation Environment ». *IEEE Communications surveys* (2000), 2–12.
- [101] Seong-Jun Oh et Kimberly M. Wasserman. « Adaptive resource allocation in power constrained CDMA mobile networks ». Dans *IEEE WCNC* (septembre 1999), pp. 510–514.
- [102] Seong-Jun Oh et Kimberly M. Wasserman. « Optimality of greedy power control and variable spreading gain in multi-class CDMA mobile networks ». Dans *Mobicom' 99* (Seattle, Washington, USA, 1999), pp. 102–112.

- [103] Seong-Jun J. Oh et Kymberly M. Wasserman. « Dynamic spreading gain control in multiservice CDMA networks ». *IEEE J. Selected Areas in Comm.* **17**, numéro 5 (mai 1999), 918–927.
- [104] Guillermo Owen. *Game theory*. San Diego, CA : Academic, 1995.
- [105] Larry C. Palmer et Lawrence W. White. « Demand Assignment in the ACTS LBR System ». *IEEE Trans. Communications* **COM-38** (mai 1990), 684–692.
- [106] Jung M. Park, Uday R. Savagaonkar, et Edwin K.P. Chong et al. « Efficient Resource Allocation for QoS Channels in MF-TDMA Satellite Systems ». Dans *IEEE Military Communications Conference* (octobre 2000), pp. 1–5. Section U19.6.
- [107] Michael Pióro, Gábor Malicskó, Gábor Fodor, et Tomasz Szymanski. « Optimal Link Capacity Dimensioning in Proportionally Fair Networks ». Dans *LNCS 2345* (2002), E. G. et al., éditeur, pp. 277–288. Networking' 02.
- [108] Sudhir Ramakrishna et Jack M. Holtzman. « A Scheme for Throughput Maximization in a Dual-Class CDMA System ». *IEEE J. Selected Areas in Comm.* **16** (août 1998), 830–844.
- [109] Kadangode Ramakrishnan et Sally Floyd. « A Proposal to add Explicit Congestion Notification (ECN) to IP ». Internet draft RFC 2481, janvier 1999.
- [110] Motakuri V. Ramana. « An Exact Duality Theory for Semi-definite Programming and its Complexity Implications ». *Math. Programming* (1996).
- [111] R. Ramaswamy et Pulak Dhar. « Comments on An Efficient SS/TDMA Time Slot Assignment Algorithm ». *IEEE Tran. Communications* **32** (Sept. 1984), 1061–1065.
- [112] Alvin E. Roth. *Axiomatic Models of Bargaining*. New York : Wiley, 1979.
- [113] Dan Rubenstein, James F. Kurose, et Donald F. Towsley. « The Impact of Multicast Layering on Network Fairness ». Dans *Sigcomm* (1999), pp. 27–38. Egalement rapport de recherche 99-08, disponible sur <http://www-net.cs.umass.edu/papers/papers.html>.
- [114] Scott Shenker. « Fundamental Design Issues for the Future Internet ». *IEEE J. on Selected Areas in Comm.* **13**, numéro 7 (1995), 1176–1188.
- [115] A. K. Sinha. « A model for TDMA burst assignment and scheduling ». *COMSAT Tech.Rev.* **6** (Automne 1976).
- [116] Anton Ştefănescu et Maria V. Ştefănescu. « The Arbitrated Solution for Multi-objective Convex Programming ». *Rev. Roum. Math. Pure Applicat.* **29**, numéro 7 (1984), 593–598.
- [117] Chi W. Sung et Wing S. Wong. « Power control and rate management for wireless multimedia CDMA systems ». *IEEE Trans. on Communications* **49**, numéro 7 (juillet 2001), 1215–1226.
- [118] Benjamin Teitelbaum, Susan Hares, Larry Dunn, et al. « Internet2 QBone – Building a Testbed for differentiated services ». *IEEE Network* **13**, numéro 5 (septembre/octobre 1999), 8–16.

-
- [119] Yiu K. Tham. « Burst assignment for satellite-switched and earth-station frequency-hopping TDMA networks ». Dans *IEE* (août 1990), volume 137 (I :4), pp. 247–255.
- [120] Yiu K. Tham. « On Fast Algorithms for TDM Switching Assignments in Terrestrial and Satellite Networks ». *IEEE Trans. on Communications* **43**, numéro 8 (août 1995), 2399–2404.
- [121] William Thomson. « Nash’s bargaining solution and utilitarian choices rules ». *Econometrica* **49**, numéro 2 (1981), 535–538.
- [122] Bjarne Toft. *Handbook of combinatorics*, graham, grötschel, lovász edition, volume 1 (4). North Holland, 1995, chapitre Coloring, stable sets and perfect graphs, pp. 233–288.
- [123] Tuna Toker, Eitan Altman, Jérôme Galtier, et Corinne Touati. « Slot Allocation in a TDMA Satellite System : Simulated Annealing Approach ». Dans *AIAA International Communication Satellite Systems Conference and Exhibit (ICSSC)* (Yokohama, Japon, 2003).
- [124] Corinne Touati, Eitan Altman, et Jérôme Galtier. « Fair power and transmission rate control in wireless networks ». Dans *Globecom 2002* (2002), volume 2, pp. 1229–1233.
- [125] Milan Vojnovic, Jean-Yves Le Boudec, et Catherine Boutremans. « Global Fairness of Additive-Increase and Multiplicative-Decrease with Heterogeneous Round-Trip Times ». Dans *Infocom* (2000), volume 3, pp. 1303–1312.
- [126] Liang Xu, Xuemin Shen, et Jon W. Mark. « Performance analysis of adaptive rate and power control for data service in DS-CDMA systems ». Dans *IEEE Globecom* (novembre 2001), pp. 627–631.
- [127] Masami Yabusaki et Shigefusa Suzuki. « Approximate Performance Analysis and Simulation Study for Variable-Channel-Per-Burst SS-TDMA ». *IEEE Trans. Communications* **38** (mars 1990), 318–326.
- [128] Haïkel Yaïche, Ravi R. Mazumdar, et Catherine Rosenberg. « A game theoretic framework for bandwidth allocation and pricing in broadband networks ». *IEEE/ACM Transactions on Networking* **8**, numéro 5 (2000), 667–677.
- [129] Roy D. Yates. « A Framework for Uplink Power Control in Cellular Radio Systems ». *IEEE Journal on Selected Areas in Communications* **13**, numéro 7 (septembre 1995), 1341–1348.
- [130] Kwan L. Yeung. « Efficient time slot assignment algorithms for TDM hierarchical and nonhierarchical switching systems ». *IEEE Trans. Communications* **49** (février 2001), 351–359.
- [131] James Yoh, Charles C. Wang, et Gary W. Goo. « Survey of Ka-band satellites for wideband communications ». Dans *IEEE Military Communications Conference* (1999), volume 18, pp. 120–125.
- [132] Dongmei Zhao, Xuemin Shen, et Jon W. Mark. « Quality-of-service support by power and rate allocation in WCDMA systems ». Dans *IEEE Globecom* (novembre 2001), pp. 604–608.

- [133] Tolga Örs, Zhili Sun, et Barry G. Evans. « A Meshed VSAT Satellite Architecture using an On-Board ATM Switch ». Dans *IEEE International Performance Computing and Communication Conference* (Tempe/Phoenix, Arizona, février 1997), pp. 208–214. ISBN :0-78-033874-X.

Annexe D

Publications

Brevet

- [1] J. Kelif, C. Touati, J. Galtier, E. Altman et B. Fourestie. « Stratégies d’allocations de débits et d’admission d’appels pour les mobiles d’un réseau CDMA ou W-CDMA ». dépôt de brevet en cours.

Revue internationale avec comité de lecture

- [2] C. Touati, E. Altman et J. Galtier. « Nash Bargaining Allocations with Applications to Mobile and Satellite Communications ». Dans *Tenth International Symposium on Dynamic Games and Applications, Workshop «Networking games & resource allocation»* (Petrozavodsk, Russie, 12-15 juil. 2002). Une version détaillée est en cours de révision au journal *Computer Networks* (Elsevier).
- [3] C. Touati, E. Altman et J. Galtier. « Semi-definite Programming Approach for Bandwidth Allocation and Routing in Networks ». *Game Theory and Applications* 9 (2003), 169–179. Présenté également au Workshop «Networking games & resource allocation» de Tenth International Symposium on Dynamic Games and Applications, 12-15 juil. 2002, Petrozavodsk, Russie.

Conférences internationales avec comité de lecture

- [4] C. Touati, E. Altman et J. Galtier. « Utility Based Fair Bandwidth Allocation ». Dans *IASTED International Conference on Networks, Parallel and Distributed Processing and Applications (NPDPA 2002)* (Tsukuba, Japon, Oct. 2002), pp. 126–131.
- [5] C. Touati, E. Altman et J. Galtier. « Fair power and transmission rate control in wireless networks ». Dans *IEEE Globecom 2002* (Taipei, Taiwan, Nov. 2002), volume 2, pp. 1229–1233.

- [6] C. Touati, E. Altman et J. Galtier. « Radio planning in Multibeam Geostationary Satellite Networks ». Dans *AIAA International Communcation Satellite Systems Conference and Exhibit (ICSSC 2003)* (Yokohama Japon, Avr. 2003).
- [7] T. Toker, E. Altman, J. Galtier, C. Touati, I. Buret, B. Fabre et C. Guiraud. « Slot Allocation in a TDMA Satellite System : Simulated Annealing Approach ». Dans *AIAA International Communcation Satellite Systems Conference and Exhibit (ICSSC 2003)* (Yokohama Japon, Avr. 2003).
- [8] C. Touati, E. Altman et J. Galtier. « Semi-Definite Programming for Bandwidth Allocation in Communication Networks ». Dans *ITC'18* (Berlin, Allemagne, Sept. 2003).

Rapports de recherche

- [9] C. Touati, E. Altman et J. Galtier. « On Fairness in Bandwidth Allocation ». Rapport technique RR 4269, INRIA, Sept 2001.
- [10] C. Touati, E. Altman et J. Galtier. « Fair Bandwidth Allocation between Providers in a Geostationary Satellite Network ». Rapport technique RR 4421, INRIA, Sept 2001.
- [11] S. Alouf, E. Altman, J. Galtier, J. Lalande et C. Touati. « Un algorithme d'allocation de bande passante satellitaire ». Rapport technique, INRIA, Sept 2001. (A paraître).

Les principes d'équité appliqués aux réseaux de télécommunications

Résumé :

Les travaux de cette thèse portent sur l'allocation équitable de ressources dans les réseaux de télécommunication. Il a été montré que la satisfaction relative qu'apporte une allocation à un utilisateur dépend de l'application utilisée, ce qui n'est pas pris en compte par les protocoles actuellement utilisés. Nous avons adapté les critères d'équité issus de la théorie des jeux afin de prendre en compte les particularités inhérentes aux réseaux de télécommunications, et en particulier aux besoins des applications temps réel et avons proposé des méthodes algorithmiques de résolution adaptées à des exemples de réseaux de trois types : filaires, mobiles et satellitaires. En particulier, en nous inspirant du NBS (Nash Bargaining Solution), nous avons adapté une large famille de critères d'équité (couvrant entre autres les connus d'équité max-min et d'équité proportionnelle) afin de prendre en compte les fonctions d'utilité des applications. Ces dernières représentent la satisfaction relative qu'apporte l'allocation aux utilisateurs.

Mots-clés :

Réseaux, allocation de ressources, théorie des jeux, équité, algorithmique.